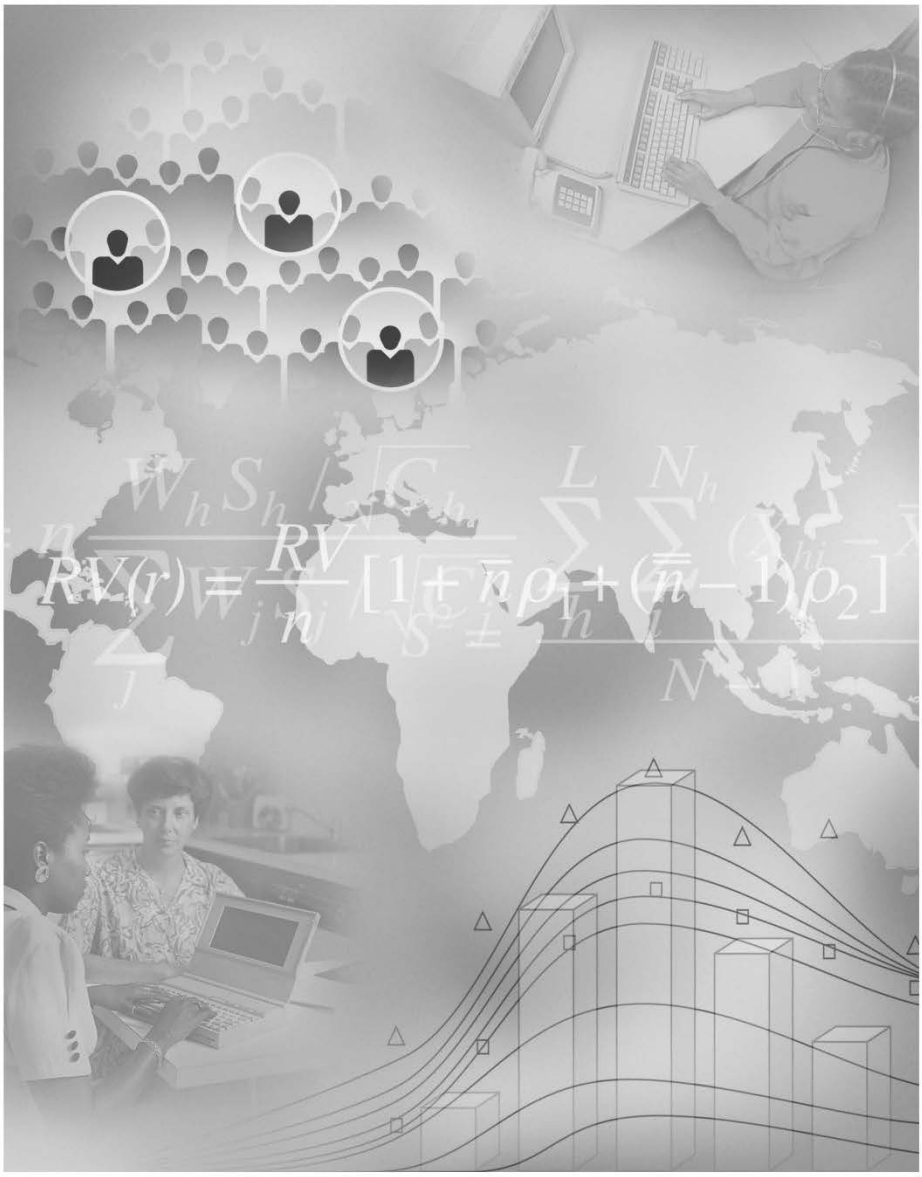
the Survey Statistician

The Newsletter of the International Association of Survey Statisticians

No. 77 **January 2018**



INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

Editors:
Natalie Shlomo and Eric Rancourt

Section Editors:
Peter Wright – Country Reports
Ken Copeland – Ask the Experts
Risto Lehtonen – New and Emerging Methods
Natalie Shlomo – Book & Software Review

Circulation/Production:
Lori Young
Nick Husek

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@manchester.ac.uk) or Eric Rancourt Eric.Rancourt@Canada.ca).

ISSN 2521-991X

In this Issue:

Letter from the Editors

The January 2018 issue of *The Survey Statistician* contains articles of interest and important information regarding upcoming conferences, contents of relevant journals, updates from the IASS Executive and more. We hope you enjoy this issue.

Natalie Shlomo will be leaving as co-editor of IASS *The Survey Statistician* due to other commitments on the ISI Executive Council. She has been editing the newsletter since 2010 and has greatly enjoyed her role and ensuring that our periodical has interesting methodological content as well as our society's news. As of the next issue in July 2018, the new co-editor will be Danutė Krapavickaitė (danute.krapavickaite@vgtu.lt) and we welcome her to the editorial board. Please send to Danutė or Eric (Eric.Rancourt@Canada.ca) your feedback and comments on how we can make improvements.

In the *New and Emerging Methods Section* (edited by the Scientific Secretary Risto Lehtonen), Alina Matei (University of Neuchatel) contributed an article titled: On Some Reweighting Schemes for Non-ignorable Unit Nonresponse. In the article, the author presents different estimators under this setting and provides a simulation study comparing weighting approaches and discusses advantages and disadvantages.

The *Ask the Experts* Section, written and edited by Ken Copeland, the following question is addressed: How do Establishment Surveys Differ from Household Surveys? The author provides a number of distinguishing features that are particular to Establishment Surveys.

In the *Book and Software Review Section* (edited by Natalie Shlomo), Brady West (University of Michigan) has reviewed the book: Adaptive Survey Design by B. Schouten, A. Peytchev and J. Wagner published by Chapman and Hall/CRC (2017).

If you would like to contribute an article to New and Emerging Methods Section, please contact Risto Lehtonen (risto.lehtonen@helsinki.fi). 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). You are also welcome to submit your own questions with an answer if you are aware of an important topic of interest. If you are interested in writing a book or software review, please contact Danutė Krapavickaitė (danute.krapavickaite@vgtu.lt). Finally, if you would like to contribute brief articles or editorials to the newsletter, please send them directly to the editors of the newsletter, Eric Rancourt and Danutė Krapavickaitė.

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 contributions. We also thank the editor of the section, Peter Wright (Peter.Wright2@Canada.ca) of Statistics Canada for contacting all country representatives and coordinating the country reports. Please get in touch with Peter if there has been a change in the country representative in order to keep our contact list up-to-date. We ask all country representatives to contribute articles on your

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

This issue of *The Survey Statistician* includes a letter and updates from our IASS President, Peter Lynn and from our Scientific Secretary, Risto Lehtonen. We would like to highlight that the IASS now has a twitter account: @iass_isi. Please start connecting this twitter account to your home institutions and start tweeting.

The *News and Announcement* section includes important information about submitting invited session proposals at the upcoming WSC to be held in Kuala Lumpur in 2019. The online submission system has gone live at http://www.isi2019.org. Please also send your proposals to Cynthia Clark at czfclark@cox.net. In this section, there is also an article containing highlights from the fifth biennial European Establishment Statistics Workshop, organized by the European Network for Better Establishment Statistics (ENBES) as well as updates on the situation of Andreas Georgiou with a statement of support that was circulated among professional societies of which IASS is a signatory. Finally, we have added a news item on 2017 award winners in the area of statistics and survey statistics and wish to congratulate these members for outstanding achievements.

We thank Lori Young from Statistics Canada for collating the advertisements of upcoming conferences and for preparing the tables of contents in the *In Other Journals* section. This is a very time-consuming and detailed task but the information she gathers is deeply appreciated by the members. We also thank Lori in her role as Production Editor and for all her hard work in collating the articles into this issue of *The Survey Statistician*. In addition, we would like to thank Nick Husek from the Australian Bureau of Statistics for final editing of the newsletter and distribution.

Please take an active role in supporting the IASS *The Survey Statistician* 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,

As this is my first letter to the membership of the IASS, I must start by thanking you for electing me to this position. It is humbling that you trust me to take care of the IASS for two years. I will do my best, but the IASS is really in all of our hands, not just mine, so I hope that you will get engaged with the work of the association and that together we can grow the IASS and increase its impact.

I can hardly believe that six months have already passed since I took over the reins from Steve Heeringa at the World Statistics Congress (WSC) last June. The week in Marrakesh was intensely productive. The first meeting I attended – half an hour after arriving in the country – was a joint meeting with the International Statistical Institute (ISI) Executive Committee and the Presidents of all the other associations of the ISI. This was an eye opener for me, not only giving me insight into how the ISI works, but also providing examples of good initiatives and great ideas from other associations that could inform what we do in the IASS. More of that later. We also held meetings of both the outgoing and incoming IASS Executive Committees and Steve and I had a meeting with ISI President Pedro Silva and Director of the ISI Permanent Office Ada van Krimpen to discuss how the permanent office can better support the IASS and how the IASS can better support the objectives of the ISI and of course, the IASS General Assembly took place.

A couple of important decisions were taken at the General Assembly. We decided to slightly increase the IASS membership dues. This is partly to pre-empt a likely increase that will be necessary when the ISI completes a review of their costs in administering the associations and partly a recognition that there is little or no slack in the IASS finances. One way to keep membership fees down, of course, is to spread the cost over more people. So, please take every opportunity to suggest to relevant colleagues that they should join the IASS. If your colleagues are skeptical as to why they should join, you could point them towards a short video that I made while in Marrakech:
https://www.youtube.com/watch?v=HpkNt9liYe8.

A second decision taken at the General Assembly was to reconstitute the IASS Council. This body ceased to formally exist when the IASS adopted new statutes in 2013, reflecting the move of the legislative seat of the association from France to the Netherlands. Unfortunately, to do this will take time as it constitutes an amendment to the statutes, something which can only be done by the General Assembly during a WSC, following a formal proposal with supporting arguments which must be circulated at least 30 days in advance. So, I will ensure that such a formal proposal is made in good time for WSC 2019!

The IASS statutes state that we should hold a General Assembly each Year, so I will therefore convene one in 2018. As well as meeting a formal requirement, this should help us to better keep in touch and keep up the momentum with our various activities. We will use web-based video-conferencing, which should hopefully enable a good number of members to participate. The date and time will be announced soon.

In September we put out a call for requests for support for workshops and conferences taking place in 2018.

The Executive Committee agreed that the IASS should support four events in 2018:

- Conference on Small Area Estimation (celebration of 75th birthday of Danny Pfefferman), Shanghai, June 2018. https://www.sae2018.com

- Second International Conference on the Methodology of Longitudinal Surveys, Essex, July 2018. https://www.understandingsociety.ac.uk/mols2

- Baltic–Nordic–Ukrainian Workshop on Survey Statistics, Jelgava, Latvia, August 2018. https://wiki.helsinki.fi/display/BNU/Events

- Francophone Survey Sampling Colloquium, Lyon, October 2018. http://sondages2018.sfds.asso.fr/

You can find further details of these events elsewhere in this edition of the *Survey Statistician*. As well as providing modest financial support, our support of these events benefits both parties in other ways. Awareness of the IASS is increased through publicity to the participants and on the event website; and the event gains credibility from having the backing of a respected scientific association. If you are involved in organising an international conference or workshop in 2019, on a topic relevant to the IASS, look out for the call for requests for support later this year and please consider applying!

I am delighted that the ISI has been successful in again obtaining funds from the World Bank Trust Fund for Statistical Capacity Building to support the participation of Statisticians from developing countries in conferences, workshops and short courses. The IASS will be applying to the ISI for some of these funds to be allocated to the above-mentioned events that we are supporting in 2018.

One of the first things I did as President was to set up an IASS twitter account. Communication with, and beyond, our global membership is vital. Follow us at @iass_isi.

While it is important to ensure that the IASS continues to function well, we should also remind ourselves from time to time why we do these things. We are living in particularly uncertain times at present. We read on a daily basis of threats to international security, threats to democracy and human rights, and indeed threats to the very future existence of the human race due to the failure of governments to respond adequately to climate change. In these circumstances, it is vital that high-quality, relevant survey statistics continue to be produced, disseminated and understood. Good statistics form the basis of informed debate and informed policy making. Ultimately, that is why we all do what we do.

I wish all IASS members a successful and productive 2018.

Peter Lynn,
IASS President

Report from the Scientific Secretary

The 61th World Statistics Congress of the ISI in July 2017 in Marrakech was a great success for the entire statistics community. IASS was strongly involved in the preparation of ISI WSC 2017, as was reported by Denise Silva in The Survey Statistician, July 2017 issue. Altogether 23 sessions submitted by the IASS community, or organized jointly with another ISI association, were accepted in the programme for Invited Paper Sessions or Special Topic Sessions. This indicates the high interest and success of the IASS community and individual members to contribute to the scientific contents of the event.

The 62th World Statistics Congress of the ISI will be held from 18 to 23 August 2019 in Kuala Lumpur, Malaysia. Preparations for ISI WSC 2019 have already been started, by opening Call for Proposals of Invited Paper Sessions (IPS) and Special Topic Sessions (STS). Instructions for IASS related proposals are published in this TSS issue. They are prepared by Cynthia Clark (czfclark@cox.net), who is our representative on the Scientific Programme Committee (SPC). The submission period for IPS proposals is from 1 January 2018 to 31 March 2018, and from 15 April 2018 to 31 August 2018 for IPS proposals. Detailed guidelines for proposal preparation are published at the ISI WSC 2019 website, see http://www.isi2019.org/call-for-proposal/.

The IASS has been active also in the organization of Short Courses for WSC events. For ISI WSC 2019, it is good time to start thinking about short course proposals. Proposals (with topics, lecturers and organizers) can be sent to me or to the (forthcoming) ISI Short Course Committee.

As in the previous years, supporting scientific conferences and workshops in survey statistics has been an important activity of the IASS. Financial support can be crucial for organizing regional events in particular. Offering reduced registration fees for IASS members has been a precondition for IASS support, increasing attractiveness of IASS membership. The IASS supported conferences and workshops have supplemented nicely the WSC and Satellite Meetings scheme. One of the IASS supported events, the SAE2017 conference in Paris, was organized as a satellite event to the Marrakech WSC.

In 2017 we have supported financially the following scientific events.

The first was a conference honoring Professor J.N.K. Rao on the occasion of his 80th birthday. The conference was held in 24-27 May 2017 in Kunming, China and was hosted by the School of Mathematics and Statistics at Yunnan University. More information on the event is available at:
https://niasra.uow.edu.au/content/groups/public/@web/@inf/@math/documents/mm/uow223641.pdf

The 5th Italian Conference on Survey Methodology (ITACOSM 2017) took place in 14-16 June 2017 in Bologna, Italy. The event was hosted by the Department of Statistical Sciences of the University of Bologna. The website contains abstracts and presentation slides, and more. See details at https://events.unibo.it/itacosm2017.

The 2017 Small Area Estimation Conference (SAE 2017) was arranged as a Satellite Meeting for the ISI WSC 2017 in 10-12 July 2017 in Paris, France. The conference was organized by Ensai (École Nationale de la Statistique et de l'Analyse de l'Information), the CREST (Centre de Recherche en Économie et Statistique) and the ILB (Institute Louis Bachelier). More details, including a Book of Abstracts, are available at http://sae2017.ensai.fr/.

The Baltic-Nordic-Ukrainian Network on Survey Statistics organized its annual Workshop on Survey Statistics Theory and Methodology in 21-24 August 2017 in Vilnius, Lithuania. The workshop was hosted by the Faculty of Fundamental Sciences of the Vilnius Gediminas Technical University and was devoted to celebrate the 80th birthday of Professor Carl-Erik Särndal (see a separate text by Danutė Krapavickaitė in this TSS issue). Access to more details, and Workshop Proceedings, are available at http://vilniusworkshop2017.vgtu.lt/.

EESW17, the fifth biennial European Establishment Statistics Workshop, was organized by the European Network for Better Establishment Statistics (ENBES) in August 30 - September 1, 2017 in Southampton, UK. The workshop was hosted by the University of Southampton (see a separate article by Desislava Nedyalkova in this TSS issue). The ENBES website at https://statswiki.unece.org/display/ENBES offers information on the workshop, including presentations.

Last but not the least, the 4th International Workshop on Surveys for Policy Evaluation and the 5th Brazilian School on Sampling and Survey Methodology (ESAMP V) was held in 17-20 October 2017 in Mato Grosso, Brazil. The website of the event is http://www.redeabe.org.br/esamp2017/.

For 2018, the IASS is continuing to support a limited number of scientific conferences and workshops on survey statistics. This far, the Executive Committee has decided to support the following events in 2018.

- SAE 2018 Conference on Small Area Estimation (A Celebration of 75th Birthday of Professor Danny Pfeffermann), to be arranged in 16-18 June 2018 in Shanghai, China. More details can be found at the conference website https://www.sae2018.com.

- Second International Conference on the Methodology of Longitudinal Surveys will be held in 25-27 July 2018 in Essex, UK. The conference website is https://www.understandingsociety.ac.uk/mols2.

- Baltic–Nordic–Ukrainian (BNU) Workshop on Survey Statistics will be organized in 21-25 August 2018 in Jelgava, Latvia. More information will be available at https://wiki.helsinki.fi/display/BNU/Events.

- Francophone Survey Sampling Colloquium is to be held in 24-26 October 2018 in Lyon, France. The website of the event is http://sondages2018.sfds.asso.fr/.

The IASS Executive Committee will publish later in this year an announcement for applications to support scientific events that will be organized in 2019.

The Survey Statistician is an important media for the IASS community for sharing news and information on events and meetings and also for introducing recent research on important methodological topics. The article by Alina Matei (University of Neuchâtel) in the New and Emerging Methods section, entitled *On some reweighting schemes for nonignorable unit nonresponse* discusses a particular type of nonresponse in surveys: a situation where unit response probabilities depend on the variable of interest that may be missing. This challenging situation is met increasingly often in sample surveys. The article introduces some recent methods to treat this case. The floor is open for papers on new and emerging methods in various areas of survey statistics and related areas.

Please do not hesitate to contact me if you want more information or want to submit an article to the New and Emerging Methods section.

Risto Lehtonen
risto.lehtonen@helsinki.fi

News and Announcements

Request for Proposals for Invited Program Sessions
62nd ISI WSC in Malaysia, August 2019

The planning has begun for the 62nd ISI World Statistics Conference which will be held in Kuala Lumpur, Malaysia from 18-23 August 2019. The official call for Invited Program Sessions is now on the ISI conference website: http://www.isi2019.org/call-for-proposal/. You are invited to submit your proposal directly to the website. The deadline for Invited Program Session proposals is March 31, 2018.

Cynthia Clark, a member of the IASS Executive Committee, is the IASS representative on the Scientific Program Committee. Proposals relevant to the International Association of Survey Statisticians should be designated as such on the submitted proposal. If your proposal is relevant to other ISI sections, you may so designate. Please also send your proposal to Cynthia Clark, czfclark@cox.net. The proposals designated as IASS will be sent to the section for prioritization by the IASS Executive Committee. You should also indicate if you are willing to have your proposal considered as a Special Topic Contributed Session. Those proposals will also be reviewed by the IASS Executive Committee. Any proposals that have initially been sent to Cynthia Clark should now be posted to the ISI website. A copy of the final proposal should also be sent to Cynthia and labeled as SUBMITTED.

Please note that there are specific guidelines for participation in ISI sessions. These will have to be adhered to in the final program. They are provided for your information.

Guidelines for session organisers, chairs, presenters, and discussants

The following participation guidelines apply to organisers, chairs, presenters and discussants of Invited Paper Sessions (IPS), Special Topic Sessions (STS) and Contributed Paper/Poster Sessions (CPS) of the 61st WSC. In the description below, both STS and CPS are referred to as Contributed Papers.

1. Each individual can present only one paper as an oral presenter during the 61st WSC. Exceptions will be granted only in unusual circumstances, and requests must be approved by the Chairs of the Scientific Programme Committee (SPC) and Local Programme Committee (LPC). All requests should be sent to contact@isi2017.org.

2. A co-author of a paper being presented by someone else is not counted as a presenter. Hence it is possible for an individual to be a co-author of multiple papers being presented at the 61st WSC.

3. A participation as discussant is not counted as a participation as presenter.

4. An individual can serve as a chair or a discussant in more than one session provided that there is no schedule conflict.

5. Within the same session, any individual, including the session Organiser, can take up the role of either presenter or chair or discussant, but should not assume any two such roles.

6. Each individual can organise at most one STS. Exceptions must be approved by the Chair of the Local Programme Committee (LPC). All requests should be sent to contact@isi2017.org.

7. All Organisers, Chairs, Presenters and Discussants of any session must be registered participants of the WSC. Exceptions may be granted to organisers who are unable to attend, but not to individuals in any of the other roles.

Highlights from the 2017 fifth biennial European Establishment Statistics Workshop (EESW17)

EESW17, the fifth biennial European Establishment Statistics Workshop, organized by the European Network for Better Establishment Statistics (ENBES) was hosted by the University of Southampton, UK, from August 30 to September 1, 2017.

ENBES was established in 2009 and aims to promote the exchange of ideas on methodology, practices, approaches and tools in the field of business statistics. The field of establishment statistics merits special attention because of a large variation in unit sizes and structures, multinational units, intense population dynamics, and a response process different from that when interviewing persons.

The EESW17 program included 35 papers and posters in 12 sessions, with a total of 48 participating business survey methodologists from official statistics, academia and the private sector. In the following we give highlights of the topics discussed.

The opening session was on sampling designs and included a comparison of sampling designs, use of auxiliary data in the estimation from samples, and the impact of profiling on sampling and estimation. Next followed two sessions on communication in data collection. In particular, the major role of communication and data collection was illustrated through various subjects such as questionnaire design, use of paradata and editing.

Another issue of relevance was covered in two topics on measuring response burden and sample coordination as a tool for spreading the response burden across the statistical units. The speakers discussed how to optimally measure response burden and how to coordinate surveys taking into account the heterogeneous business populations and their dynamics. Tools for sampling design and coordination were presented.

That the choice of unit types about which to produce statistics can have a large impact on the estimates was highlighted in the two sessions on units. Statistical unit types, created within the statistical system, led to discussions about their validity, conceptualization, definition, and implementation. The contributions covered the role of the business register, combining different statistical units in order to improve the quality of the statistical outputs, improving profiling of the enterprises and issues in data integration, for instance when producing statistics for domains/breakdowns (e.g., geographic areas) to which unit types cannot be trivially assigned. A statement on the unit problem, by a working group invited by ENBES, has also been presented at the workshop.

Another interesting topic was that of data visualization and communication with users. Users nowadays are less attracted by traditional output formats, such as tables. The speakers showed visualisations with different effectiveness, developed new, innovative, visualisations, and used gamifications to present statistical output to users.

The benefits of a small workshop such as EESW are the ease of following all the presentations, the ample time allocated for discussion and last but not least the wide range of topics with lots of new ideas as well as some further developments on current subjects.

ENBES's website www.enbes.org provides a link to papers and presentations from EESW17.

We would like to thank our sponsors the International Association of Survey Statisticians (IASS), European Free Trade Association (EFTA) and the UK Office for National Statistics (ONS) for their support, as well as the University of Southampton for hosting the workshop. We encourage you to attend the next workshop which is to be hosted by EUSTAT on 24-27 September 2019 in Bilbao, Spain.

Desislava Nedyalkova,
ENBES Membership and Communications Secretary

AWARDS IN THE FIELD OF STATISTICS, SURVEY STATISTICS AND METHODOLOGY IN 2017

The IASS membership congratulate the following distinguished colleagues for outstanding recognition in the year of 2017:

Wayne Fuller – 2017 Samuel S. Wilks Award for Outstanding Contributions to Statistics (ASA)

Danny Pfeffermann – 2017 West Medal in memory of John Howard West for outstanding contributions to the development or communication of official statistics (Royal Statistical Society)

Jon Rao – 2017 Award for Outstanding Contribution to Small Area Estimation (ISI Satellite Meeting on Small Area Estimation)

Donald Rubin – 2017 Waksberg Award for Statistical Contributions to Survey Methodology (Survey Methodology)

Roderick Little and Donald Rubin – 2017 Karl Pearson Prize for their book, Statistical Analysis with Missing Data (ISI)

Michael Brick – 2017 Monroe G. Sirken Award in Interdisciplinary Survey Methods Research (ASA)

Wendy Martinez, Neil Horton and John Eltinge – 2017 Founders Award (ASA)

Peter Miller – 2017 AAPOR Award for Exceptionally Distinguished Achievement (AAPOR)

Don Dillman, Jolene Smyth and colleagues – 2017 Warren J. Mitofsky Innovators Award (AAPOR)

Donald Rubin – 2017 C.R. and Bhargavi Rao Prize for Outstanding Research in Statistics (Penn State)

SIGN-ON STATEMENT TO HALT LEGAL PROCEEDINGS AGAINST GEORGIOU AND HIS COLLEAGUES

The American Statistical Association circulated a statement for individuals and organizations to sign in support of Andreas Georgiou—the former head of the Hellenic Statistical Authority (ELSTAT)—and his colleagues, who continue to face charges of wrongdoing by the Greek government for revising inaccurate and misleading deficit and debt figures for the years up to and including 2009.

ISI and all of the Associations of ISI signed the statement of support which was due to be completed by December 31st. Specifically, the statement calls on Greek authorities to immediately halt legal proceedings against Georgiou and his colleagues and notes the detriment of the prosecutions on the country's economy and the field of science.

Background information regarding the case can be found here: https://www.isi-web.org/images/news/Court%20proceedings%20against%20Andreas%20%20Georgiou.pdf

An interview with Georgiou recently featured in *Significance* magazine can be found here: http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2017.01052.x/full.

Ask the Experts

"How do Establishment Surveys Differ from Household Surveys?"
Kennon R. Copeland, NORC at the University of Chicago
December 2017

Statistical and data collection methods for surveys of establishments differ in a number of key aspects from those employed for surveys of households, such as greater use of stratification, and use of multiple respondents from a sampling unit. These differences are driven by various factors related to the population heterogeneity, structure of units, existence of registers, population dynamics, availability of records, and multiple sources for information (see, e.g., Riviere). While establishment surveys have been conducted since at least the nineteenth century, comprehensive and systematic treatment of establishment survey methods outside of documentation specific to individual surveys is a relatively recent occurrence (e.g., Office of Management and Budget).

Focus on methods and applications for establishment surveys is rare in statistical and survey research conferences. As a result, a series of conferences (International Conference on Establishment Surveys (ICES) was initiated to serve the need for convening researchers to cover a broad spectrum of survey methods for business, farms, and institutions. An edited volume was published from selected papers from the first ICES, held in Buffalo, NY, in 1993 (Cox, et al), and a special issue in the Journal of Official Statistics was produced from selected papers from ICES-IV, held in Montreal, Canada in 2012 (Smith and Phipps). The fifth ICES was held in 2016 in Geneva, Switzerland, and ICES-VI is currently being planned for 2020.

POPULATION HETEROGENEITY

The population of establishments is extremely skewed, with a large number of very small units and a small number of very large units. For example, U.S. firms with fewer than 5 employees accounted for 61.75% of all U.S. firms in 2015 yet accounted for only 4.74% of total U.S. employment, while firms with 500+ employees accounted for only 0.33% of firms yet accounted for 52.50% of total employment (U.S. Census Bureau). This distribution stands in contrast to that of households, for which there is much less heterogeneity in terms of size.

This distribution means that a small portion of the population of establishments will drive values for information of interest related to size, revenue, output, etc. The result is extensive use of size stratification for establishment surveys (in addition to use of geography and establishment type) and the application of over-sampling of the largest units and under-sampling of the smallest units. In some surveys, certainty strata may be established so as to include in the sample all establishments above a specified size. Given the small influence on population totals associated with the smallest

establishments, some surveys exclude from the sample establishments below a specified size, with estimates for the excluded strata generated through the use of statistical models.

Given the influence of the largest units, nonresponse follow-up will tend to have greater variability for establishment surveys than for household surveys. Greater emphasis needs to be placed on the larger establishments to lessen the opportunity for nonresponse bias. Nonresponse adjustment may be more likely to utilize prior data for an establishment and/or use of modelled administrative data than would be the case for non-respondent households.

STRUCTURE OF UNITS

For populations of persons, units have a fairly limited hierarchy – person, household, housing unit – and the highest level is geographically compact. Establishments have much more complicated hierarchies – e.g., establishment, firm, enterprise – and at levels above establishment (single physical location, one predominant activity) both geography and type of activity may be non-compact. As a result, one first consideration in designing an establishment survey is the level (e.g., establishment, firm, enterprise) at which data are to be collected.

Classification of the unit may be problematic as well. Whereas housing units may be simply thought of as existing or not, and persons within a housing unit can be simply thought of as being determined to be in scope or not for the survey by asking a few screening questions, establishments may logically fit within multiple classifications (e.g., industry). Rules must be established for how establishments are to be classified and the information required to be collected in order to make the classification.

With establishments, it is often the case that approval from some senior officer must be obtained before survey data may be reported. There is also the consideration of coordinated responses, and the possibility that, even though information is desired at the establishment level, approval and reporting must be made at the firm level. Very loosely analogous to household surveys and the development of proxy reporting rules, data collection procedures for establishment surveys must allow for handling situations as to obtaining approval and from where data are to be obtained.

EXISTENCE OF REGISTERS

Often, there exists a register or administrative list of enterprises, firms, and establishments within a given population of interest, which contains information about each unit that is useful in stratification and sample selection (such as geography, size, and characteristics). National statistical organizations often have access to more current, comprehensive, and detailed registers than do non-governmental organizations, which must typically rely upon companies with provide business listings (e.g., Dun & Bradstreet in the U.S.). This information allows for creation of a sampling frame with the capability of carrying out detailed stratification.

For household surveys, it is generally the case that information is known only at the housing unit level with information on characteristics of the household beyond geography (e.g., income, population) being inferred through statistical modeling.

POPULATION DYNAMICS

While both household and establishment populations experience the addition and subtraction of units, the issue of “births” (new units) and “deaths” (no longer existing units) in establishment populations can present more complications for generating population estimates than is the case for household populations. For example, it is known that a substantial portion of employment growth in the U.S. results from births; however, there is a lag between their formation and their availability for inclusion in the survey. To this end, models are often generated to account for activity from births to provide more accurate population estimates. To obtain complete coverage and represent new housing units, it is not uncommon to carry out area sampling at an initial stage of selection for a household survey, whereas the use of area sampling for establishment surveys is rare.

AVAILABILITY OF RECORDS

Establishment surveys typically collect quantitative data related to performance of the population. These data may be employment, wages, sales, revenue, purchases, production, inventory, admissions, discharges, procedures, teachers, students, or the like. Summary information on the population is required to assess level and trends in performance at geographic and characteristic (e.g., industry, size, age of establishment, ownership status).

As such, the expectation is that each unit should have available records from which the information of interest can be obtained and reported. Unlike for household surveys, where respondents are often asked to answer questions from memory or based upon their opinion, attitude, perception, or behavior, respondents to establishment surveys are expected to refer to available records when reporting survey data. Often, data reporting may be customized to the data record systems available for individual sample units.

One other consideration associated with collecting records information from establishments is the cost, in terms of time, resources, and money, to the unit in obtaining and reporting the information. Given availability of administrative data on establishments, there may be more use of such to replace data collection for selected items.

MULTIPLE SOURCES FOR INFORMATION

It is not uncommon that no one respondent within an establishment can provide all the information required for the survey. This may be a case of different data being held or generated by different departments with an establishment, or of selected data being available only at some higher level in the hierarchy (e.g., only maintained by the firm as opposed to the establishment).

Again, very loosely analogous to household surveys and development of proxy reporting rules, data collection procedures for establishment surveys must recognize the possibility of the need for multiple respondents within a unit, rules for identifying and following up with the coordinating respondent, collections methods for allowing dispersion of/access to the instrument by multiple respondents.

SUMMARY

In sum, establishment surveys are primarily distinguished from household surveys through the use of size stratifications, the hierarchical and often disparate geographical and approval relationship for units, the existence of registers for creating sampling frames, the importance of "births" and "deaths" on population estimate trends, the reliance on records for reporting by units, and the possibility of multiple reporters being required to obtain all required data for a given unit.

References:

Cox, B.G., Binder, D.A., Chinnappa, B.N., Christianson, A., Colledge, M.J., and Kott, P.S. 1995. *Business Survey Methods*. John Wiley & Sons, Inc. New York, NY.

Office of Management and Budget. 1988. Quality in Establishment Surveys, Statistical Policy Working Paper 15. Springfield VA: National Technical Information Service.

Riviere, P. 2002. "What Makes Business Statistics Special?" *International Statistics Review*, 70:1, 145-159.

Sadeghi, A., Talan, D., and Clayton, R. 2016. "Establishment, firm, or enterprise: does the unit of analysis matter?" *Monthly Labor Review*, U.S. Bureau of Labor Statistics, https://doi.org/10.21916/mlr.2016.51.

Smith, P., and Phipps, P. 2014. Special Issue on Establishment Surveys, *Journal of Official Statistics*, 30:4.

U.S. Census Bureau, 2017. 2015 Statistics of U.S. Businesses. https://www.census.gov/data/tables/2015/econ/susb/2015-susb-annual.html.

New and Emerging Methods

On some reweighting schemes for nonignorable unit nonresponse

Alina Matei[1]

A particular type of nonresponse in surveys concerns situations where unit response probabilities depend on the variable of interest that may be missing: that means nonreponse is not missing at random. This type of nonresponse is specific to surveys with sensitive questions related for example to attitudes and income. From the inference point of view, the mechanism leading to this type of nonresponse is known as nonignorable. Nonignorable nonresponse means here that no model can explain the nonresponse mechanism. In this context, the construction of valid estimators is challenging. However, some solutions useful to reduce unit nonresponse bias were proposed in the literature and are discussed. Monte Carlo results are provided in order to compare them.

Keywords: unit nonresponse, not missing at random, reweighting, nonresponse bias, logistic regression, generalized calibration.

1 Introduction

Missing data arise in surveys under two forms: unit nonresponse (a sampled unit does not respond to the entire survey) and item nonresponse (a sampled unit does not respond to a particular item). Nonresponse can have negative effects on survey estimates, like the occurrence of nonresponse bias or the increase of their variance since the number of respondents diminishes.

We focus here on unit nonresponse and consider a survey for which possible respondents form a finite population with labels $U = \{1,2,\ldots,N\}$. A random sample s of units is drawn without replacement from this population in order to make inference. Some sampled units do not respond to the entire survey. Let y_i be the value of the variable of interest y measured on unit $i \in U$ and x_i a vector of covariates associated to the same unit. It is assumed that y_i is missing for nonrespondents, but x_i is known for all units selected in the sample, respondents and nonrespondents. A response indicator variable $R_i, i \in s$ is introduced, taking value 1 if unit i answers the survey, and 0 if not. It is also assumed that unit $i \in s$ answers the survey with the probability $p_i = P(R_i = 1 \mid s)$. The distribution of $R_i, i \in s$ is called the response mechanism.

[1] Institute of Statistics, University of Neuchâtel, Rue Bellevaux 51, 2000, Neuchâtel, Switzerland and Institute of Pedagogical Research and Documentation (IRDP) Neuchâtel, Switzerland, alina.matei@unine.ch

Three types of nonresponse are defined in the statistical literature:

- missing completely at random (MCAR): p_i does not depend on y_i and/or x_i;
- missing at random (MAR): p_i depends on available information (here on x_i, but not on y_i);
- not missing at random (NMAR): p_i depends on y_i or on the value of another missing variable in the survey, for any unit $i \in s$.

From the inference point of view, the mechanisms leading to MCAR and MAR are called *ignorable*, while that leading to NMAR is known as *nonignorable* (for a discussion about ignorable and nonignorable nonresponse, see Särndal and Lundström, 2005, p. 103). Lohr (1999) noted on page 339 that 'ignorable means that a model can explain the nonresponse mechanism and that the nonresponse can be ignored after the model accounts for it, not that the nonresponse can be completely ignored and complete-data methods used.' Lohr (1999), p. 351 also underlined that ignorable nonresponse is the case where 'conditionally on measured covariates, nonresponse is independent of the variable of interest'. Negating this statement, one obtains that nonignorable nonresponse is the case where even conditionally on measured covariates, nonresponse is still dependent on the variable of interest. This corresponds to the case where nonresponse is NMAR, but an usual model with covariates x cannot account for nonresponse. In the spirit of the previous definition of Lohr (1999), p. 339, nonignorable nonresponse means here that no model can explain the nonresponse mechanism and that the nonresponse cannot be ignored since no model can account for it. For such cases, some solutions to decrease nonresponse bias were proposed in the literature and are discussed in this paper. Note that some of these solutions do not imply statistical models, but some algorithmic approaches. For other methods useful to diminish the bias due to nonignorable unit nonresponse see Kim and Shao (2013).

NMAR nonresponse is typical for surveys with sensitive questions, concerning attitudes, income, drug consumption, etc. The construction of valid estimators is challenging in these surveys. However, it can concern more surveys, because, as noted by Särndal and Lundström (2005), p. 104, if p_i and x_i 'are indeed strongly related, a not uncommon situation is that' x_i and y_i 'are also related, perhaps strongly so. To imagine then that' p_i and y_i 'are unrelated is counterintuitive and unlikely to hold in a finite population.'

We focus on unit nonresponse and bias nonresponse reduction. In general, nonresponse bias is never completely eliminated, but one tries to reduce it as much as possible. In what follows, we present in Section 2 three reweighting schemes useful to reduce bias when NMAR nonresponse occurs. Section 3 shows and discuss some simulation results based on these reweighting schemes, while in Section 4 we draw our conclusions. In what follows, in the presence of nonresponse a unit is denoted by i, and by k in its absence.

2 Three reweighting schemes

Let $\pi_k = P(k \in s)$ and $r \subseteq s$ be the set of respondents to the survey. We assume that the units in r respond to the survey independently of the others. The response is treated as an additional phase of random sampling, that is known as the 'quasi randomization' approach (Oh and Scheuren, 1983). We consider that the response probabilities have the following form

$$p_i = P\left(R_i = 1 \mid y_i\right) = \frac{1}{1+\exp\left[-\left(\beta_0 + \beta_1 y_i\right)\right]}, i \in s. \quad (1)$$

Note that $P(R_i = 1 \mid y_i)$ depends on y_i that may be missing, assuming that nonresponse is not missing at random. Under the 'quasi randomization' approach, the inference is based on the sampling distribution, keeping the variable of interest and the covariates fixed, and on the modeled response distribution of R_i. We want to estimate the mean of the variable of interest y: $\overline{Y} = \sum_{i \in U} y_i / N$. Without nonresponse a widely used estimator of $\overline{Y}$ is the Hájek estimator given by $\hat{\overline{Y}} = (\sum_{k \in s} y_k / \pi_k) / (\sum_{k \in s} 1/\pi_k)$. Nonresponse may be treated by reweighting (adjustment), using the adjusted weights $1/(\pi_i p_i), i \in r$. Consequently, the following adjusted estimator for $\overline{Y}$ is derived

$$\hat{\overline{Y}}_p = \sum_{i \in r}\left(\frac{y_i}{\pi_i p_i}\right) / \sum_{i \in r}\left(\frac{1}{\pi_i p_i}\right).$$

In practice, the response probabilities p_i are never known; they are estimated. This is a challenging task since p_i depends on y_i. Once p_i estimated by $\hat{p}_i$, the following estimator is used

$$\hat{\overline{Y}}_{\hat{p}} = \sum_{i \in r}\left(\frac{y_i}{\pi_i \hat{p}_i}\right) / \sum_{i \in r}\left(\frac{1}{\pi_i \hat{p}_i}\right). \quad (2)$$

Given the assumed form of p_i in (1), we may think to fit a logistic regression model on the indicators $R_i, i \in s$: $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 y_i$. Since y_i is only observed for respondents $(R_i = 1)$, it is impossible to obtain directly estimates for β_0, β_1 and p_i using maximum likelihood (or weighted maximum likelihood), which represents the usual method to fit a such model. To overcome this problem, some methods to estimate or to approximate p_i were introduced in the literature and are presented below.

2.1 Two ways of using logistic regression

In the main form of p_i, Cassel et al. (1983) suggested to replace y_i by x_i, where x is a vector of covariates known for all units in the sample; the elements in x should be well correlated with y.

This replacement leads to the following assumed form for response probabilities

$$p_{i,x} = P(R_i = 1 \mid x_i) = \frac{1}{1+\exp\left[-\left(\alpha_0 + x_i^{\mathrm{T}}\boldsymbol{\alpha}\right)\right]}, \tag{3}$$

where α_0 and $\boldsymbol{\alpha}$ are parameters. Estimation of parameters can be done using maximum likelihood with (R_i, x_i), $i \in s$. Let $\hat{\alpha}_0$ and $\hat{\boldsymbol{\alpha}}$ be the estimates of α_0 and $\boldsymbol{\alpha}$, respectively. The estimated response probabilities are $\hat{p}_{i,x} = 1/\left\{1+\exp\left[-\left(\hat{\alpha}_0 + x_i^{\mathrm{T}}\hat{\boldsymbol{\alpha}}\right)\right]\right\}$.

A second proposition was formulated by Laaksonen and Chambers (2006). These authors proposed to impute the unknown values of y_i for $i \in s \setminus r = \bar{r}$. Let y_i^* be the imputed value of y_i for $i \in \bar{r}$ obtained by using a specific imputation method, and set $y_i^* = y_i$ for $i \in r$. The following form of the response probabilities is assumed

$$p_{i,y^*} = P(R_i = 1 \mid y_i^*) = \frac{1}{1+\exp\left[-\left(\delta_0 + \delta_1 y_i^*\right)\right]}, \tag{4}$$

where δ_0 and δ_1 are parameters. Estimation of parameters can be done using maximum likelihood with (R_i, y_i^*), $i \in s$. Let $\hat{\delta}_0$ and $\hat{\delta}_1$ be the estimates of δ_0 and δ_1, respectively. It follows that the estimated response probabilities are

$$\hat{p}_{i,y^*} = 1/\left\{1+\exp\left[-\left(\hat{\delta}_0 + \hat{\delta}_1 y_i^*\right)\right]\right\}.$$

The use of Expression (3) or of Expression (4) introduces a supplementary bias in estimation of $\overline{Y}$ since the assumed form of the response probabilities is different to the true one given in (1). Moreover, knowledge of nonrespondents' covariates is required in the first approach. For the second proposition, the imputation of missing y_i values is also challenging, since nonresponse depends on the variable to be imputed. In both methods, after the p_i estimation, the adjusted type Hájek estimator given in (2) is used. The variance of both estimators can be estimated using the reverse approach (Fay, 1991, Shao and Steel, 1999).

2.2 Generalized calibration

First, we take a look at calibration and calibration in the presence of nonresponse. In the absence of nonresponse, Deville and Sarndal (1992) proposed that:

1. A set of weights w_k is constructed by modifying the initial weights $d_k = 1/\pi_k, k \in s$ such that a distance between d_k and w_k is minimized, while satisfying the following calibration equations

$$\sum_{k\in s} w_k x_k = \sum_{k\in U} x_k, \tag{5}$$

assuming $\sum_{k\in U} x_k$ is known;

2. the final weights are $w_k = d_k F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_k) = F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_k)/\pi_k$,

where $F(.)$ is a function with suitable properties, like $F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_i) = 1 + \hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_i$ or $F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_i) = \exp(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_i)$, and $\hat{\boldsymbol{\lambda}}$ is the estimation of a vector of parameters $\boldsymbol{\lambda}$ obtained by solving the calibration equations in (5).

The calibration estimator of $\overline{Y}$ is in our case

$$\hat{\overline{Y}}_{cal} = \left(\sum_{k\in s} d_k F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_k) y_k\right) / \left(\sum_{k\in s} d_k F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_k)\right).$$

In the presence of nonresponse two different approaches are available (see e.g. Särndal and Lundström, 2005) and are reviewed below.

1. The two-step approach, where

(a) one estimates the response probabilities (using for example logistic regression)

$$p_{i,x} = P(R_i \mid x_i) = \frac{1}{1+\exp\left[-\left(\alpha_0 + x_i^{\mathrm{T}}\boldsymbol{\alpha}\right)\right]},$$

and forms the intermediate adjusted weights $d_i / \hat{p}_{i,x}$;

(b) one obtains the final weights by calibration $w_i = d_i F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_i) / \hat{p}_{i,x}$ such that the following calibration equations are satisfied:

$$\sum_{i\in r} w_i x_i = \sum_{k\in U} x_k \ \textit{or}\ \sum_{i\in r} w_i x_i = \sum_{k\in s} d_k x_k, \tag{6}$$

assuming that $\sum_{k\in U} x_k$ *or* $\sum_{k\in s} d_k x_k$ are known.

2. The one-step approach, where

(a) one obtains the final weights $w_i^c = d_i F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_i), i \in r,$ such that the calibration equations (6) are satisfied.

(b) $F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} x_i)$ is an adjustment factor that is implicitly an estimation of p_i^{-1}.

Note that the same variables x_i are used in the adjustment factor $F(.)$ and in the calibration equations. In both approaches, the calibration estimator of $\overline{Y}$ is respectively

$$\hat{\overline{Y}}_{cal,2step} = \left(\sum_{i\in r} w_i y_i\right) / \left(\sum_{i\in r} w_i\right) \textit{ and } \hat{\overline{Y}}_{cal,1step} = \left(\sum_{i\in r} w_i^c y_i\right) / \left(\sum_{i\in r} w_i^c\right).$$

Generalized calibration (Deville, 2000 ; Kott, 2006) uses the framework of the one-step approach calibration, but allows to use a set of variables z that can be different to x.

It follows that the final weights are $w_i^g = d_i F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} z_i)$, and thus $1/p_i = F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} z_i)$,

where x_i is the vector of q calibration variables, and z_i the vector of q 'model' variables. The following calibration equations are fulfilled $\sum_{i\in r} w_i^g x_i = \sum_{i\in r} d_i F(\hat{\boldsymbol{\lambda}}^{\mathrm{T}} z_i) x_i = \sum_{k\in U} x_k$ or $\sum_{i\in r} w_i^g x_i = \sum_{k\in s} d_k x_k$. The generalized calibration of $\overline{Y}$ is

$$\hat{\overline{Y}}_{gcal} = \left(\sum_{i\in r} w_i^g y_i\right) / \left(\sum_{i\in r} w_i^g\right).$$

The variance of $\hat{\overline{Y}}_{gcal}$ can be estimated using the reverse approach (Fay, 1991, Shao and Steel, 1999) or by jackknife (Kott, 2006).

Remark 1 *The values of the components of z_i should be known only for respondents. Thus, z_i can simply contain y_i; see Deville (2000) ; Kott (2006). Consequently, y is the variable of interest, but also a 'model' variable (Kott, 2006) in the 'response model' or an 'instrument' (see e.g. Särndal and Lundström, 2005; Lesage et al., 2017).*

Remark 2 *It is important to note that:*

1. *The functional form of $F(\cdot)$ in the one-step approach calibration and in generalized calibration determines the form of the response probabilities, and influences significantly the properties of the final estimator (see also Haziza and Lesage, 2016).*

2. No explicit model is specified for the treatment of nonresponse; it is given by the procedure: p_i is implicitly assumed to be $1/F(\boldsymbol{\lambda}^{\mathrm{T}} z_i)$. From this point of view, the procedure is an algorithm, and does not use a statistical model to account for nonresponse.

3. For the assumed form of p_i given in Expression (1), $F(\cdot) = 1 + \exp(\cdot)$. This functional form of $F(\cdot)$ ensures that the estimated response probabilities range between 0 and 1. In practice, the functional form of $F(\cdot)$ is a question of "guessing" since there is no way to determine its real form such that $F(\boldsymbol{\lambda}^{\mathrm{T}} z_i) = p_i^{-1}$. Haziza and Lesage (2016) showed by Monte Carlo simulation that an incorrect functional form of $F(\cdot)$ in one-step calibration may result in significant biases. For generalized calibration such results are not available, but we suspect that the same problem occurs when $F(\cdot)$ is misspecified.

3 Monte Carlo results

Monte Carlo simulation is performed assuming a model for the outcome variable y under complete response and a form for the response probability, which is allowed to depend on the outcome as in Expression (1). Since we focus here on nonresponse, a census was used, by setting $d_k = 1$, for all $k \in U$. Thus, 10,000 sets r were directly drawn from U using Poisson sampling with probabilities p_i.

Four settings with $N = 500$ and different correlation degrees between x and y were employed:

1. Setting 1: $x_i : N(2,4^2), y_i = 0.3 + 0.5x_i + \varepsilon_i, \varepsilon_i : N(0,1)$, iid, with a correlation between y and x of 0.88,
2. Setting 2: $x_i : N(2,4^2), y_i = 0.3 + 0.1x_i + \varepsilon_i, \varepsilon_i : N(0,1)$, iid, with a correlation between y and x of 0.38,
3. Setting 3: $x_i : N(2,4^2), y_i = 0.3 + 0.05x_i + \varepsilon_i, \varepsilon_i : N(0,1)$, iid, with a correlation between y and x of 0.20,
4. Setting 4: $x_i : N(2,4^2), y_i : N(0,1)$, with $i = 1,\ldots N$, with a correlation between y and x equals approximately to 0.

In each setting, $p_i = \exp(0.6 + 0.3y_i)/\left(1 + \exp(0.6 + 0.3y_i)\right)$, $i = 1,\ldots N$, $x_i = (1, x_i)^T$ and $z_i = (1, y_i)^T, i = 1,\ldots, N$. The correlation between $p = (p_i)_{(1,\ldots,N)}$ and y was about 0.9 in all settings, while the mean of p varied between 0.64 and 0.73. Figure 1 shows the scatter plots of the four settings, respectively.

For each simulation, the population mean of y was estimated through the Hájek estimator adjusted for nonresponse with different choices of $\hat{p}_i$:

- $\hat{\bar{Y}}_y$: logistic regression with y as covariate (it cannot be computed in practice),
- $\hat{\bar{Y}}_x$: logistic regression with x as covariate,
- $\hat{\bar{Y}}_{y^*}$: logistic regression with y^* (y_i^* for $i \notin r$ was imputed using the model $y_i = \gamma_0 + \gamma_1 x_i + \varepsilon_i$ with $\varepsilon_i \sim N(0,1)$ iid),
- $\hat{\bar{Y}}_{cal,1step}$: one-step calibration using x as calibration variable,
- $\hat{\bar{Y}}_{gcal}$: generalized calibration using x as calibration variable and y as 'model' variable,
- $\hat{\bar{Y}}_{unif}$: estimator using $\hat{p}_i = n_r / N$, where n_r is the number of respondents,
- $\hat{\bar{Y}}_{p_{true}}$: estimator using the true response probabilities p_i (it cannot be computed in practice).

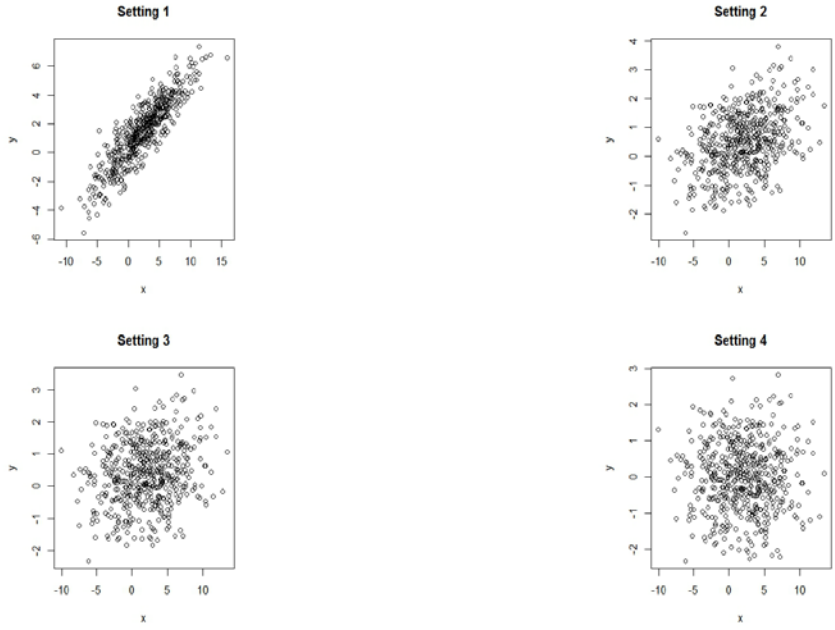


Figure 1: Scatter plots of the four settings used in Monte Carlo simulation

The function $F(x) = 1 + \exp(x)$ was used for one-step calibration and generalized calibration, ensuring that the assumed form of the response probabilities is respected. All the computations were carried out in R using the package 'sampling' (Tillé and Matei, 2016); see also the vignette called 'Calibration and generalized calibration' of this package. The functional form of $F(\cdot)$ was accommodated in the raking method of Deville and Sarndal (1992) (where the functional form of $F(x)$ is $\exp(x)$) by calibrating on $\sum_{k\in U} x_k - \sum_{i\in r} x_i d_i$ and afterwards by coming back to the original form of $1 + \exp(x)$.

For an estimator generically denoted by $\hat{\bar{Y}}$ of $\bar{Y} = \sum_{k\in U} y_k / N$, we computed the following measures:

- the Monte Carlo relative bias: $RB_{MC} = B/\bar{Y}$, where $B = E_{sim}(\hat{\bar{Y}}) - \bar{Y}$, $E_{sim}(\hat{\bar{Y}}) = \sum_{j=1}^{M} \hat{\bar{Y}}_j / M$, $\hat{\bar{Y}}_j$ is the estimate $\hat{\bar{Y}}$ obtained at the j th run, and M is the number of runs,
- the Monte Carlo variance: $Var_{MC} = \frac{1}{M-1} \sum_{j=1}^{M} \left[\hat{\bar{Y}}_j - E_{sim}\left(\hat{\bar{Y}}\right) \right]^2$,
- the square root of the Monte Carlo mean square error: $RMSE_{MC} = \sqrt{B^2 + Var_{MC}}$.

Tables 1, 2, 3, and 4 show the results corresponding to each setting, respectively. When the correlation between x and y is important, the generalized calibration provides a small relative bias and a small variance (see Table 1). For the same setting, the estimators $\hat{\bar{Y}}_x$ and $\hat{\bar{Y}}_{y^*}$ based on maximum likelihood show a larger bias and a similar estimated variance. Compared to $\hat{\bar{Y}}_x$, $\hat{\bar{Y}}_{y^*}$ performs better in terms of relative bias, because $y_i, i \in r$ is used as covariate in the logistic model and the imputation model is the true one here. For the others settings, where the correlation between x and y decreases, $\hat{\bar{Y}}_{gcal}$ provides a high variance compared to all the other estimators, while its relative bias is reduced even for a weak correlation, like in Setting 3. The estimators $\hat{\bar{Y}}_y$ and $\hat{\bar{Y}}_{p_{true}}$ perform the best in terms of bias and variance for all settings, but they cannot be compute in practice. Note that $\hat{\bar{Y}}_y$ shows a smaller Var_{MC} than $\hat{\bar{Y}}_{p_{true}}$ in all settings.

Table 1: Setting 1, correlation degree between x and y is about 0.88.

Estimator	RB_{MC}	Var_{MC}	$RMSE_{MC}$
$\hat{\bar{Y}}_y$	0.001	0.001	0.031
$\hat{\bar{Y}}_x$	0.062	0.001	0.098
$\hat{\bar{Y}}_{y^*}$	0.040	0.001	0.069
$\hat{\bar{Y}}_{cal,1step}$	0.062	0.001	0.098
$\hat{\bar{Y}}_{gcal}$	0.001	0.001	0.031
$\hat{\bar{Y}}_{unif}$	0.258	0.004	0.386
$\hat{\bar{Y}}_{p_{true}}$	0.001	0.007	0.083

Table 2: Setting 2, correlation degree between x and y is about 0.38.

Estimator	RB_{MC}	Var_{MC}	$RMSE_{MC}$
$\hat{\bar{Y}}_{y}$	<0.001	<0.001	0.005
$\hat{\bar{Y}}_{x}$	0.173	0.001	0.091
$\hat{\bar{Y}}_{y^*}$	0.161	0.001	0.085
$\hat{\bar{Y}}_{cal,1step}$	0.173	0.001	0.091
$\hat{\bar{Y}}_{gcal}$	-0.005	0.006	0.078
$\hat{\bar{Y}}_{unif}$	0.203	0.001	0.105
$\hat{\bar{Y}}_{p_{true}}$	<0.001	0.001	0.034

Table 3: Setting 3, correlation degree between x and y is about 0.20.

Estimator	RB_{MC}	Var_{MC}	$RMSE_{MC}$
$\hat{\bar{Y}}_{y}$	<0.001	<0.001	0.004
$\hat{\bar{Y}}_{x}$	0.219	0.001	0.092
$\hat{\bar{Y}}_{y^*}$	0.215	0.001	0.091
$\hat{\bar{Y}}_{cal,1step}$	0.219	0.001	0.092
$\hat{\bar{Y}}_{gcal}$	-0.012	0.023	0.153
$\hat{\bar{Y}}_{unif}$	0.227	0.001	0.096
$\hat{\bar{Y}}_{p_{true}}$	<0.001	0.001	0.032

Table 4: Setting 4, correlation degree between x and y is about 0.

Estimator	RB_{MC}	Var_{MC}	$RMSE_{MC}$
$\hat{\bar{Y}}_y$	-0.125	<0.001	0.006
$\hat{\bar{Y}}_x$	58.404	0.001	0.129
$\hat{\bar{Y}}_{y^*}$	58.402	0.001	0.129
$\hat{\bar{Y}}_{cal,1step}$	58.402	0.001	0.129
$\hat{\bar{Y}}_{gcal}$	51.510	0.203	0.464
$\hat{\bar{Y}}_{unif}$	58.335	0.001	0.129
$\hat{\bar{Y}}_{p_{true}}$	-0.009	0.001	0.034

4 Conclusions

We reviewed three reweighting schemes useful to decrease the bias for nonignorable unit nonresponse. Methods based on logistic regression (maximum likelihood or weighted maximum likelihood) allow a parametric modelisation of the response probabilities. For these models is also possible to perform goodness-of-fit tests, ensuring a control of the form of p_i. The quantities $1/\hat{p}_i$ can be very small or very large, making the corresponding estimator unstable. In practice, like for MAR, after determining $\hat{p}_i$, the sample can be partitioned in weighting classes formed according to the quantiles of the $\hat{p}_i$'s (Eltinge and Yansaneh, 1997). For a unit i inside a given class C, instead of $\hat{p}_i$ one uses the response rate observed within class C. For the MAR case, theoretical properties of the estimator using maximum likelihood with x as covariates were studied by Kim and Kim (2007). Such properties are not available for NMAR, but simulation results (see Section 3) showed that the estimated variance of the mean estimator is reduced when the estimated p_i are used instead of the true p_i (the last result was showed for MAR by Kim and Kim , 2007).

Generalized calibration is used to adjust for NMAR unit nonresponse. As the results in Section 3 show, if the goal is to estimate unknown parameters of the population, like totals or means, it can provide better results in terms of mean square error than the maximum likelihood solutions if the calibration variables are very well correlated with y; see Kott and Liao (2012), Kott (2014) and Kim and Riddles (2012).

Nevertheless, generalized calibration can increase the variance (and even the nonresponse bias) if the calibration variables are not well correlated with the model variables. For the behavior of the generalized calibration estimator, see also Lesage et al. (2017) and Osier (2013). As we have underlined, under the quasi-randomization approach, generalized calibration is an algorithmic tool useful to reduce the unit-nonresponse bias, but it does not allow to check the correctness of

the assumed form of the response probabilities. The reviewed methods present advantages, but also drawbacks, and a clear difference between them is not available in the literature.

Acknowledgements:

The author would like to thank to Professor Risto Lehtonen for his encouragement in writing this article.

References:

Cassel, C. M., Särndal, C.-E., and Wretman, J. H. (1983). Some uses of statistical models in connexion with the nonresponse problem. In Madow, W. G. and Olkin, I., editors, *Incomplete Data in Sample Surveys, vol. 3*, pages 143–160. Springer.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382.

Deville, J.-C. (2000). Generalized calibration and application to weighting for non-response. In *Compstat - Proceedings in Computational Statistics: 14th Symposium Held in Utrecht, The Netherlands*, pages 65–76, New York. Springer.

Eltinge, J. L. and Yansaneh, I. S. (1997). Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U. S. consumer expenditure survey. *Survey Methodology*, 23:33–40.

Fay, R. E. (1991). A design-based perspective on missing data variance. In *Proceedings of the 1991 Annual Research Conference*, pages 429–440. US Bureau of the Census.

Haziza, D. and Lesage, E. (2016). A discussion of weighting procedures for unit nonresponse. *Journal of Official Statistics*, 32(1):129–145.

Kim, J. K. and Kim, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canadian Journal of Statistics*, 35(4):501–514.

Kim, J. K. and Riddles, M. K. (2012). Some theory for propensity-score-adjustment estimators in survey sampling. *Survey Methodology*, 38(2):157–165.

Kim, J. K. and Shao, . (2013). *Statistical Methods for Handling Incomplete Data*. Chapman and Hall/CRC.

Kott, P. S. and Liao, D. (2012). Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine. *Survey Research Methods*, 6(2):105–111.

Kott, P. S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 32:133–142.

Kott, P. S. (2014). Calibration weighting when model and calibration variables can differ. In Mecatti, F., Conti, P. L., and Ranalli, M. G., editors, *Contributions to Sampling Statistics, Mecatti F., Conti P.L. and Ranalli M.G. Eds.*, pages 1–18. Springer.

Laaksonen, S. and Chambers, R. (2006). Survey estimation under informative nonresponse with follow-up. *Journal of Official Statistics*, 22(1):81–95.

Lesage, E., Haziza, D., and D'Hautlfoeuille, X. (2017). A cautionary tale on instrument vector calibration for the treatment of unit nonresponse in surveys. *In revision, available at http://www.crest.fr/ckfinder/userfiles/files/Pageperso/xdhaultfoeuille/DHHL.pdf.*

Oh, H. L. and Scheuren, F. (1983). Weighted adjustment for nonresponse. In Madow, W. G., Nisselson, H., and Olkin, I., editors, *Incomplete Data in Sample Survey*, volume 2, pages 143–184, New York. Academic Press.

Osier, G. (2013). Dealing with non-ignorable non-response using generalised calibration: a simulation study based on the Luxemburgish Household Budget Survey. Technical report, STATEC Luxembourg, *http://www.statistiques.public.lu/catalogue-publications/economie-statistiques/2013/65-2013.pdf.*

Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonneglible sampling fractions. *Journal of the American Statistical Association*, 94:254–265.

Tillé, Y. and Matei, A. (2016). *sampling: Survey Sampling*. R package version 2.8.

Book and Software Review

Review of *Adaptive Survey Design* by B. Schouten, A. Peytchev and J. Wagner

Brady T. West, Survey Research Center, Institute for Social Research, University of Michigan-Ann Arbor (bwest@umich.edu)

The survey research community continues to grapple with significant difficulties in producing high-quality, population-based survey estimates. Survey response rates are declining in all modes of data collection, and the costs required to recruit and measure sampled units that are representative of a target population of interest are continuing to climb. These challenges have introduced tremendous opportunities for survey methodologists to research more efficient methods for collecting survey data. Within the past decade, techniques known as *Adaptive Survey Design (ASD)* and *Responsive Survey Design (RSD)* have blossomed in response to these challenges. Survey methodologists developing these techniques have made a concerted effort to incorporate more data-driven decision-making into dynamic survey designs that aim to increase efficiency. The new Chapman & Hall / CRC book *Adaptive Survey Design* is an important attempt by three widely-renowned survey methodologists and statisticians to describe state-of-the-art approaches to implementing these types of data-driven survey designs. The three authors have been leading methodological research on these designs, and they have crafted a clear and coherent text that provides survey managers and survey statisticians alike with an important and pragmatic guide to understanding and implementing these types of designs.

The content of the book is smartly organized into five distinct sections that the authors have designed to appeal to different members of the book's rather wide target audience. Sections I, II, and III, which introduce key concepts and preparatory steps underlying ASD as well as strategies for implementing ASD in practice, have been written for survey managers tasked with guiding survey operations and implementing these types of designs. Survey statisticians would also benefit from reviewing these sections to solidify their conceptual understanding of ASD from an operational point of view. Section IV has been written with more technically-oriented survey statisticians in mind, and describes the important statistical and sampling theory underlying the tools and indicators that are essential for making these designs as efficient as possible. This section also introduces opportunities for additional statistical refinement of these approaches. Finally, Section V looks to the future, presenting important research and enhancement opportunities and communicating the thoughts of the authors with respect to what ASDs might look like further down the road as the current survey research environment continues to evolve. While one could read the book cover-to-cover to gain a full appreciation of all dimensions of ASD, this structure results in an easy reference guide for survey researchers with different backgrounds

who are interested in reviewing what is known about specific aspects of these designs.

Section I is defined by Chapter 2, and introduces the essential concepts of ASD. I particularly enjoyed the attention given by the authors to the subtle distinction between ASD, which focuses on the tailoring of data collection procedures to different strata (or subgroups) of sample units in a way that is expected to increase efficiency, and RSD, which refers to survey designs with multiple phases, where design protocols are modified for *all* sample units that are still being worked when earlier phases end (again to increase efficiency). The authors clearly describe how responsive designs can also include adaptive features (e.g., trying *different* protocols for different sampled units once an initial phase under a given protocol has no longer proven productive). In this sense, much like the Bayesian-Frequentist debate, survey researchers should not place themselves firmly in one camp or another, but rather understand the features of each approach that would be the best fit for a given survey research project.

Section II is defined by Chapters 3, 4, and 5, and focuses on important guidelines for preparing to implement ASD. Chapter 3 focuses on the key ASD feature of stratification, or identifying subgroups of the full sample that may respond well to different design features. The authors define three important goals of stratification, elaborate on practical guidance for how to form strata, and then present examples of stratification (in the ASD context) from two major national surveys. The authors also review the literature on different sources of information for forming strata, which makes for a nice contribution, and the chapter (like each of those that follow) ends with a summary of important practical points related to this topic. Chapter 4 describes interventions and modifications of design features in the ASD context, and provides a nice table (4.1) that reviews what we know from the literature on intervening in an ASD context (including what interventions have been implemented and what the effects of those interventions were). This careful attention to what we know (and what we don't know) from a relatively nascent literature is another hallmark of this book. The authors again present examples from major national surveys and make practical recommendations. Finally, Chapter 5 emphasizes the importance of developing good models of the probability of responding in the ASD context, given that predicted response propensities are important for stratifying sampled units, and describes practical approaches to fitting response propensity models with additional examples. This chapter also introduces model-based tools (such as the Fraction of Missing Information, or FMI) for monitoring the risk of nonresponse bias as a data collection proceeds and units in different strata are responding to different design protocols.

Section III is defined by Chapters 6, 7, and 8, and begins in Chapter 6 with a discussion of the importance of cost modeling in the ASD context. Costs are often either ignored or given very little weight in the literature, and this chapter brings the issue to the forefront, given that efficient ASD strives to maximize survey productivity without elevating costs. Chapter 6 also introduces important logistical considerations for implementing ASD, such as the development of technical systems and infrastructure for monitoring field outcomes in different strata. Examples from the literature of estimating costs and monitoring ASD outcomes are provided throughout. Chapter 7 turns the focus to optimizing ASD, given that there may be multiple outcomes of interest that are targeted by design modifications (e.g., cost and response rates). This chapter introduces three main approaches that might be used

to optimize these designs: statistical optimization, simulation, and trial-and-error (when optimal solutions may be unknown and different protocols need an initial evaluation). Each of these approaches is illustrated with examples from the literature.

Finally, Chapter 8 addresses the question of how robust ASDs are, given that key design parameters are often unknown or estimated with a great deal of uncertainty. This chapter introduces brand new measures that have been proposed in the literature for evaluating the robustness of ASDs, including strategies for their use and interpretation. Chapter 8 also introduces the possibility of improving ASDs that use uncertain design parameters with Bayesian methodology, which is a very important direction for future research in this area.

Section IV is defined by Chapters 9 and 10, and presents a more statistical overview of ASD, including coverage of various data quality indicators that can be used for monitoring and intervention and the use of ASD to adjust for nonresponse bias. The exposition on the various indicators in Chapter 9 will be particularly appealing to survey statisticians, as clear definitions of the various indicators (e.g., the R-Indicator) that have been proposed for tracking the risk of nonresponse bias are presented, and the strengths and weaknesses of the indicators are described from a statistical point of view. While this material is a bit more technical, I did not see it as entirely unreasonable for any researchers with graduate training in survey methodology. Chapter 10 addresses the specific question of how ASD can be used to adjust for nonresponse bias, which has been the focus of a recent literature review (Tourangeau et al. 2017). This type of bias is typically attacked during post-processing adjustments to survey weights, and Chapter 10 talks about how ASD can be used to reduce the bias during data collection. The authors present empirical examples from the literature, in additional to theoretical evidence of when bias reduction would be expected from using ASD. The authors also discuss point estimation and variance estimation approaches that should be employed when ASD is used in practice.

Section V concludes the book by looking to the future. Chapter 11 focuses on the question of whether ASD can be used to address measurement error problems, given that much of the literature has focused on the use of ASD for addressing nonresponse bias. This chapter was also a bit technical, but examples of initial work in this area are presented and discussed. Chapter 12 concludes with a broad overview of the role that ASD will play in the future of survey research, considering a broad array of designs that might benefit from this technique, and proposes several directions for future research from both statistical and operational perspectives. I found this closing chapter to be a nice summary that sets an agenda for future work in enhancing and studying these types of designs in an ever-changing survey research climate.

In sum, this is an outstanding practical handbook for any survey researchers who are (wisely) trying to incorporate data-driven decision-making into the improvement of their survey designs. The days of successfully fielding surveys under a fixed design protocol and then reviewing potential modifications after data collection has concluded are long gone, and this book synthesizes what we now know about the benefits and drawbacks of ASD.

Brady T. West

References:

Tourangeau, R., Brick, J.M., Lohr, S., and Li, J. (2017) Adaptive and responsive survey designs: a review and assessment. *Journal of the Royal Statistical Society (Series A)*, 180(1), 203-223.

Country Reports

ARGENTINA

Reporting: Veronica Beritich

A New National Victimization Survey

On the initiative of the Ministry of National Security, the National Institute of Statistics and Censuses (Instituto Nacional de Estadística y Censos- INDEC) conducted the fieldwork of the first National Victimization Survey during the months of March, April and May of 2017 with the aim of releasing information on the situation of public security in Argentina. Its geographic scope encompasses the 23 provinces and the City of Buenos Aires. It had provincial and national coverage. The target population was people aged 18 and over, residing in private homes located in towns of 5,000 and more inhabitants.

Its main objectives were to get comparable indicators at national and provincial levels on the prevalence and incidence of crime during the year 2016; to identify the security measures taken by the population to prevent crime; to collect information on the characteristics of the crime and the context of victimization; and polling the social perception of insecurity, the performance of security forces and judicial institutions, and assistance to victims.

To meet the objectives of the survey, a questionnaire was designed divided into two blocks: the household block and the individual block. The selected sample consisted of 46,765 dwellings. Every household in each of the selected dwellings had to answer the household block. Within each of these households, one of the household members aged 18 and over was selected at random to answer the individual block.

The household block had to be answered by the head of the household or, if not possible, by a household member aged 18 and over. The informant answered for all the members of the household and about the characteristics of the dwelling and the household.

The individual block inquired about the perception of security, the security measures taken by the home and the interviewee, the performance of the security forces, the occurrence of crimes against the home and against people, and the characteristics of the last victimization suffered for each type of crime.

The National Victimization Survey surveyed fourteen types of crimes, differentiated into two groups: crimes against the home and crimes against persons. Victims are persons who, individually or collectively, have suffered physical or mental harm, emotional suffering, economic loss or substantial diminution of their fundamental rights through acts or omissions that violate existing criminal laws including laws proscribing criminal misuse of power.

Prevalence is the proportion of households or persons who were victims of at least one crime during the year 2016. This indicator counts only once for each victim who suffered several types of crime or many times the same type of crime during that year.

Preliminary estimates indicate that, during 2016, 27.5% of households have been victims of a crime against the household or against any of its members, and that 67.7% of the total number of crimes surveyed has not been reported.

General information on this survey can be found at www.indec.gov.ar.
For further information, please contact ces@indec.mecon.gov.ar.

AUSTRALIA

Reporting: Rory Tarnow-Mordi and Alexander Hanysz

The Intelligent Coder: Developing a machine-learning classification system

The ABS has developed the Intelligent Coder, a text classification application suited to the needs of a National Statistical Office (NSO) in classifying short free text responses to large classification hierarchies, such as ANZSCO or ANZSIC, the Australian and New Zealand Standard Classification of Occupation and Industry Classification respectively.

A large amount of effort is expended by NSOs in developing complete hierarchical descriptions of statistical classifications of interest, like industry, occupation, education, commodity, language or country of original. However it is unreasonable to expect survey respondents to be able to volunteer their relevant code. It falls to the NSO itself to receive respondents' descriptions of their relevant characteristics and map these descriptions to the standard classification code.

The original and most widely accepted way of mapping descriptions to classifications is using clerical coding: where officers are trained to have a full understanding of a set of classification hierarchies. These officers then manually assign classification codes to responses. Clerical coding is expensive and time-consuming, so automated solutions must be pursued.

The ABS has long used an index-based coder for automated classification. This involves the creation of an index file: a set of rules that map the presence of particular words and phrases to the code that should be assigned. This is an attempt to mechanise the heuristics that a clerical coder might use to assign codes, and it succeeds in speeding up the classification process, as a large numbers of records are able to be classified very quickly.

The identification of patterns in responses and codifying these patterns is a procedure for which there are automated options; instead of manually creating an index file an automated procedure could be used which, given a set of examples, determines the optimal rules for classification. This allows the creation and refinement of index files to proceed much more quickly, as long as coded example records exist. The Intelligent Coder is this solution.

The Intelligent Coder represents text as points in vocabulary space, implements a hierarchical multi-class classification algorithm, and replicates the classification algorithm to ensure that generalisation to unseen data can be judged. Text responses are processed to a numeric vector by the bag-of-words approach, where a vocabulary of all unique words is listed from the text data available. Then an individual text record is represented as a binary vector of the same length as the vocabulary list, with 1 for each vocabulary word that is contained in the record, and 0 for words that are not. This can be thought of as representing a record as a point in vocabulary space. The

bag-of-words approach does not respect the order, the importance, or the context of words in a record, but the presence or absence of words captured by the bag-of-words approach probably captures most of the distinguishing information in records – descriptions provided by respondents tend to be semantically simple and terse.

In lieu of implementing a natural multiclass classification algorithm, the Intelligent Coder classifies records by combining a set of binary support vector machine classifiers. Specifically, a record begins at the root of the classification tree and is recursively classified to the most likely child node, where “most likely” is judged by two factors: the set of binary classifiers that combine to classify to the set of child nodes, the confidence that the coder has for that child node. The binary classifiers can be combined by creating a binary classifier for all pairs of child nodes, and a record is assigned to the child node that is assigned by the most classifiers.

The confidence that the coder has is created by bagging: resampling from the training data with replacement and creating an independent coder for each resample. These coders then vote on each record – this vote is used to evaluate the confidence of the classification.

The Intelligent Coder was trained and tested on a set of text responses to sample survey questions about occupation and industry collected between 2014 and 2016, which had been classified using an index-based coder with clerical coding for remaining records. Initial results showed that with little effort the Intelligent Coder could increase the rate of automated coding by 20% without a degradation in accuracy.

Useful references:

Clarke, F. R., and S. J. Brooker. "Use of Machine Learning for Automated Survey Coding." Proceedings of the 58th ISI World Statistics Congress. 2011.
Hastie, T., Tibshirani, R., and Friedman, J. "The Elements of Statistical Leaning: Data mining, Inference, and Prediction." Springer Series in Statistics. 2009.

For further information please contact Alexander Hanysz, alexander.hanysz@abs.gov.au

BOSNIA AND HERZEGOVINA

Reporting: Edin Šabanović

Pilot longitudinal EU-SILC 2017 finalized: Data sets sent to Eurostat

Statistics on Income and Living Conditions Survey (SILC) in European Union is produced by the survey based on a sample of households that aims to gather both household- and person-level information. This survey has rules that apply to both types of statistical units and it collects data on income and living conditions, employment, health and material deprivation of households and their members. The overall objective of the SILC survey is to collect, produce and disseminate information on income level and structure, as well as to measure poverty and living standard in the country, which are calculated according the EU methodology and regulations. Since the EU-SILC has cross-sectional and longitudinal component, both kinds of survey analysis were performed.

The pilot longitudinal ILC survey in Bosnia and Herzegovina was conducted in 2017 on the basis of the previously conducted cross-sectional pilot survey from 2015. The main objective of the second pilot survey was to test survey methodology, data collection method and field work organization in the longitudinal survey pattern.

The response rate in this pilot survey was 76.2% and data about 291 households and 872 individuals were collected by using Computer Assisted Personal Interview (CAPI) data collection method. Data entry application was made in Blaise software, while the data analysis was done in SPSS. The micro data files were transmitted to Eurostat, as well as the Quality Report. Pilot survey results will be used for the preparation of the first full-scale Survey on Income and Living Conditions in 2018.

The pilot longitudinal ILC survey in Bosnia and Herzegovina was fully funded by IPA 2014 Multi-Beneficiary Program on statistical cooperation and supported by technical assistance of Eurostat experts and GOPA consultants.

For more information, contact Edin Šabanović (edin.sabanovic@bhas.gov.ba), Sector for Statistical Methodology, Standard, Planning, Quality and Coordination, Agency for Statistics of Bosnia and Herzegovina.

CANADA

Reporting: Christiane Laperrière

Evaluating Robust Estimators for the Survey of Household Spending

The Canadian Survey of Household Spending (SHS) is an annual voluntary survey that collects detailed information on household expenditures. In 2010, the SHS has been redesigned with the aim to better adapt the collection methods and the reference periods to the capacity of the respondent to provide accurate information. The new methodology combines a questionnaire with recall periods based on the type of expenditure (1, 3 or 12 months, last payment, and four weeks) and a daily expenditure diary that the household completes for two weeks following the interview. As well, data collection is now continuous throughout the year. This methodology is similar to the one used by most household spending surveys in other countries. All expenditure variables in the interview and diary are annualized by multiplying them by a factor appropriate for the reference period. It is possible for some households to report large values for certain types of expenditures, which will have a significant impact on the estimates once annualized. Even when the large expenditure values are verified and considered valid, they can render the classical estimators quite unstable. Therefore, it may be desirable to use robust estimators instead.

An evaluation was conducted with the 2015 SHS data to define robust estimators using the concept of conditional bias, as defined in Beaumont et al. (2013). These estimators will be biased but will have smaller variances than the classical estimators in the presence of influential units (which are defined as units that have a large impact on the sampling error when they are excluded from the population). Expenditure estimates were computed using the robust estimator defined by the conditional bias theory. To do so, the theory had to be applied to the complex two-stage design of the SHS. The robust estimator can be written as a weighted sum of a modified version of the variable of interest which corresponds to the original expenditure variable that has been corrected for influential values. This correction factor is a function of the conditional bias of the unit. With SHS data, it was possible that for some influential households, the correction factor would transform the original expenditure value into a negative value. This was not deemed to be a desirable property. Indeed, for analysis purposes, users are interested in knowing when a household reports a positive amount for a specific expenditure and so it was important that the correction to influential values maintained the sign (positive or negative) of the original expenditure. Some modifications were therefore put in place to the implementation of the robust estimators to guarantee this property.

Overall, results from this evaluation were positive and the robust estimators show interesting potential for the SHS and other surveys where influential values can largely impact the estimates. Further investigation would be needed in order to fully evaluate the impact on the estimates in comparison to the usual estimates produced by the SHS, as well as the impact on the variance estimation method.

References

BEAUMONT, J.-F., HAZIZA, D., & RUIZ-GAZEN, A. (2013). A UNIFIED APPROACH TO ROBUST ESTIMATION IN FINITE POPULATION SAMPLING. *BIOMETRIKA*, 555-569.

ESTONIA

Reporting: Kristi Lehto, Kaja Sõstra, Ene-Margit Tiit

Index-based Methodology in Demographic Statistics

Concerning demographic data, the situation nowadays is somewhat controversial: the census data have lost part of their credibility due to high mobility of the population, which causes under-coverage of census. At the same time, there are different alternative sources of data about population: number of registers collecting data on different groups of population, big surveys containing rich material and also so-called big data. The problem is how to use these data sources to estimate the population size and other census variables.

Current estimation of population size

The task is to estimate currently (yearly) the number of residents in the country, where the main problem is that the external migration is not always registered. The problems of estimating population size and estimating external migration are connected: when there exists a methodology for estimating the population size for each year, it is also possible to estimate net migration including also non-registered migration. Knowledge of the exact population size is very important also in the preparation process for the register-based census.

The problem was solved in Estonia using the novel methodology of **residency index**. The prerequisite for usage of this methodology is the existence of ID-codes for all Estonian people and large number (about 20) of administrative registers where all people are identified by ID-codes. With help of these registers for all people, living in Estonia (or recently left from here) each year the signs of life were generated showing if the person had been active in the register during current year. In average, a person gets 3—5 Signs of life per year (e-g. visiting doctor, getting salary or social support etc). The residency index for a person j in year k is calculated by the following formula

$$I(j,k) = dI(j,k-1) + g\sum_{i=1}^{m} a_i\, L(i,j,k)$$

Where $L(i,j,k)$ equals 1 if the person j was active in register i in year k and 0 otherwise, a_i are weights, $I(j,k—1)$ is the person's index last year and d and g are non-negative parameters, $d+g$ =1. The values of parameters and weights, are estimated statistically. The value of index is truncated by 0 and 1 and also the threshold c has been estimated to make the decision:

$$I(j,k) \geq c => person\ j\ is\ resident\ at\ year\ k;$$
$$I(j,k) < c => person\ j\ is\ non-resident\ at\ year\ k;$$

The index-based methodology for estimation of population size is in Statistics Estonia in use since 2013, in 2016 it was implemented also for calculation of external migration. Also this methodology is encouraging for estimating the size of transnational community of Estonians that has been assessed very roughly since.

LITHUANIA

Reporting: Danutė Krapavickaitė

SPECIAL REPORT
Workshop on Survey Statistics Theory and Methodology

Workshop on Survey Statistics Theory and Methodology was organized by Baltic-Nordic-Ukrainian network on survey statistics in Vilnius, Lithuania, on August 21-24, 2017.

The event was dedicated to the 80th birthday of Carl-Erik Särndal. The celebrant gave a keynote lecture "Responsive – or adaptive – design for surveys with considerable nonresponse". Other keynote speakers were Alina Matei talking about non-ignorable nonresponse, calibration and nonresponse adjustment using R package sampling; Anton Grafström gave a talk "Spatially balanced sampling and its use as a variance reduction technique"; Lilli Japec presented a lecture **"**Big data in statistics production". 39 participants attended the Workshop with the presentations. The most popular topic was dealing with non-response in sample surveys.

The event was surmounted by a panel discussion session "Challenges of inference in survey statistics" concentrated on inference for fixed populations and sub-populations under a rapidly changing survey environment of official statistics. The moderator **Risto Lehtonen** motivated the session by pointing out that we are celebrating the 40th anniversary of the publication of *Foundations of Inference in Survey Sampling*, written by Carl-Erik Särndal together with Claes-Magnus Cassel and Jan Håkan Wretman. The book constitutes a landmark contribution to the theoretical principles of inference in survey sampling and was published in an era when the prevailing principles of survey inference were challenged. The message of the book is still timely and deserves attention in official statistics.

Indeed, the 70's was a period of intensive debates between the randomization (or design-based) approach and the prediction-based (or model-based) approach, initiated by Richard Royall with his 1970 Biometrika article on prediction-based sampling theory. Since Jerzy Neyman's 1934 path-breaking contribution to the representative method, randomization theory dominated survey sampling inference and, as Ken Brewer wrote in 2013, "It came as a considerable shock to the finite population sampling establishment when Royall (1970) issued his highly readable call to arms for the reinstatement of purposive sampling and prediction-based inference". Carl-Erik Särndal contributed with a number of important articles including his 1978 paper "Design-based and model-based inference in survey sampling" (with an extensive discussion section) and the 1979 paper. "Prediction theory for finite populations when model-based and design-based principles are combined" (with Cassel and Wretman), both published in the Scandinavian Journal of Statistics. In a joint 1983 article "Six approaches to enumerative survey sampling", Särndal and Brewer considered the role of model as the most important distinguishing factor between the inferential approaches considered and concluded by saying: "...we see

merits in both model-based and randomization-based inferences, and look for fruitful interactions between them".

In the 70's and 80's, there thus were serious attempts to derive principles for a combined model-based and randomization-based inference. Parallel to these efforts, new methodological developments were introduced under the randomization approach. In a series of papers by Carl-Erik Särndal and his colleagues, a basis was created for the famous Springer 1992 book by Särndal, Swensson and Wretman on design-based model-assisted survey sampling. By presenting a sound and applicable apparatus for generalized regression estimation and calibration, the book became soon - and it still is - a standard source in official statistics methodology, all over the world.

Today, new data sources and computational approaches emerge such as big data, data science and machine learning. The increasing versatility of the data and methodology landscape calls for clear understanding within official statistics of the inferential basis underlying the various approaches and methods. Therefore, I believe, it is important to introduce topics on inference in NSI staff training programs and university courses on modern survey statistics. In addition to the design-based approach, model-based and Bayesian approaches would be useful to be discussed.

Carl-Erik Särndal extended the discussion making three remarks.

1. Debates in the 1970's and their aftermath. Risto Lehtonen recalled "the period of intensive debates" of the 1970's, in which I took some part as a young statistician. Arguments "for or against" were sometimes harsh, almost as if the opposing parties were claiming one theory "true" and the other "false". Such sharp distinction may be possible in a science such as physics, but not in the social science context catered to by survey statistics for human or business populations. At best, there is firmly established practice, more or less well supported by a certain theoretical base.

In 1970, the randomization theory school could look back at a spectacularly victorious period. The challenge posed by the model-based views of Ken Brewer and Richard Royall and his collaborators came as a shock to "the finite population sampling establishment" at the time. That Royall's papers were not, like many honest efforts, forgotten and neglected as "exotic radical ideas", this may seem surprising. But these ideas had substance; Brewer and Royall inspired many, myself included.

Nevertheless, Morris Hansen was a forceful defender of the randomization theory school. In a famous co-authored article – Journal of the American Statistical Association 1983 - he intervened strongly against the model based mode of inference. My impression is that he was genuinely convinced that official statistics derived from purely model-based inference risked to give poor accuracy and be misleading to users of official statistics.

In the wake of the 1970's debates, randomization inference came to be called "design based inference", in contrast to the competing "model based inference". It was natural that a medium ground would appeal to those many "classical" survey statisticians who could not help but feeling an implicit presence of models - relationships between survey variable and auxiliary variables – in classical methods such as the ratio estimator. Hence came the term "model assisted" - and design based - inference, as

formally presented and elaborated in the 1992 book that Bengt Swensson, Jan Wretman and I had taken about nine years to write.

It is fair to say that from that period on, “design based estimation” was, predominantly, “model assisted design based estimation”. The generality of the concept allowed the assisting models used to produce the predicted values to be increasingly more sophisticated, and to profit from advances in general statistical theory, such as nonparametric regression, linear mixed models and several others. What is the best way to intertwine the fundamental ideas “randomization” and “modelling” in one recipe for inference, if we say that a completely “pure” variety of either kind is not realistic? The question has fascinated survey statisticians. Brewer, in later years, called his brand Combined inference. Another that has been proposed is Design assisted model based inference.

2. Later demands on inference for Survey Statistics. The 1970’s debates on inference in surveys had a heavy impact on the later developments. As society evolves, the field adapts to the new conditions. By way of example, modern society demands statistics for various not-so-large subpopulations, administrative areas of a country for example. This need lay behind the development of Small area statistics as a field with extensive reliance on model based thinking. Meanwhile, modern society has brought drastic decline of response rates for probability samples. The selected persons exercise their right to refuse participation or to have life styles that make them hard or impossible to contact. This lay behind the surge of nonresponse adjustment methods. Subfields such as Small area statistics and Nonresponse treatment remain in focus at conferences and in specialized treatises, books and/or extensive reviews. They have “models” as a more or less prominent undercurrent.

The design based inference is ideal when its (quite demanding) conditions are met, but it cannot cope well with very small number of observations in a subpopulation of interest, or with large percentage of data missing from the drawn probability sample. Thus, small area statistics and coping with nonresponse have in common that they expose, regretfully some would say, the vulnerability of the design based paradigm. These fields of research adapt by turning to a more or less extensive use of model assumptions.

What about “intensive debate” in the recent decades? Should methods with increased presence of models be denounced because they fail to live up to the ideal randomization theory norm? The profession has responded: certainly not. It would be hard for survey statisticians to refuse to comply with pressure from superiors to supply numbers also for small parts of the finite population. It is equally hard to insist that statistics from probability samples affected by high rates of nonresponse must not be published at all, for risk of very large bias. The show must go on.

3. The field of Survey Statistics in the future – where will the action be? Since the 1970’s, the Survey Statistics literature – at least its theoretical parts – has focused on the estimation phase of a survey: clever estimation theory. What will be the preoccupying issues in the future when theoreticians engage in debate, while managers in the National Statistical Institutes try to resolve “the best way to produce official statistics” on time and on budget?

Survey Statistics theory and practice will foreseeably turn more to the nature of the data collection underlying the estimation. I mean, instead of probability sampling from high quality survey frames, there is the prospect – or rather the threat - of data collections seen as practical, rapid and cost efficient from more or less carefully assembled (web)panels, or from the fuzzy Big Data cloud. With such sources of input for the estimation, satisfactory theory of quite different inclination and orientation is needed and may be forthcoming. To develop it is a responsibility for survey statisticians.

Thomas Laitila considers that randomization theory is spotless in the sense statistical inference can be drawn without assumptions. This is also why it is suitable for use in production of official statistics.

However, the applicability of the theory is restricted by e.g., nonresponse, which in recent years has reached levels far beyond what can be accepted. Arguments are raised for using alternative data sources and new methods involving applications of e.g., frequentist model-based and Bayesian inference. These are indeed well founded and fine theories with justifying areas of application, areas where the randomization theory is less suitable. The central issue to consider is how statistics are to be interpreted. The sources of randomness behind the theories are not the same and statistics are thereby to be interpreted differently. For example, frequentists provide confidence intervals while Bayesians provide credible intervals, and they are not to be given the same meaning. Another issue is model assumptions which are used to a larger extent within model-based and Bayesian inference. Issues like these must be considered before alternative inference theories are implemented in production of official statistics.

Alternative data sources and new methods will be employed in official statistics production within new survey designs. Instead of providing the main data source upon which inference is made, survey sampling can be used in special surveys used for validating/correcting the information from alternative data sources. Also, new data collection techniques in the future may in some cases relieve surveys from the nonresponse problem. So, the randomization theory will be needed in future statistics production, and new survey designs will in some cases call for extensions of the randomization theory.

Imbi Traat spoke about likelihood-based inference in sample surveys. Maximum likelihood estimation methods are attractive whenever models with unknown parameters are involved. This important technique is often used in classical statistics. Its development to sample survey data in finite populations is non-trivial both conceptually and technically. However, step by step, researchers have adjusted this inferential method for sample surveys. Likelihood-based approach grounds the inference on the probabilistic nature of sample data. Danny Pfeffermann and his co-authors have developed the concept of informative sampling and have grounded the inference on the conditional distribution of the study variable, given sampled. Due to selection mechanism this distribution is different from the population distribution. Recent advances on maximum likelihood estimation for sample surveys are presented in the book by Raymond L. Chambers, et al. (2012). They consider survey data consisting of study and auxiliary variables, inclusion and response indicators. They construct the likelihood that accounts for all random processes generating these data and use it for making inference in sample surveys.

The discussion arose much thoughts for participants about the future of the Survey Statistics.

The organizers of the Workshop are grateful to the International Association of Survey Statisticians and the Nordplus Higher Education program of the Nordic Council of Ministers for financial support, which made this event possible. The next Workshop will be held in Latvia, Jelgava, on August 21-24, 2018.

For more information please visit the website of Workshop on Survey Statistics Theory and Methodology http://vilniusworkshop2017.vgtu.lt/, or contact Danutė Krapavickaitė at danute.krapavickaite@vgtu.lt.

References:

Brewer, K. R. W. & Särndal, C.-E. (1983). Six approaches to enumerative survey sampling. In *Incomplete Data in Survey Sampling 3*, Session VIII, 363-368, New York: Academic Press.

Cassel C.-M., Särndal C.-E., Wretman J. H. (1977) *Foundations of inference in survey sampling*. John Wiley & Sons, Inc., New York.
Cassel C.-M., Särndal C.-E., Wretman J. H. (1979) Prediction Theory for Finite Populations When Model-Based and Design-Based Principles Are Combined: With an Application to a Study of Choice of Transportation Mode across the Öresund Straits. *Scandinavian Journal of Statistics* Vol. 6, No. 3, pp.97-106.

Chambers R. L., Steel D. G., Wang S., Welsh A. (2012). *Maximum Likelihood Estimation for Sample Surveys*. Chapman and Hall/CRC.

Royall, *R. M.* (*1970*). On *finite population sampling theory* under certain linear regression models. Biometrika, 57 , 377-367.

Särndal C.-E., Thompsen I., Hoem J. M., Lindley D. V., Barndorff-Nielsen O., Dalenius T. (1978). Design-Based and Model-Based Inference in Survey Sampling (with Discussion and Reply). *Scandinavian Journal of Statistics.* Vol. 5, No. 1, pp. 27-52.

MALAYSIA

Reporting: Riyanti Saari for Sri Dr. Mohd Uzir Mahidin

Innovation Award of Prime Minister's Department (AIJPM)

Innovation Award of Prime Minister's Department (AIJPM) was introduced in 2009 to inculcate a culture of innovation and creativity among public servant. This is in line with the sixth strategic thrust in Eleventh Malaysia Plan (RMKe-11): Innovation and Productivity. AIJPM enables the Government to recognise the innovation undertaken by the agencies and departments in Prime Minister's Department (JPM) which give high impacts and improve the quality of products, services and work processes. The Department of Statistics Malaysia (DOSM) participates in AIJPM every year. In 2017, DOSM sent four innovation projects for AIJPM 2017 such as:

- **Statistics Data Warehouse (StatsDW)**

The StatsDW provides information in a dynamic and interactive form and accessible to all users for free through the DOSM's portal (eDatabank, Visualisation & Time Series and Location Intelligence).

- **Interactive Distributive Trade (i-DT)**

i-DT is an interactive portal for disseminate distributive trade data in Malaysia which provides interactive facilities that will help users to get live distribution data.

- **Malaysia Social Statistics System (MySocialStats)**

This system is developed to facilitate users to update, validate and verify data and online data approval.

- **BCI (Building Material Cost Index) e-Data Collection**

This system is developed particularly for DOSM's state offices to obtain building material cost data by monthly from each selected outlet.

For AIJPM 2017, **StatsDW project** won the first place out of 30 projects under the ICT Category. This achivement generates the new momentum in producing high quality innovations and creativity in DOSM.

NEW ZEALAND

Reporting: Scott Ussher and Sean Broughton

Overview of New Zealand's Redeveloped Labour Force Survey

In 2016, Stats NZ redeveloped the Household Labour Force Survey (HLFS) to improve the relevance and quality of New Zealand's labour market data. Additionally, these improvements better align the HLFS with current best practice and with standards set by the International Labour Organization.

The updated HLFS includes more information about the nature of people's employment conditions and their work arrangements. This data can be used to better understand different patterns of employment.

New content includes:

- employees' employment relationships (eg., permanent or temporary work arrangement)
- length of time employed in current job
- employment agreement type (eg., collective or individual)
- union membership
- preference for different types of employment (eg., temporary employment or self-employment)
- improved identification of self-employed people.

An important change is the publication of the underutilisation rate as a key statistic, using data from the updated HLFS.

New topics added as supplements/modules

The redeveloped HLFS makes it easier to include extra topics as additional supplements or modules. In the June 2017 quarter, the HLFS began including data on the disability status of respondents. Disabled people twice as likely to be unemployed and subsequent news stories present, for the first time, a set of labour market statistics broken down by disability status. Stats NZ will continue to derive labour market outcomes for disabled people and non-disabled people in the June quarter each year, using the Washington Group Short Set questions.

The content of the New Zealand Income Survey was integrated into the HLFS as an annual (June quarter) module.

In the September 2017 quarter, the HLFS included the module *Childcare in New Zealand 2017*. This was New Zealand's third major survey of childcare. Previous surveys were carried out in 1998 and 2009.

Childcare in New Zealand 2017 collected information about the use of early childhood education (ECE), out-of-school services (OSS), informal care arrangements (eg., grandparents looking after children), subsidy use, and the relationship between using

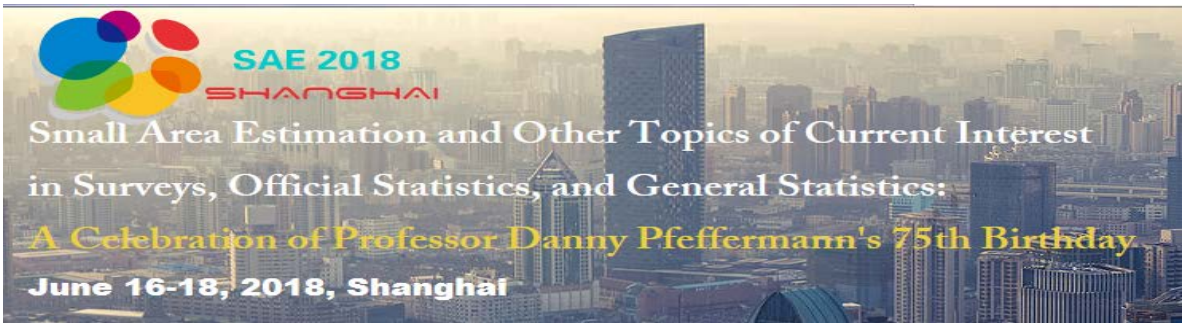
ECE and/or OSS, and work and/or study arrangements. Results from the survey will be released in the coming months.

Further information

See Stats NZ's website for further information on New Zealand's HLFS or the household surveys programme 2016–20, or contact Scott.Ussher@stats.govt.nz or Sean.Broughton@stats.govt.nz.

Upcoming Conferences and Workshops

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



Organized by: Small Area Estimation SAE 2018
When: June 16 – 18, 2018
Where: Shanghai, China
Homepage: https://www.sae2018.com/

Q2018 is the 9th European Conference on Quality in Official Statistics
Organized by: European Conference on Quality in Official Statistics
Where: Krakow, Poland
When: June 26-29, 2018
Homepage: https://ec.europa.eu/eurostat/cros/Q2018_en

Second International Conference on the Methodology of Longitudinal Surveys
Organized by: Understanding Society
Where: Essex, United Kingdom
When: July 25 - 27, 2018
Homepage: https://www.understandingsociety.ac.uk/mols2

The Joint Statistical Meetings (JSM) 2018
Organized by: Lead With Statistics Vancouver Convention Centre
Where: Vancouver, British Columbia, Canada
When: July 28 - August 2, 2018
Homepage: http://ww2.amstat.org/meetings/jsm/2018/conferenceinfo.cfm

BALTIC-NORDIC-UKRAINIAN
NETWORK ON SURVEY STATISTICS

Baltic-Nordic-Ukrainian Network Workshop on Survey Statistics 2018
Organized by: Baltic-Nordic-Ukrainian Network
Where: Jelgava, Latvia
When: August 21-24, 2018
Homepage: https://wiki.helsinki.fi/display/BNU/Events

Royal Statistical Society 2018 International Conference
Organized by: The Royal Statistical Society
Where: Cardiff, Wales
When: September 3-6, 2018
Homepage:
http://www.rss.org.uk/RSS/Events/RSS_Conference/2018_International_Conference/RSS/Events/Conference/RSS_2018_International_Conference.aspx?hkey=a556f955-22c6-4502-8956-5994c7f7c047

IAOS-OECD Conference 2018
Organized by: The International Association for Official Statistics (IAOS) and the Organisation for Economic Co-operation and Development (OECD)
When: September 19 – 21, 2018
Where: Paris, France
Homepage: http://www.oecd.org/iaos2018/

Organized By: 10ème Colloque francophone sur les sondages
Where: l'Université de Lyon
When: 24 au 26 octobre 2018
Homepage: http://sondages2018.sfds.asso.fr./

BigSurv 2018
Organized by: International Conference on Big Data Meets Survey Science
When: October 25-27, 2018
Where: Barcelona, Spain
Homepage: https://www.bigsurv18.org/

62nd ISI World Statistics Congress 2019 (ISI WSC 2019)
Organized by: International Statistical Institute (ISI) 62nd ISI World Statistics Congress
When: August 18-23, 2019
Where: Kuala Lumpur, Malaysia
Homepage: http://www.isi2019.org/

In Other Journals

Journal of Survey Statistics and Methodology

Volume 5, Issue 4 (December 2017)
https://academic.oup.com/jssam/issue

Survey Statistics

Adjusting Measurement Bias in Sequential Mixed-Mode Surveys Using Re-Interview Data *Thomas Klausch; Barry Schouten; Bart Buelens; Jan Van Den Brakel*

Dependent Latent Effects Modeling for Survey Estimation with Application to the Current Employment Statistics Survey
Julie Gershunskaya; Terrance D. Savitsky

Survey Methodology

Following Up with Nonrespondents via Mode Switch and Shortened Questionnaire in an Economic Survey: Evaluating Nonresponse Bias, Measurement Error Bias, and Total Bias
Joseph W Sakshaug; Stephanie Eckman

Assessing the Effect of Social Integration on Unit Nonresponse in Household Surveys
Ashley Amaya; Jeffrey R. Harring

Design and Results for a Survey of Nonrespondents in a Longitudinal Cohort of Young Adults
Patrick M. Marek; Arthur V. Peterson, Jr.; Maxine Henning

Does Granting Linkage Consent in the Beginning of the Questionnaire Affect Data Quality?
Stephanie Eckman; Georg-Christoph Haas

Survey Methodology

Statistics Canada Statistique Canada

Canada

Volume 43, Number 2 (December 2017)
http://www.statcan.gc.ca/pub/12-001-x/12-001-x2017001-eng.htm

Journal of Official Statistics

Volume 33, Issue 4 (December 2017)
https://www.degruyter.com/view/j/jos.2017.33.issue-4/issue-files/jos.2017.33.issue-4.xml

Survey Practice

Volume 10, Number 4 (2017)
http://www.surveypractice.org/index.php/SurveyPractice/index

Table of Contents

Survey Research Methods SRM

Journal of the European Survey Research Association

Home About Login Register Search Current Archives Editorial Board

Volume 11, Number 4 (2017)
https://ojs.ub.uni-konstanz.de/srm/index

Table of Contents

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

Volume 33, Issue 4 (2017)
https://content.iospress.com/journals/statistical-journal-of-the-iaos/33/4

2030 agenda for sustainable development: Implications for official statistics
Pfeiffer, Annette | Middeke, Fabienne | Tambour, Martin

Building a pluralistic ecosystem of data to leave no one behind: A human rights perspective on monitoring the Sustainable Development Goals
Feiring, Birgitte | Thornberry, Francesca | Hassler, Adrian

Balancing Input-Output tables with Bayesian slave-raiding ants
Gonzales Martínez, Rolando

How have National Statistical Institutes improved quality in the last 25 years?
Marker, David A.

The supply and use table as a useful tool for improving data quality in palestine
Khasib, Amina

Promoting statistical literacy in Abu Dhabi
Baniyas, Raba'a

Stratification of populated areas using a method of robust principal components
Bussi, Javier | Marí, Gonzalo | Méndez, Fernanda

Hybrid distance functions for K-Means clustering algorithms
Dalatu, Paul Inuwa | Fitrianto, Anwar | Mustapha, Aida

Using big data from Wikipedia page views for official tourism statistics
Donovan, Conor | Flaherty, Eóin T. | Healy, Eimear Quinn

International Statistical Review
Revue Internationale de Statistique

Volume 85, Issue 3 (December 2017)
http://onlinelibrary.wiley.com/doi/10.1111/insr.v85.3/issuetoc

Interview with Nancy Reid
Ana-Maria Staicu

A Tutorial on Palm Distributions for Spatial Point Processes
Jean-François Coeurjolly, Jesper Møller and Rasmus Waagepetersen

Variance Estimation of Imputed Estimators of Change for Repeated Rotating Surveys
Yves G. Berger and Emilio L. Escobar

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

Volume 10, Issue 3 (December 2017)
http://www.tdp.cat/issues16/vol10n03.php

Table of Contents

A Multi-dimensional Privacy-aware Evaluation Function in Automatic Feature Selection
Yasser Jafer, Stan Matwin, Marina Sokolova

Measuring Rule Retention in Anonymized Data - When One Measure Is Not Enough
Sam Fletcher, Md Zahidul Islam

Instance-Based Learning with l-diversity
Koray Mancuhan, Chris Clifton

Volume 7 Issue 3 (2016 - 2017)
http://repository.cmu.edu/jpc/vol7/iss3/

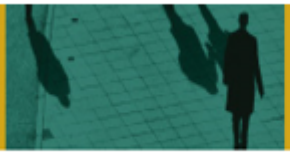
How Will Statistical Agencies Operate When All Data Are Private?
John M. Abowd

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith

On the Meaning and Limits of Empirical Differential Privacy
Anne-Sophie Charest and Yiwei Hou
Practical Data Synthesis for Large Samples
Gillian M. Raab, Beata Nowok, and Chris Dibben

A New Data Collection Technique for Preserving Privacy
Samuel S. Wu Dr, Shigang Chen, Deborah L. Burr, and Long Zhang

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS



Journal of the Royal Statistical Society
Statistics in Society
Series A

Vol 181 Volume 181 Issue 1 (January 2018)
http://onlinelibrary.wiley.com/doi/10.1111/rssa.2018.181.issue-1/issuetoc

Articles:

Forecasting daily political opinion polls using the fractionally cointegrated vector auto-regressive model
Morten Ørregaard Nielsen and Sergei S. Shibaev

Spatiotemporal trends in teen birth rates in the USA, 2003–2012
Diba Khan, Lauren M. Rossen, Brady Hamilton, Erin Dienes, Yulei He and Rong Wei

The repayment of unsecured debt by European households
Charles Grant and Mario Padula

Estimating the cumulative incidence function of dynamic treatment regimes
Idil Yavuz, Yu Chng and Abdus S. Wahed

Parsimonious higher order Markov models for rating transitions
S. Baena-Mirabete and P. Puig

Modelling illegal drug participation
Sarah Brown, Mark N. Harris, Preety Srivastava and Xiaohui Zhang

Point, interval and density forecasts of exchange rates with time varying parameter models
Angela Abbate and Massimiliano Marcellino

Can conversational interviewing improve survey response quality without increasing interviewer effects?
Brady T. West, Frederick G. Conrad, Frauke Kreuter and Felicitas Mittereder

A re-evaluation of fixed effect(s) meta-analysis
Kenneth Rice, Julian P. T. Higgins and Thomas Lumley

Data set representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice
Jamie C. Moore, Gabriele B. Durrant and Peter W. F. Smith

Separating risk from heterogeneity in education: a semiparametric approach
Jacopo Mazza and Hans van Ophem

Methods for estimating complier average causal effects for cost-effectiveness analysis
K. DiazOrdaz, A. J. Franchini and R. Grieve

Do environmental concerns affect commuting choices? : Hybrid choice modelling with household survey data
Jennifer Roberts, Gurleen Popli and Rosemary J. Harris

Journal of the American Statistical Association

ASA
AMERICAN STATISTICAL ASSOCIATION

Volume 112, Issue 519 (October 2017)
http://www.tandfonline.com/toc/uasa20/112/519?nav=tocList
Applications and Case Studies

Discussion:

Statistical Significance and the Dichotomization of Evidence
Blakeley B. McShane & David Gal

A p-Value to Die For
Donald Berry

The Substitute for p-Values
William M. Briggs

Some Natural Solutions to the p-Value Communication Problem—and Why They Won't Work
Andrew Gelman & John Carlin

Statistical Significance and the Dichotomization of Evidence: The Relevance of the ASA Statement on Statistical Significance and p-Values for Statisticians
Eric B. Laber & Kerby Shedden

Rejoinder: Statistical Significance and the Dichotomization of Evidence
Blakeley B. McShane & David Gal

Article:
Optimal Seed Deployment Under Climate Change Using Spatial Models: Application to Loblolly Pine in the Southeastern US
Alfredo Farjat, Brian J. Reich, Joseph Guinness, Ross Whetten, Steven McKeand & Fikret Isik

Mining Massive Amounts of Genomic Data: A Semiparametric Topic Modeling Approach
Ethan X. Fang, Min-Dian Li, Michael I. Jordan & Han Liu

Mortality Rate Estimation and Standardization for Public Reporting: Medicare's Hospital Compare
E. I. George, V. Ročková, P. R. Rosenbaum, V. A. Satopää & J. H. Silber

"When, Where, and How" of Efficiency Estimation: Improved Procedures for Stochastic Frontier Modeling
Mike G. Tsionas

Set-Based Tests for the Gene–Environment Interaction in Longitudinal Studies
Zihuai He, Min Zhang, Seunggeun Lee, Jennifer A. Smith, Sharon L. R. Kardia, V. Diez Roux & Bhramar Mukherjee

A Geometric Approach to Visualization of Variability in Functional Data
Weiyi Xie, Sebastian Kurtek, Karthik Bharath & Ying Sun

Finding Common Modules in a Time-Varying Network with Application to the Drosophila Melanogaster Gene Regulation Network
Jingfei Zhang & Jiguo Cao

MWPCR: Multiscale Weighted Principal Component Regression for High-Dimensional Prediction
Hongtu Zhu, Dan Shen, Xuewei Peng & Leo Yufeng Liu

Constructing Predictive Microbial Signatures at Multiple Taxonomic Levels
Tao Wang & Hongyu Zhao

Sparse Simultaneous Signal Detection for Identifying Genetically Controlled Disease Genes
Sihai Dave Zhao, T. Tony Cai, Thomas P. Cappola, Kenneth B. Margulies & Hongzhe Li

Multiscale Spatial Density Smoothing: An Application to Large-Scale Radiological Survey and Anomaly Detection
Wesley Tansey, Alex Athey, Alex Reinhart & James G. Scott

Theory and Methods:

Empirical Likelihood for Random Sets
Karun Adusumilli & Taisuke Otsu

Automatic Optimal Batch Size Selection for Recursive Estimators of Time-Average Covariance Matrix
Kin Wai Chan & Chun Yip Yau

Continuous Time Analysis of Fleeting Discrete Price Moves
Neil Shephard & Justin J. Yang

Joint Estimation of Quantile Planes Over Arbitrary Predictor Spaces
Yun Yang & Surya T. Tokdar

On Modeling and Estimation for the Relative Risk and Risk Difference
Thomas S. Richardson, James M. Robins & Linbo Wang

Parsimonious Tensor Response Regression
Lexin Li & Xin Zhang

Estimation of Monotone Treatment Effects in Network Experiments
David Choi

Generalized Scalar-on-Image Regression Models via Total Variation
Xiao Wang, Hongtu Zhu & for the Alzheimer's Disease Neuroimaging Initiative

A Functional Varying-Coefficient Single-Index Model for Functional Response Data
Jialiang Li, Chao Huang, Zhub Hongtu & for the Alzheimer's Disease Neuroimaging Initiative

Clustering Huge Number of Financial Time Series: A Panel Data Approach With High-Dimensional Predictors and Factor Structures
Tomohiro Ando & Jushan Bai

Generalized Additive Models for Gigadata: Modeling the U.K. Black Smoke Network Daily Data
Simon N. Wood, Zheyuan Li, Gavin Shaddick & Nicole H. Augustin

Robust Permutation Tests For Correlation And Regression Coefficients
Cyrus J. DiCiccio & Joseph P. Romano

Estimation in the Semiparametric Accelerated Failure Time Model With Missing Covariates: Improving Efficiency Through Augmentation
Jon Arni Steingrimsson & Robert L. Strawderman

Tukey g-and-h Random Fields
Ganggang Xu & Marc G. Genton

Semiparametric Inference in a Genetic Mixture Model
Pengfei Li, Yukun Liu & Jing Qin

Extrinsic Local Regression on Manifold-Valued Data
Lizhen Lin, Brian St. Thomas, Hongtu Zhu & David B. Dunson

Bayesian Calibration of Inexact Computer Models
Matthew Plumlee

Estimating Population Size With Link-Tracing Sampling
Kyle Vincent & Steve Thompson

An Effective Semiparametric Estimation Approach for the Sufficient Dimension Reduction Model
Ming-Yueh Huang & Chin-Tsang Chiang

On Theoretically Optimal Ranking Functions in Bipartite Ranking
Kazuki Uematsu & Yoonkyung Lee

Joint Selection in Mixed Models using Regularized PQL
Francis K. C. Hui, Samuel Müller & A. H. Welsh

Fixed-k Asymptotic Inference About Tail Properties
Ulrich K. Müller & Yulong Wang

A Group-Specific Recommender System
Xuan Bi, Annie Qu, Junhui Wang & Xiaotong Shen

Review article:

Two-Level Orthogonal Screening Designs With 24, 28, 32, and 36 Runs
Eric D. Schoen, Nha Vo-Thanh & Peter Goos

We are very pleased to welcome the following new members!

Title	First name	Surname	Department	Country
Dr.	James	Chipperfield	Australian Bureau of Statistics	Australia
Dr.	Roberto	Olinto Ramos	IBGE	Brazil
Mr.	Anil	Arora	Statistics Canada	Canada
Mr.	Jørgen	Elmeskov	Statistics Denmark	Denmark
Mrs.	Marjo	Bruun	Statistics Finland	Finland
Mr.	Dieter	Sarreither	Statistiches Bundesamt	Germany
Dr.	Faustina	Frempong-Ainguah	Regional Institute For Population Studies	Ghana
Dr.	Girish	Chandra	Division of Forestry Statistic	India
Mrs.	Sigalit	Mazeh	International Rel. & Statistical Coordination	Israel
Prof.	Giorgio	Alleva	ISTAT	Italy
Mr.	Farzan	Madadizadeh	Dept.of Epidemiology and Biostatistics	Republic of Iran, Islamic
Mr.	Hyungsoo	Park	Statistics Korea	Republic of Korea
Dr.	Pieter Cornelis J.	Everaers	European Commission - Eurostat	Luxembourg
Mr.	Ieong Meng	Chao	Direcção dos Serviços de Estatística e Censos	Macao, SAR China
Mr.	Cosme	Vodounou	AFRISTAT	Mali
Mr.	Deepuk	Bahadoor	Statistics Mauritius	Mauritius
Dr.	Julio A.	Santaella Castell	INEGI	Mexico
Dr.	Emilio	Lopez Escobar	Numérika-Medición y Análisis Estad. Avanzado, SC	Mexico
Dr.	Omar	De La Riva Torres		Mexico
Ms.	Liz	MacPherson	Statistics New Zealand	New Zealand
Ms.	Christine Benedicht	Meyer	Statistics Norway	Norway
Mrs.	Alda de Caetano	Carvalho	Inst. Nacional de Estatística (INE)	Portugal
Mr.	Stefan	Lundgren	Statistics Sweden	Sweden
Ms.	Renee	Picanso	National Agriculture Statistics Service	United States
Mr.	Charles	Rothwell	National Center of Health Statistics	United States
Dr.	Graham.	Kalton	WESTAT Inc.	United States
Dr.	Paul	Biemer	Research Triangle Institute	United States
Dr.	Mick P.	Couper	Survey Research Center	United States
Dr.	Stanislav	Kolenikov		United States

IASS Executive Committee Members

Executive Officers: (2017 – 2019)

President:	Peter Lynn (UK)	**Plynn@essex.ac.uk**
President-elect	Denise Silva (Brazil)	**Denisebritz@gmail.com**
Vice-Presidents:		
Scientific Secretary:	Risto Lehtonen (Finland)	**Risto.lehtonen@helsinki.fi**
VP Finance:	Jean Opsomer (USA)	**Jean.Opsomer@colostate.edu**
Chair of the Cochran-Hansen Prize Committee and IASS representative on the ISI Awards Committee:	Anders Holmberg, (Norway/Sweden)	**Anders.holmberg**@ssb.no
IASS representative on the 2019 World Statistics Congress Scientific Programme Committee:	Cynthia Clark (USA)	**Czfclark@cox.net**
Ex Officio Members	Ada van Krimpen	**An.vankrimpen@cbs.nl**

IASS Twitter Account @iass_isi

Institutional Members

2 International Organisations

AFRISTAT
EUROSTAT

28 Bureaus of Statistics

AUSTRALIA – AUSTRALIAN BUREAU OF STATISTICS
BRAZIL – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE)
CANADA – STATISTICS CANADA
DENMARK –STATISTICS DENMARK
FINLAND – STATISTICS FINLAND
GERMANY –STATISTICHES BUNDESAMT
GHANA – UNIVERSITY OF GHANA
IRAN – ISLAMIC REPULIC OF
ISRAEL- INTERNATIONAL REL. & STATISTICAL COORDINATION
INDIA – INDIAN COUNCIL OF FORESTRY RESEARCH AND EDUCATION PO NEW FOREST
ITALY –INSTITUTO NAZIONALE DI STATISTICSA (ISTAT)
KOREA, REPUBLIC OF – STATISTICS KOREA
LUXEMBOURG – EUROPEAN COMMISSION – EUROSTAT
MACAO, SAR China – DIREÇCAO DOS SERVIÇOS DE ESTATISTICA E CENSOS
MALI – AFRISTAT
MAURITIUS – STATISTICS MAURITIUS
MEXICO – DONATO MIRANDA FONSECA 68col. Adolfo López Mateos
MEXICO –INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (INEGI)
MEXICO – NUMÉRIKA-MEDICION Y ANALISIS ESTAD. AVANZADO, SC
NEW ZEALAND – STATISTICS NEW ZEALAND
NORWAY – STATISTICS NORWAY
PORTUGAL –INSTITUTO NACIONAL DE ESTADÍSTICA (INE)
SWEDEN – STATISTICS SWEDEN
UNITED STATES – NATIONAL AGRICULTURE STATISTICS SERVICE
UNITED STATES – NATIONAL CENTER FOR HEALTH STATISTICS
UNITED STATES – RESEARCH TRIANGLE INSTITUTE
UNITED STATES - SURVEY RESEARCH CENTER
UNITED STATES – WESTAT INC.

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

CHANGE OF ADDRESS FORM

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

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

Name: Mr./Mrs./Miss/Ms. _____ First name: _____

E-mail address *(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/