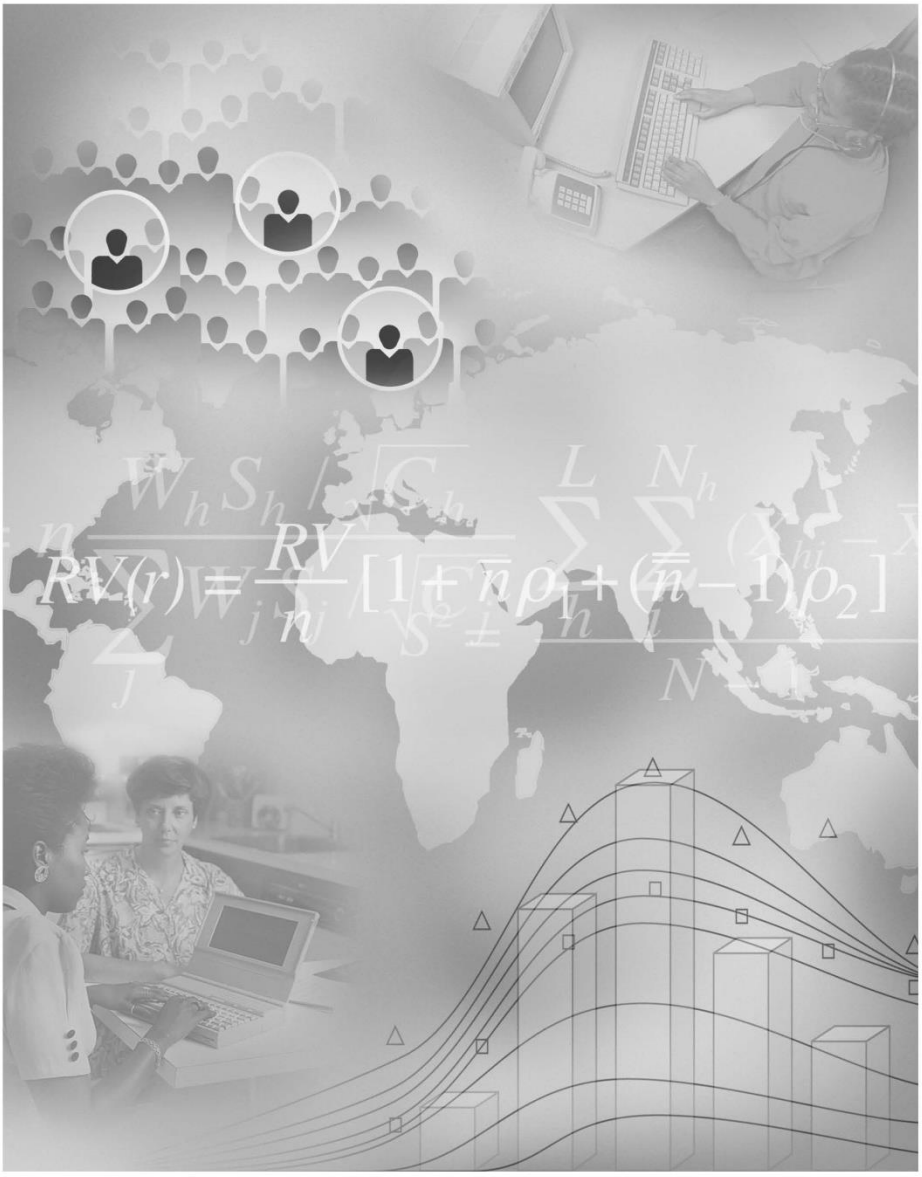
the Survey Statistician

The Newsletter of the International Association of Survey Statisticians

No. 70 **July 2014**



International Statistical Institute

Institut International de Statistique

Editors **ISSN 2521-991X**

Natalie Shlomo and Frank Yu

Section Editors

Pierre Lavallée — Country Reports
Robert Clark — Ask the Experts
Mick Couper — New and Emerging Methods
Natalie Shlomo — Book and Software Review

Circulation/Production

Olivier Dupriez
Courtney Williamson
Henry Chiem

The Survey Statistician is published twice a year in English 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-iass.org/home/services/the-survey-statistician/ .

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

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

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

In This Issue

Letter from the President

Dear Colleagues,

This is my second formal letter as President of our association, to be published *in the Survey Statistician*, and I shall devote it to telling you of what we were able to achieve during the last six months, and also what we failed (big) to achieve so far. I repeat some of the news that I told you in my interim letter because I am not sure that you all read that letter. It also makes this letter more impressive…

- The transition of IASS from a French-registered association to the new Dutch version of the IASS has been basically completed. I know that most of you don't care too much about this transition and I don't expect it to have any direct impact on our activities, but at least you won't be bothered any more with ballots and boring meetings at the ISI WSC, related to this transition. I like to thank again our outgoing Executive Director, Ms. Catherine Meunier, who is still working on the final details of the transition, mostly to do with money transfer and alike. As I wrote to you in my previous letter, Ms. Shabani Mehta from Statistics Netherlands (s.mehta@cbs.nl) serves now as the IASS liaison at the ISI and she provides big help in all aspects.

- The redesign of the new IASS website (http://isi-iass.org/home/) has been completed and the old website is no longer in use. Our big thanks are due to Olivier Dupriez (odupriez@worldbank.org) for all his sole work on this important enhancement. Please visit the website regularly, to see all the news regarding the association, and use it for advertisement of important events, job (postdoc) advertisements, on-line discussions etc. We actually haven't added yet a "job positing" page for the simple reason that no request to advertise a new job has been received so far, but I am sure that once the process starts, this corner will be very active, so don't hesitate to be the first one to use it. Another addition is a "directory of members", which contains a searchable directory of members with their email address, accessible only to IASS members. Feel free to visit the website to update your profile, add a bio etc. As I wrote in my previous letter, we plan to have the "Ask the experts" section on the web as an online continuous process, in addition to the section appearing in the Survey Statistician, which is limited to publication twice a year. Please write to the new editor of the Section, Dr. Ken Copeland, (copeland-kennon@norc.org) and send your questions, discussions, and comments. In the new website you will find also registration forms to the IASS, which you can use for recruiting new members (see below). Finally, and as I mentioned in the past, we would like to translate the important contents of the website to several languages, mostly French, but hopefully also Spanish and Chinese, thus making the site transparent to non-English speaking members. We haven't really made any significant progress on this task so far, and we are waiting for **volunteers** to help us with this important task. Please write to Olivier with other ideas that you may have regarding the new website.

- In my previous letter I mentioned our intended participation in the ISI Statistical Capacity Building effort in developing countries. I am very pleased to inform you that with the big help from the ISI and the World Bank, we were able to organize a course in Mozambique on Analysis of Complex Survey Data Using R. Our big thanks go to our members, Pedro Silva and Marcel Vieira for volunteering to teach this course (in Portuguese), which will be held during the first week of September, 2014. We are aiming to have up

to 25 participants in the course, with at least 5 participants from other Portuguese speaking African Countries. Pedro and Marcel will also teach a course in the following week at the University of Pretoria, South Africa, on Analysis of Complex Health Survey Data Using Stata, this time in English. Details of both courses can be obtained by writing to Pedro or Marcel (or to me). I am dreaming of having another course in another region before I finish my role as President next year, so please help me to accomplish my dream by volunteering to teach such a course. Just travelling to a country or two as a consultant to help the local National Statistical Institutes in their statistical activities will also be most welcome. I shall help organizing any such activity and getting the needed financial support, so don't hesitate and write to me.

- We did amazingly well in having 13 IASS invited paper sessions (IPS) accepted for the next WSC in Rio De Janeiro. Below is the list of our sessions. For details of all the IPS sessions selected for Rio, look at http://www.isi2015.org/images/isi2015/IPS2015.pdf.

1. Methodologies relating to Big Data applications. **Organizer:** Siu-Ming Tam.

2. Statistical Disclosure Control for official statistics in the 21st century. **Organizer:** Gemma van Halderen

3. Adaptive Survey Design. **Organizer:** Barry Schouten.

4. Sampling Frame and Nonsampling Error Issues in Internet Surveys. **Organiser:** Öztas Ayhan.

5. New developments in use of model-based methods in official statistics. **Organizer:** Paul Smith.

6. Small area estimation for business and economic data. **Organizer**: Susana Rubin-Bleuer.

7. What is a Census during times of changing methodologies and technologies? **Organizer**: Arona Pistiner.

8. Recent advances in empirical likelihood approaches under complex sampling. **Organizer**: Yves Berger.

9. Using remote sensing for agricultural statistics. **Organizer**: Elisabetta Carfagna.

10. Estimation and inference methods based on integrated statistical data. **Organizer**: Li-Chun Zhang.

11. Statistical implications of changing ILO international standards for employment and unemployment. **Organizer**: Tite Habiyakare.

12. Cross national comparability of national statistics. **Organizer:** Ineke Stoop (IAOS lead, and IASS co-sponsor).

13. Bayesian Analysis of complex survey data: **Organizer:** Sahar Zangeneh.

I congratulate and thank all the organizers for putting up such interesting and diverse sessions, and Christine Bycroft, our programme committee chair, for her big successful efforts in having all these sessions accepted.

I remind you that we shall also have two special sessions in Rio, with no other IPS sessions held at the same time. The first session is an "IASS President's Invited Speaker Session", with Jon Rao and Wayne Fuller presenting a joint paper on "Sample Surveys, Past, Present and Future Directions" (I guess that small area estimation will be mentioned during the talk). The second special session is a "Journal Papers Session" and I have invited the editors of the Journal of Survey Statistics and Methodology (JSSM) and Survey Methodology to select papers for this session.

- As I hope that you are all aware of, we have advertised the 2015 Cochran-Hansen prize competition for Young Survey Statisticians from Developing and Transitional Countries. A committee chaired by Risto Lehtonen (risto.lehtonen@helsinki.fi) is waiting for applications for this competition. If you haven't done so already, please encourage your colleagues/students to apply. The deadline for submission of papers is 15 February 2015. See the website for more details. We are also funding, jointly with the ISI and the World Bank travel awards for the participation of **four** young IASS members from developing and transition countries at the 2014 Statistics Canada Symposium (http://www.statcan.gc.ca/conferences/symposium2013/).The four young statisticians selected for this award are:

  Omotola Dawodu from Nigeria, Diego Andres Perez Ruiz from Mexico, Andres Gutierrez from Colombia and Sâmela Batista Arantes from Brazil. See the website for more details of the four lucky winners. I am grateful to Mick Couper and Steve Heeringa for making the selections. Special thanks are due to **Statistics Canada** for waiving the registration fees of the four participants and letting them attend one of the short courses planned as part of the symposium.

- As I have already informed you, Frank Yu retired as co-editor of "the Survey Statistician", and has been replaced by Eric Rancourt, eric.rancourt@statcan.gc.ca. Natalie Shlomo, (natalie.shlomo@manchester.ac.uk) will continue as co-editor. Also, Robert Clark retired from his role as editor of the Section "Ask the Expert" of "the Survey Statistician" and will be replaced by Ken Copeland, (copeland-kennon@norc.org). Ken will also run this section on the new website. We are all very grateful to Frank, Natalie and Robert for their long and very successful service and wish Eric and Ken our best wishes for successful terms in their respective editorial roles.

- We continue sponsoring conferences and workshop that are appropriate for our members. Following is a list of conferences that we have sponsored or agreed to sponsor so far. Our sponsorship consists of **1200-1300 Euros** for each event.

    - **III** International Workshop on Surveys for Policy Evaluation, held in November 2013 at the Federal University of Brasília (Brazil).

    - **8th** French Colloquium in Survey Sampling, to be held at the University of Burgundy, Dijon, November 18-20 2014. sondages2014.sfds.asso.fr

    - **8th** The 2014 Baltic-Nordic-Ukrainian Network on Survey Statistics Workshop, to be held in Tallinn, Estonia, August 25-28, 2014. http://www.stat.ee/workshop-on-survey-statistics-theory-and-methodology-2014

    - The 2015 European Establishment Statistics Workshop to be held in Poznan, Poland, September 2015. http://enbes.wikispaces.com/eesw15

    - **2nd** International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts, to be held in Chicago, USA, July 2016. Scientific sponsorship (no financial request at this point in time).

    - **4th** Italian Conference on Survey Methodology- ITACOSM 2015 - Roma, 24-26 June 2015.

- We shall soon have to elect a new president, vice presidents, scientific secretary and council members, to take over after the 2015 WSC. I have started the process of forming a nomination committee for this purpose.

So, where have **I** failed? In my attempts to persuade you to recruit new members to our association. I am sure that you all agree how important it is that we have a big (and hence strong) association, the reasons for which I detailed in my previous letter, which you can find on our web). Nonetheless, as you can observe in the following tables, the registration rates are even lower than last year. We had 466 members by 23/07/2013, but only 432 members by 30/06/2014, out of which only 24 members are under the age of 39 (167 members are over 60). Do you need a better proof that the future of the association is at risk? Over the last 6 months I personally approached 6 people and my success rate was 5 out of 6. So, **it really depends only on us**. Let me repeat my proposal of a "Member-brings- a Member" plan. I remind you that the membership fees are extremely low: thirty (30) Euros for members from developed countries and fifteen (15) Euros for members from developing countries. If each one of us recruits only one new member (and many of us can bring in many more new members), we achieve our goal of doubling our membership. Don't rely on other members to recruit new members. **Do it yourself.** You can find a registration form on our new website.

IASS membership by continent and year

Regions	IASS	IASS	IASS
	12/11/2013	23/07/2013	30/6/2014
Europe	206	205	189
Africa	53	49	51
Canada	22	22	20
United States	84	84	76
Central & S. America	25	26	23
Asia	54	54	51
Australia & Oceania	24	26	22
Totals	**468**	**466**	**432**

IASS membership by gender and age, 2014

Age Group	Male	Female	Total
0 - 29	2	0	2
30 - 39	13	9	22
40 - 49	29	8	37
50 - 59	37	21	58
60 - 69	72	21	93
70 or older	65	9	74
Unknown	93	53	146
Total	**311**	**121**	**432**

This is the time of big conferences and holidays, so let me finish this long letter by wishing you and your families a very pleasant and relaxing summer and winter, if you are in the southern hemisphere.

Danny Pfeffermann,
President, IASS

Letter from the Editors

The July 2014 issue of The Survey Statistician contains articles of interest and important information regarding upcoming conferences, journal contents, updates from the IASS Executive and more. We hope you enjoy this issue, and we would be happy to receive your feedback and comments on how we can make improvements.

Unfortunately, we are saying goodbye to Frank Yu who has been co-editing the newsletter since 2010. His hard work and dedication over the years have ensured a newsletter that is relevant and informative. He is responsible for developing the three methodological sections of the newsletter: the New and Emerging Methods Section, the Book/Software Review Section and the Ask the Expert Section. Frank will be replaced by Eric Rancourt and we welcome him to the editorial team. In addition, we are also saying goodbye to Robert Clark who edited the Ask the Expert Section over the past many years. He was instrumental in ensuring the most relevant topics to be addressed by experts. Robert will be replaced by Ken Copeland.

In the New and Emerging Methods Section (edited by the Scientific Secretary Mick Couper), Li-Chun Zhang from the University of Southampton and Statistics Norway has contributed an article titled 'Data Integration'. In the article, he addresses the potential impact of reusing and combining data on the official statistical system and discusses theoretical and methodological issues including types of errors. In the Ask the Experts Section (edited by Robert Clark), J. Michael Brick from Westat, Inc. answers questions related to non-probability sampling based on a review that was carried out by the Task Force for the American Association of Public Opinion Research (AAPOR) "to examine the conditions under which various survey designs that do not use probability samples might still be useful for making inferences to a larger population." For the Book and Software Review Section, David Haziza from the University of Montreal has contributed a review of the book: Statistical Methods for Handling Incomplete Data, Chapman and Hall/CRC (2013) authored by Jae Kwang Kim and Jun Shao. Please let Mick Couper (mcouper@umich.edu) know if you would like to contribute to the New and Emerging Methods Section in the future. If you have any questions which you would like to be answered by an expert, please send them to Ken Copeland (copeland-kennon@norc.org). If you are interested in writing a book or software review, please get in touch with Natalie Shlomo (natalie.shlomo@manchester.ac.uk).

The Country Report Section has always been a central feature of *The Survey Statistician* and we thank all country representatives for their contribution and coordination of the reports. We also thank the editor of the section, Pierre Lavallée (pierre.lavallee@statcan.gc.ca) for his continuing efforts to obtain timely reports from the different countries. We ask all country representatives to please share information on your country's current activities, applications, research and developments in survey methods. Please contact Geoff Lee (geoff.lee99@bigpond.com) if there is any change or addition to country representatives.

This issue of The Survey Statistician includes the letter from our IASS President, Danny Pfeffermann, containing important updates and developments, including

supported IASS conferences and the IASS invited sessions for the WSC in Rio de Janeiro 2015.

We thank Marcel Vieira for putting together the list of conferences for inclusion in the newsletter. Please send to Marcel (marcel.vieira@ice.ufjf.br) any conference announcements that you would like advertised in the next Survey Statistician to be issued in January 2015. We also thank Courtney Williamson and Henry Chiem for collating the advertisements of upcoming conferences and for preparing the tables of contents in the In Other Journals section.

As always, we have many thanks for everyone working hard to put *The Survey Statistician* together and in particular Henry Chiem and Courtney Williamson of the Australian Bureau of Statistics for their invaluable assistance.

Please take an active role in supporting the IASS newsletter by volunteering to contribute articles, book/software reviews and country reports. 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.

Frank Yu frank_yu@tpg.com.au

Eric Rancourt eric.rancourt@statcan.gc.ca

Natalie Shlomo natalie.shlomo@manchester.ac.uk

News and Announcements

Statistics Netherlands Symposium in honour of Jelke Bethlehem

Surveys: What is their future?
September 19, 2014, Leiden, Netherlands

Location:
Naturalis Biodiversity Center
Darwinweg 2, 2333CR Leiden

This year Jelke Bethlehem will retire from service at Statistics Netherlands (CBS) and will be awarded with the CBS medal of honour. To celebrate his contributions to survey methodology and survey practice, Statistics Netherlands (CBS) is organizing an international symposium. The symposium comes at a time where traditional survey research and practice are challenged by big data, survey data arising from non-probability sampling, and various new means of communication. The symposium looks ahead and gives the platform to a number of researchers from survey statistics and social sciences.

The symposium is accessible to anyone interested, but timely registration beforehand is needed at the website of the Dutch Platform for Survey Research (NPSO), www.npso.net.

Presenters:
Jelke Bethlehem is Senior Researcher at the Department of Process Development and Methodology, Statistics Netherlands (CBS) and is Research Professor by special appointment at Leiden University
Mick Couper is Research Professor at the Survey Research Center, University of Michigan, Ann Arbor, USA
Joop van Holsteyn is Professor at the Department of Political Science, Leiden University.
Bert Kroese is Director of the Division of Process Development, IT and Methodology, Statistics Netherlands (CBS)
Jim O'Reilly is Senior Blaise Consultant at Westat, USA
Carl-Erik Särndal is Professor Emeritus of Université de Montreal and affiliated to Statistics Sweden
Barry Schouten is Senior Researcher at the Department of Process Development and Methodology, Statistics Netherlands (CBS) and holds a position by special appointment at Utrecht University
Chris Skinner is Professor at the London School of Economics, UK
Ineke Stoop is Senior Researcher at the Netherlands Institute for Social Research (SCP)
Wim Vosselman is former Manager of the Department of Process Development and Methodology, Statistics Netherlands (CBS)
Kees Zeelenberg is Director of Methods and Statistical Policies, Statistics Netherlands (CBS)

Memorial Session in honour of David Binder at the Statistical Society of Canada Annual Meeting, University of Toronto, May 25-28, 2014

A special session to honour the memory of David Binder was held at the SSC Annual Meeting at the University of Toronto on May 27th, 2014. The session was organized by Mary Thompson and was well attended with many of David's friends and colleagues.

David was a long time employee of Statistics Canada and was very active in the Statistical Society of Canada, where he served as President and as Executive Director. He was also active in other statistical societies. He devoted much of his career on developing methods to make valid statistical inferences where the data is obtained from surveys with complex survey designs. David sadly passed away on June 3rd, 2012.

The invited speakers and the title of their talks were:

Abdellatif Demnati, Statistics Canada
Variance Estimation from Complex Survey Data Using Linearization Method: Impact of David Binder
Milorad Kovacevic, United Nations Development Programme Measuring Multidimensional Poverty and Inequality
Natalie Shlomo, University of Manchester and **Rodolphe Priam**
Calibration of Small Area Estimates in Business Surveys

Ask the Experts

What are the conditions under which various survey designs that do not use probability samples might still be useful for making inferences to a larger population?

J. Michael Brick

1. INTRODUCTION

Researchers routinely collect data collection using a variety of methods and techniques. For making inference that can be generalized to a finite population, probability sampling is generally accepted as the most appropriate method. With a probability sample, every unit in the population has a known, non-zero chance of being sampled, and these probabilities are the basis for the inferences. Almost all official statistics have used this methodology for many years. It is the theory that is described in most sample survey textbooks

But probability sampling is not the only method for drawing samples and making inferences. Quota samples that only require samples meet target numbers of individuals with specific characteristics such as age and sex have been used for many years, especially in market research. In the last decade or so, widespread access to the Internet has resulted in a huge shift to online surveys. Many of these surveys draw their samples from “opt-in” panels comprised of large numbers of people to have “opted in” to do surveys. These types of surveys typically are not probability samples.

The popularity of the online, opt-in surveys is largely driven by cost. The cost per completed interview is generally a much lower than it would be for a probability sample, even if the probability sample uses a lower cost method such as mail. At the same time, probability samples around the world have been suffering due to rising nonresponse and concerns about the coverage of sampling frames, especially with the rise of cell phones. There are genuine concerns about the validity of inferences from a probability sample with significant undercoverage and high nonresponse. Is it still a probability sample when the basic sampling theory requires full coverage and response?

The next section briefly summarizes a review of non-probability sampling by a task force that I co-chaired. The last section discusses some recent work that continues this discussion.

2. AAPOR TASK FORCE

Reg Baker and I chaired a Task Force for the American Association of Public Opinion Research (AAPOR) “to examine the conditions under which various survey designs that do not use probability samples might still be useful for making inferences to a

larger population." The task force completed its report in early 2013 and the full report can be downloaded www.aapor.org. The Journal of Survey Statistics and Methodology published a summary of the report, with comments from five experts in the field.
Our main conclusions are listed below. The details supporting these conclusions are omitted because of space limitations, but they are more informative than the headlines alone. I encourage you to read the report or the journal article.

i. Unlike probability sampling, there is no single framework that adequately encompasses all of non-probability sampling.
ii. Researchers and other data users may find it useful to think of the different non-probability sample approaches as falling on a continuum of expected accuracy of the estimates.
iii. Transparency is essential.
iv. Making inferences for any probability or non-probability survey requires some reliance on modeling assumptions.
v. The most promising non-probability methods for surveys are those that are based on models that attempt to deal with challenges to inference in both the sampling and estimation stages.
vi. One of the reasons model-based methods are not used more frequently in surveys may be that developing the appropriate models and testing their assumptions is difficult and time-consuming, requiring significant statistical expertise.
vii. Fit for purpose is an important concept for judging survey data quality, but its application to survey design requires further elaboration.
viii. Sampling methods used with opt-in panels have evolved significantly over time, and, as a result, research aimed at evaluating the validity of survey estimates from these sample sources should focus on sampling methods rather than the panels themselves.
ix. If non-probability samples are to gain wider acceptance among survey researchers there must be a more coherent framework and accompanying set of measures for evaluating their quality.
x. Although non-probability samples often have performed well in electoral polling, the evidence of their accuracy is less clear in other domains and in more complex surveys that measure many different phenomena.
xi. Non-probability samples may be appropriate for making statistical inferences, but the validity of the inferences rests on the appropriateness of the assumptions underlying the model and how deviations from those assumptions affect the specific estimates.

3. RECENT DEVELOPMENTS

Much of the survey research community continues to hold that only probability sampling can be used to make inferences to a population and that online non-probability sampling is not appropriate when inference is the goal (e.g., Bethlehem and Cooben, 2013). I would hope that even those who hold this opinion realize that there needs to be room for continued research into using online sample sources for all of the reasons cited above.

Second, applications continue to be published, and not all take the approach suggested by the task force to control both sampling and improve weighting. Here are two examples that differ drastically from the task force suggestion. Barratt, Ferris and Lenton (2014) explore an online sample for a rare subpopulation. They find that an online sample of this population differs in important ways from an offline probability sample of the same group. They do not describe their estimation methods,

but suggest the online sample can be useful in addition to a probability sample. Wang et al. (2014) go in the opposite direction and use a sample of Xbox users that they know is not representative of voters in the U.S. elections. They rely fully on estimation methods and show they can produce estimates with small biases despite the problems with the sample. While we might not see these as the road to the type of applications that could easily be generalized, it does show the field is dynamic and will continue to advance in many ways.

There also are efforts to examine and control the quality of online surveys. For example, draft guidelines for online sampling quality are being developed by ESOMAR, the World Association for Social, Opinion and Market Research, and the Global Research Business Network.

This is clearly an area that is undergoing explosive and unpredictable growth. By the time the ink dries on any review like this one, something new is likely to be out there. As this work continues, it is still unclear whether online sampling will gain a firmer theoretical basis and become more acceptable for official statistics.

Acknowledgements: I would like to thank Reg Baker for reviewing and making improvements in this article.

References

Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, and Roger Tourangeau. "Summary Report of the AAPOR Task Force on Non-probability Sampling." Journal of Survey Statistics and Methodology 1, (2013): 90-143.

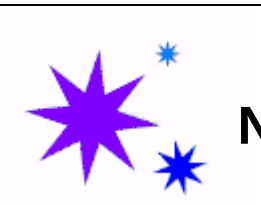
Barratt, Monica J., Jason A. Ferris, and Simon Lenton. "Hidden Populations, Online Purposive Sampling, and External Validity Taking off the Blindfold." Field Methods (2014): 1525822X14526838.

Bethlehem, Jelke, and Fannie Cooben. “Web Panels for Official Statistics?” Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong. Downloaded on May 1, 2014 from http://2013.isiproceedings.org/Files/IPS064-P1-S.pdf.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman.” Forecasting Elections with Non-Representative Polls.” (preprint of article submitted to International Journal of Forecasting). Downloaded on May 1, 2014 from http://www.stat.columbia.edu/~gelman/research/unpublished/forecasting-with-nonrepresentative-polls.pdf.

Ask the Experts - Call for Questions

If you’d like to ask the experts a question, please contact Ken Copeland at copeland-kennon@norc.org.



New and Emerging Methods

Data integration

Li-Chun Zhang
University of Southampton (l.zhang@soton.ac.uk)
Statistics Norway (lcz@ssb.no)

1 Basu's elephants and data integration

The circus owner is planning to ship his 50 elephants and so he needs a rough estimate of the total weight of the elephants. As weighing an elephant is a cumbersome process, the owner wants to estimate the total weight by weighing just one elephant. Which elephant should he weigh? So begins the famous example of Basu (1971) on the foundation of survey sampling.

To make it a good story, one needs to give some evidence that "weighing an elephant is a cumbersome process". In another, less well-known Chinese story dates to the third century AC, a wonder-boy once gave the following method: (i) take the elephant on a boat of suitable size, mark the waterline on the outside of the boat, take the elephant out of the boat, (ii) start piling stones into the boat till the mark is reached, (iii) unload and weigh the stones one by one and add up.

A *data integration* approach to the problem may be the following: (a) repeat step (i) for *all* the elephants to obtain 50 marks, (b) start piling stones into the boat till the *highest* mark is reached, (c) unload and weigh the stones one by one, note the weight difference *between* each two marks, till all the stones are unloaded (i.e. 49 + 1 values), (d) add up the weight differences appropriately to obtain the weights of *all* the 50 elephants.

Depending on how well-behaved the circus elephants are, this could take somewhat more time than weighing the largest elephant Jumbo alone, but certainly much less than repeating (i) - (iii) *separately* 50 times. Moreover, the circus owner could do well to keep the boat and the weights associated with each mark, so that some simple interpolation and extrapolation may very well save the trouble of moving and weighing any stones at all the next time around, and his boat can offer a quick solution to anyone else who needs to weigh something that is rather heavy.

The moral is to *reuse* and to *combine* data. Data integration provides an over-arching outlook to the methodological re-engineering for the official statistical system. On the one hand, this has to do with the ever-and-rapid increasing demand of statistical evidence that cannot be met by the data directly collected using designed sample surveys and censuses. On the other hand, this is driven by the need to reduce both the cost and burden associated with such purposeful, targeted data collection. A long-term perspective is necessary. In the example above, getting the weights of all the 50 elephants is actually more than what the circus owner had asked for on this

occasion. But hopefully the extra effort of constructing the boat is rewarded by future reuses.

While the general trend is clear, it should be noted that, from a scientific point of view, there will be both benefits and drawbacks of data integration compared to designed purposeful data collection. In the brief personal account that follows, I shall first give some general remarks about the outlook in Section 2, and then describe in Section 3 some prominent theoretical and methodological issues, roughly ordered by types of errors, that I have encountered in practice. Some of the statistical challenges and research opportunities will be discussed along the way.

2 Looking beyond the sampling paradigm

2.1 Total error framework

When it comes to sample surveys, there exist a number of total survey error frameworks (e.g. Biemer and Lyberg, 2003; Weisberg, 2005; Groves et al. 2009). The plurality is somewhat unavoidable given the complexity and extensive human-intervention involved in the production process, which invites different conceptualization and organization of all the potential errors. Yet no one today actually calculates and disseminates a "total error" of the final survey estimate. Quantifying the "total error" has become something of a holy grail, perpetually alluring but practically unattainable at the same time.

Nevertheless, the total survey error perspective is valuable as it enables communication, and promotes to a wholesome understanding of the statistical production chain. For the same reason, I proposed a two-phase total error framework for integrated statistical data (Zhang 2012). The first phase, adapted from the survey-error model of Groves et al. (2009) with some minor modifications, applies to the different input-source datasets each on its own. The second phase concerns the processes of integration, where many types of errors arise that are either distinct to data integration or are now much more prominent even though the error may also exist in the sample survey context. Some of these errors will be discussed in Section 3. Here I make just two quick remarks.

Big Data has emerged as a potential source for official statistics (e.g. Daas and Puts, 2014). The human, transaction or sensor generated Big Data are similar to the 'traditional' administrative data in that they are records of events/happenings, instead of responses to direct probing in sample survey or census. Transformation of information organized around the first-phase "objects" (Zhang, 2012), such as twitter messages, credit card transactions or logs of mobile network roaming, to the second-phase units, such as person, household or business is necessary as the first step of integration. The two-phase model is equally needed for the integration of Big Data.

Next, it is generally helpful to distinguish between a process as the source of an error and the process at which this error may be assessed or adjusted. For instance, a delay in the registration of an employee benefit, which occurs at the input administrative source, may potentially be evaluated and adjusted as a measurement error for activity status, using a probabilistic framework (e.g. Zhang and Fosen, 2012) at the second phase. It is of course important and necessary to improve the quality at the "source". But *statistical methods* of evaluation and adjustment are generally needed not only in the meantime but *always*, as long as the data are not error-free.

2.2 Beyond the sampling paradigm

In what may be called the *sampling paradigm* of combining survey and administrative (or census) data, the non-survey data provide mainly sampling frame and auxiliary information, for reducing both sampling and non-sampling errors. Under the data integration perspective, the role of non-survey data is considerably extended, and sample survey data is no longer a necessity.

The term *register-based* refers often to statistics tabulated on *statistical register data* processed from purely administrative data. It may be that no additional survey data are available at all, such as when health statistics are exclusively compiled based on clinical records. Or, relevant survey data may be available, but are only used 'indirectly' to define the processing rules of the administrative data, or to assess the accuracy of the statistical register but not to adjust the register-based results. For example, past census and sample survey households have been used for such purposes in the context of statistical register of households (Zhang, 2011).

It is important to notice that combining data from multiple sources is generally necessary for the register-based approach. In particular, integration with one or several base registers (Wallgren and Wallgren, 2006), including Population Register, Business Register and Immobility Register (of building, property and land), is almost always required to obtain the target population frame and to improve the data quality. For instance, part of the input to register-based education statistics are university exam results, which may be utilized for deriving, say, the variable highest level-of-education. Matching these data to the Population Register is necessary in order to verify the in-scope target population, and it helps to combine multiple exam results of the same person and check their plausibility, *etc.*, and it enables these data to be 'linked' with other relevant sources of educational data for the stated purpose.

3 Some errors of data integration

Sampling, nonresponse and measurement errors have by far received most attention under the total survey error framework. Examples of additional errors that are sometimes mentioned include coverage, specification and processing error (e.g. Biemer and Lyberg, 2003). Meanwhile, data integration brings forward some distinct errors that are either new (e.g. error of progressive data, Section 3.5) or old but require re-newed approaches (e.g. linkage error, Section 3.1).

3.1 Linkage error

In the simplest situation of two linking datasets A and B, *linkage* error occurs if (I) record *a* in A and *b* in B that correspond to two different units appear as a matched record *ab* in the linked dataset AB, or (II) record *a* in A and *b* in B that correspond to the same unit do not appear as a matched record *ab* in AB. Unless a unique and error-free identifier exists, or can be constructed in both datasets, and can be used to link them, linkage errors are by and large unavoidable.

In a seminal paper, Fellegi and Sunter (1969) outline the theory of probability record linkage (RL), which remains the foundation of the current practice (e.g. Herzog et al., 2007). Sadinle and Fienberg (2013) generalize the framework to multiple datasets. Let $\Delta = \{(a,b); a \in A, b \in B\}$ be the set of all possible matches. Let (M,U) be a bi-partition of Δ, where M contains all the true matches, and U all the false ones. Let γ_{ab} denote a score, or pattern, of comparison between *a* and *b*. Probability RL under the Fellegi-Sunter (FS) paradigm is based on the ratios

$$R_{ab} = \frac{f(\gamma_{ab}|(a,b) \in M)}{f(\gamma_{ab}|(a,b) \in U)}$$

where f is a probability density or mass function depending on the outcome space of γ_{ab}. Provided it is feasible to define the distributions $P((a,b) \in M)$ and $P((a,b) \in U)$ in a meaningful manner, R_{ab} can be directly related to the conditional probability $P[(a,b) \in M|R_{ab} = r]$.

For statistical inference based on RL, however, one needs the *joint* distribution of the linkage errors (I) and (II), and a different framework is required. Let ω be an $n_A \times n_B$ *linkage* matrix (Neter et al., 1965), where $\omega_{ab} = 1$ if $a \in A$ is matched to $b \in B$, and 0 otherwise, and n_A and n_B are the sizes of A and B, respectively. Any possible matched dataset, with the associated variables for estimation or analysis, jointly corresponds to a distinct ω. Let Ω be the outcome space of ω, i.e. with each matrix as a single element of Ω. Modelling of the distribution $P_\Omega(\omega)$ provides then the means for proper statistical inference of data linkage.

Illustration: Let $A = \{a_1,a_2\}$ and $B = \{b_1,b_2\}$. The FS paradigm can produce 4 ratios $\{R_{ab};a,b = 1,2\}$ and potentially 4 probabilities $\{P[(a,b) \in M|\gamma_{ab}];a,b = 1,2\}$. Whereas we have

$$\Omega = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right)$$

Future research needs to overcome, at least, two main difficulties. First, it is generally not enough to have *only* the matched dataset, much like it is misleading to provide only the respondent dataset in the case of nonresponse. Second, the pairwise ratios or probabilities produced under the FS paradigm are generally insufficient to identify the joint distribution $P_\Omega(\omega)$. It also remains to be seen whether the need for $P_\Omega(\omega)$ leads to fundamental changes in the practice of RL.

3.2 Coverage error

The Business Register (BR) is typically compiled based on multiple sources. The fact that a unit may appear on some but not all the input sources is an indication of potential *coverage* error, whether one decides to include that unit in the target population or not. For household population, the coverage error has attracted much attention in the censuses (e.g. Wolter, 1986).

The models for dealing with coverage errors, e.g. in the censuses, are often taken from the capture-recapture (CR) methods (e.g. Fienberg, 1972; Cormack, 1989; IWGDMF, 1995a and 1995b). Take the case of two lists A and B that both enumerate the target population U. Under a simple CR model, one may assume, among others, that the list captures have constant but different probabilities for A and B, and are independent of each other, i.e.

$$P(i \in A, i \in B|i \in U) = P(i \in A|i \in U)P(i \in B|i \in U) \quad (1)$$

The probability-product rule (1) is evidently a conditional independence assumption that corresponds to the two-way log-linear model for $i \in U$ with null first-order interactions.

Essentially the model (1) combined with post-stratification has been used for under-coverage adjustment of census enumeration (e.g. Hogan, 1993). For enumeration based on statistical registers, however, the problem of *erroneous* enumeration is often non-negligible, where an enumerated unit is said to be erroneous if it does not really belong to the target population. To take erroneous enumerations explicitly into account, one needs to consider the combined *target-list* universe, denoted by $U_+ = U \cup A \cup B$. Two assumptions that may potentially be of general use are, respectively,

$$P(i \in U | i \in A, i \in B) = P(i \in U | i \in A \setminus B)P(i \in U | i \in B \setminus A) \quad (2)$$

$$P(i \in U | i \in A, i \in B) = P(i \in U | i \in A)P(i \in U | i \in B) \quad (3)$$

(Zhang, 2014). In particular, assumption (2) corresponds approximately to a three-way log-linear model for $A \cup B$ with null second-order interactions, while assumption (3) is a member of a completely different class of models. The probability-product rule (3) is termed *pseudo conditional independence (PCI)*. Formalizing the concept of PCI in parallel to conditional independence that underpins the log-linear models for contingency tables, it is possible to develop classes of hierarchical log-linear PCI models in parallel to the standard hierarchical log-linear model.

Future research will establish whether a modelling approach to include both under- and over-coverage errors can have direct impact on the census methodology. A still more ambitious goal is to produced census-like population statistics without the census. To start with, the census enumeration may be replaced by an “improved administrative file” (i.e. register enumeration), as some countries (e.g. Israel, Switzerland) have done already. A general modelling approach also opens up the possibility for using several input registers instead of one combined register.

Applications to CR data in a range of situations can be conceived. For instance, the target population may be clandestine and dynamic, such as the active drug-users. Relevant lists may be available from the police, clinics and non-governmental organisations. Erroneous enumeration can occur in all these lists. Or, consider multiple screening procedures, each generating a list of the units with a positive test result. Only the test-positive units are subjected to a comprehensive examination, which may reveal both erroneous and under-enumerations in each list. A model for predicting the errors of each test as well as the combined test results may then be of interest.

3.3 Unit error

Due to the errors in the input source, statistical units constructed from secondary data *alone* may have non-negligible *unit* errors. For instance, register-based household statistics is produced in a number of countries in the last census round. Household unit error is the case if (i) people actually belong to the same household appear in different register-households, or (ii) people in the same register household actually belong to different households. Similarly, delineation of statistical units such as enterprise and kind-of-activity units in the BR suffers also from unit errors.

Suppose that each *target* unit can be represented as a union of smaller *base* units, which are never broken up themselves when constructing the target units. For example, persons can be base units when households are the target units, but not family because the members of a family may very well live in different households. However, family members residing at the same dwelling may in practice be treated as base units when households are the target units.

Let ω be an $n \times n$ *allocation* matrix (Zhang, 2011), for base units $j = 1,...,n$, where $\omega_{ij} = 1$ if j is belongs to the target unit $i = 1,...,m$, and m is unknown in general. The allocation matrix differs from the linkage matrix. We have $\sum_i \omega_{ij} = \sum_j \omega_{ij} = 1$ or 0 for linkage matrix ω, since a record may or may not be linked to another record, but it cannot be linked to more than one record. Whereas we have $\sum_i \omega_{ij} = 1$ and $\sum_j \omega_{ij} = 1,2,...,n$ for allocation matrix, because a base unit must belong to one and only one target unit, and a target unit must contain either one or more than one base units. Notice that the rows of an allocation matrix which have only zero's do not correspond to any target unit. Moreover, while $0 \leq \sum_{i,j} \omega_{ij} \leq \min(n_A, n_B)$ in the case of a linkage matrix, we have $\sum_{i,j} \omega_{ij} = n$ in the case of an allocation matrix.

Using the allocation matrix, it is possible to represent the household-level variables of interest as functions of the individual-level variables. Provided the distribution $P_\Omega(\omega)$, then, one can evaluate the statistical uncertainty of the household variables that is caused by the unit errors. A stratified multinomial model for the allocation matrix can be used (Zhang, 2011). For a simple illustration, let the strata be given by the number of base units within each block of allocation, the number of distinct allocation matrices within each stratum are fixed, such as

$n = 2$: $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$n = 3$: $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Under a multinomial model one simply assigns a point mass to each distinct allocation matrix.

Future research should establish similarly a unit-error theory for business units. A relevant issue that differs from household unit error is the industrial coding, such as NACE in Europe. The NACE can equally be applied to nested business units such as enterprise A and its associated local-kind-of-activity (LKA) units $B_1,...,B_K$. The NACE of the LKA units may differ from that of the enterprise A, as long as a unique code must be assigned to each business unit. As a result, a total by a given NACE code may differ depending on whether the aggregation is over the enterprises or over LKA units. It would be sensible to consider this partial identification problem (Zhang, 2012) together with the unit error. The issue does not seem to be a concern for household unit error because e.g. one does not apply the same classification to person and household.

3.4 Relevance error

Specification error refers to the difference between the quantity intended to be measured and that is obtainable by the survey instrument, e.g. the questionnaire. *Relevance* error generalises the latter to be the quantity that is obtainable via data integration, i.e. including that via the survey instrument directly. On a further subtle extension, it can refer to the difference between the quantity measured by some gold standard and what is obtainable, i.e. regardless of whether the gold standard measures exactly the theoretical quantity intended or not. For example, administrative data often entails relevance error, sometimes referred to as the definitional error, because by nature the primary data definition is meant for the

administrative purposes, which tends to differ from the survey data definition that is, or could be, implemented. We shall call the statistical variable derived from administrative sources a *proxy*, provided it has *similar definition* and the *same support* as the survey variable that is, or could have been, available. Thus, for example, the binary job-seeker status derived from administrative source is a proxy to the unemployment status collected in the Labour Force Survey (LFS). But the same job-seeker status is *not* a proxy to the LFS activity status defined as (employed, unemployed, inactive), because it has a different support. And it does not seem very meaningful to talk about relevance error without the variables' having the same support. Other typical sources of proxy variables are past censuses, or historic data of the same variable in repeated surveys.

We notice that the concept of proxy is a familiar one. For instance, indirect interviews where household members answer on behalf of absentees yields proxy to direct interview data. Or, in mixed mode collection, the data collected *via* telephone, internet, paper-questionnaire, *etc.* may be considered as proxies of each other if there are mode effects. Still another situation is micro data for public use, where the true data are available, but there may be concerns about unsecured sensitive information. Using synthetic proxy data instead of the true data may avoid disclosure risks (e.g. Rubin, 1993; Fienberg et al., 1998).

The natural approach to assess relevance error is to treat the difference between the proxy and target variables as unit-specific measurement error that does not necessarily have expectation zero. A more interesting approach is to consider the multiple proxies as manifestations of a latent construct. See e.g. Scholtus and Bakker for a structural equation modelling approach for assessing the validity of administrative and survey measures. The most general perspective is to compare the statistical distribution of the proxy and target measures (Zhang, 2012), under which the relevance error is no longer unit-specific.

Denote by Y the target variable and Z the proxy. Denote by X the additional auxiliary variables. The proxy yields *valid* inference if $f(z,x) = f(y,x)$, i.e. the joint distribution of (Z,X) is the same as that of (Y,X). Notice that, while unit-specific equality between Z and Y implies validity, validity does not require unit-specific equality. A more stringent notion is *equivalence*. The dataset $\{(z,x);s(z,x)\}$ is empirically equivalent to the target dataset $\{(y,x);s(y,x)\}$ if the empirical distribution function of $(X,Y) = (x,y)$ based on the latter is the same as that of $(X,Z) = (x,y)$ based on the former. Equivalence entails identical inference.

Table 1: Illustration of empirical equivalence

	True Value	**Proxy Value**						
Andersen	0	1	1	0	0	0	1	1
Johnson	1	0	1	0	0	1	0	1
Petersen	1	1	0	0	1	0	0	1
Number of unit-specific errors		2	2	2	1	1	3	1
Equivalence to true data		Yes	Yes	No	No	No	No	No
Relative equivalence between datasets		Yes		-	Yes			-

Illustration: Consider the binary data in Table 3.4. The three true values are listed together with all possible different sets of proxy values. For instance, the true values may be the ones collected in the survey, and the proxies processed from administrative registers. Or, the proxies may be the synthetic values released for public use. Unit by unit, none of the set of proxy values is equal to the true one. However, take e.g. the set of proxies next to the true one in Table 3.4. Because the

two have the same empirical distribution functions, they are statistically equivalent to each other, and will e.g. yield the same mean estimate and the associated variance estimate.

Future development should aim to apply the above-mentioned concepts of statistical relevance to quality assessment of statistics based on registers and data integration. The important thing is to realize that one can not remove definitional bias, but one can achieve statistical relevance.

3.5 Prediction based on progressive data

The fieldwork operation in sample surveys and censuses always has a closing date, after which the data become static and can only be altered in editing. The situation is different in most administrative data sources. Reporting and registration delays and corrections can occur a long time after the statistical reference date, whether by allowance or negligence. See e.g. Hedlin et al. (2006) for delayed introduction of birth units in the UK BR, Linkletter and Sitter (2007) for delays in Natural Gas Production reports in Texas, and Zhang and Fosen (2012) for delays in the Norwegian Employer/Employee Register. Depending on the situations, input data delays and changes may cause coverage errors or measurement errors, or both, in the integrated data.

Let t be the *reference* time point of interest and $t + d$ the *measurement* time point, for $d \geq 0$. Let $U(t)$ and $y(t)$ be the target population and value at t, respectively. For a unit i, let $I_i(t;t + d) = 1$ if the unit is to be included in the target population and 0 otherwise, i.e. based on the information available at $t + d$, and let $y_i(t;t + d)$ be the observed value for t at $t + d$. The data are said to be *progressive* if, for $d \neq d' > 0$, we *can* have

$$I_i(t; t + d) \neq I_i(t; t + d') \qquad \text{or} \qquad y_i(t; t + d) \neq y_i(t; t + d')$$

which lead to coverage errors and measurement errors, respectively, or both. Progressiveness is a distinct feature of administrative data sources compared to sample surveys, unless one is determined to *overlook* all delays and changes after a certain period.

The finite-population prediction framework as e.g. expounded in Valliant et al. (2000) needs to be extended given progressive data. Let $U_e(t;t + d)$ be a known universe of *existent* units. For instance, in repeated statistical production, one may include in $U_e(t;t + d)$ all the units that have previously been included in the population, i.e. $U_e(t;t + d) = \{i;\sum_{j=0}^{\infty} I_i(t - j;t + d) > 1\}$. The existent universe U_e admits a bipartition, denoted by $U_e = U_{e+} \cup U_{e-}$, where U_{e+} contains the units that actually belong to the target universe $U(t)$, i.e. $I_i(t) = 1$, and U_{e-} those that do not, i.e. $I_i(t) = 0$. Put $U_0(t;t + d) = U(t) \setminus U_{e+}(t;t + d)$, i.e. the target units that are not included in the existent universe. One may refer to $U_0(t;t + d)$ as the *birth delays* and U_{e-} the *death delays*. Let $Y(t) = \sum_{i \in U(t)} y_i(t)$ be the target total of interest. A general expression of a prediction-based estimator of $Y(t)$ at $t + d$ can be given as

$$\hat{Y}(t; t + d) = \sum_{i \in U_e(t;t+d)} \hat{I}_i(t)\hat{y}_i(t) + \sum_{i \in U_0(t;t+d)} \hat{y}_i(t) \qquad (4)$$

Zhang and Pritchard (2013) apply the prediction framework (4) to VAT register data in UK. Under certain assumptions, such as $y_i(t) = y_i(t;t + d)$ provided $y_i(t;t + d) > 0$, it is possible to replace $\hat{I}_i(t)$ and $\hat{y}_i(t)$ with observed values $I_i(t;t + d)$ and $y_i(t;t + d)$,

respectively, for some units in $U_e(t;t+d)$. But the progressiveness implies that this is not the case in all applications, which is an important difference to the standard finite-population prediction framework. Notice that $U_e(t;t+d)$ may vary at different d and, since $I_i(t)$ is not all known in general, so are $U_{e+}(t;t+d)$ and $U_0(t;t+d)$. Moreover, since $U_0(t;t+d)$ contains the birth delay units whose existence is unknown at $t+d$, one can not possibly predict each individual $I_i(t)$ for $i \in U_0(t;t+d)$, but only their total $Y_0(t;t+d) = \sum_{i \in U0(t;t+d)} y_i(t)$. Finally, it may be noticed that the difference from the standard finite-population prediction framework is fundamentally due to the indeterminate frame of the target population. Indeed, provided $U(t)$ is known at $t+d$, i.e. when progressiveness causes only measurement errors, then we would obtain $\hat{Y}(t;t+d) = \sum_{i \in U(t)} \hat{y}_i(t)$, which is equivalent to the standard prediction framework that allows for measurement errors. The prediction approach (4) requires modelling of $\{I_i(t), y_i(t); i \in U_e(t;t+d)\}$ and $Y_0(t;t+d)$, with or without conditioning on historic y- and I-values and other relevant auxiliaries. Zhang and Pritchard (2013) notice potential connections to the literature on estimation in the presence of nonresponse and informative sampling. The modelling of longitudinal progressive data should provide interesting topics for future research.

References

[1] Basu, D. (1971). An essay on the logical foundations of survey sampling, Part 1 (with discussion). In: Godambe and Sprott (Eds.), *Foundations of Statistical Inference*, pp. 203-242. Holt, Reinhart and Winston, Toronto.

[2] Biemer, P. and Lyberg, L. (2003). *Introduction to Survey Quality*. John Wiley & Sons, Inc.

[3] Daas, P. J.H. and Puts, M.J.H. (2014) Big Data as a Source of Statistical Information. *The Survey Statistician*, no. 69, pp. 22-31.

[4] Fellegi, I.P. and Sunter, A.B. (1969). A theory for record linkage. *Journal of the American Statistical Association*, vol. **64**, pp. 1183-1210.

[5] Fienberg, S.E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika*, vol. **59**, pp. 409 - 439.

[6] Fienberg, S. E., Makov. and R. J. Steele (1998), Disclosure limitation control using perturbation and related methods for categorical data. *Journal of Official Statistics*, vol. **14**, pp. 485-502.

[7] Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., and Tourangeau, R. (2009). *Survey Methodology*, 2nd Edition. John Wiley & Sons, Inc.

[8] Hedlin, D., Fenton, T., McDonald, J.W., Pont, M. and Wang, S. (2006). Estimating the undercoverage of a sampling frame due to reporting delays. *Journal of Official Statistics*, vol. **22**, pp. 53-70.

[9] Herzog, T.N., Scheuren, F.J. and Winkler, W.E. (2007). *Data Quality and Record Linkage Techniques*. Springer.

[10] Hogan, H. (1993). The Post-Enumeration Survey: Operations and results. *Journal of the American Statistical Association*, vol. **88**, pp. 1047-1060.

[11] IWGDMF - International Working Group for Disease Monitoring and Forecasting. (1995). Capture-recapture and multiple-record systems estimation I: History and theoretical development. *American Journal of Epidemiology*, vol. **142**, pp. 1047-1058.

[12] Linkletter, C.D. and Sitter, R.R. (2007). Predicting natural gas production in Texas: An application of nonparametric reporting lad distribution estimation. *Journal of Official Statistics*, vol. **23**, pp. 239-251.

[13] Neter, J., Maynes, E.S. and Ramanathan, R. (1965). The effect of mismatching on the measurement of response error. *Journal of the American Statistical Association*, vol. 60, pp. 1005-1027.

[14] Rubin, D. (1993). Discussion, statistical disclosure limitation. *Journal of Official Statistics*, vol **9**, pp. 461-468.

[15] Sadinle, M. and Fienberg, S.E. (2013). A Generalized Fellegi–Sunter Framework for Multiple Record Linkage With Application to Homicide Record Systems. *Journal of the American Statistical Association*, vol. **108**, pp. 385-397.

[16] Scholtus, S. and Bakker, B.F.M. (2013). *Estimating the validity of administrative and survey variables through structural equation modeling: a simulation study on robustness*. Statistics Netherlands, Discussion paper 201302.

[17] Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. John Wiley & Sons, Inc.

[18] Weisberg, H. (2005). *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. The University Chicago Press.

[19] Wolter, K. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, vol. **81**, pp. 338-346.

[20] Zhang, L.-C. (2011). A unit-error theory for register-based household statistics. *Journal of Official Statistics*, vol. **27**, pp. 415-432.

[21] Zhang, L.-C. (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica*, vol. **66**, pp. 41-63.

[22] Zhang, L.-C. and Fosen, J. (2012). A modelling approach for uncertainty assessment of register-based small area statistics. *Journal of the Indian Society of Agricultural Statistics*, vol. **66**, pp. 91-104.

[23] Zhang, L.-C. and Pritchard, A. (2013) Short-term turnover statistics based on VAT and Monthly Business Survey data sources. *ENBES workshop 2013, Nuremberg*.

[24] Zhang, L.-C. (2014). Om modelling register coverage errors. *Journal of Official Statistics, to appear*.

New and Emerging Methods – Call for Volunteers

If you're interested in contributing an article to the "New and Emerging Methods" section of a future edition of *The Survey Statistician*, please contact Mick Couper at mcouper@umich.edu.

Book and Software Review

Statistical Methods for Handling Incomplete Data
J.K. Kim and J. Shao.

Chapman and Hall, 2014. ISBN 978-1-4398-4963-7. xi + 211 pages.

Book Review by David Haziza, Université de Montréal

This book was written by two internationally reputed researchers who have contributed tremendously to the development of theory and methods for handling missing data in the last two decades.

As the authors write in their preface, missing data is frequently encountered in statistics. For example, it occurs in surveys because some of the sampled units refuse to respond to the survey or because of the inability to contact them. Dropout or noncompliance in clinical trials may also lead to missing responses for some subjects. It is well known that unadjusted estimators may be heavily biased if the respondents differ from the non-respondents systematically with respect to the study variables. In the last three decades, there has been a massive development of statistical methods all sharing the same goal: obtain valid inferences in the presence of missing data.

This text provides a rigorous treatment of statistical inference in the presence of missing data and covers many recent developments in this important area, including fractional imputation, doubly robust procedures and methods for handling missing data in longitudinal and clustered data. The text provides complete proofs of the essential results. Every time a new concept is introduced, it is accompanied with at least one (but most of the time, several) example.

The book is accessible to readers that have a good background (at the advanced undergraduate level or graduate level) in mathematical statistics including likelihood inference as well as a good background in statistical modelling including linear regression models and models for binary data. The book is suited for PhD level graduate students and researchers in the area of missing data.

The book contains a nice collection of end-of-chapter exercises, mostly theoretical, some of which consist of proving/verifying results stated but not proved in the text. These exercises will prove useful to instructors teaching a graduate level class.

Some sections/chapters are fairly technical (e.g., Section 4.2 on the theoretical properties of imputed estimators and Chapter 7 on longitudinal and clustered data), which may be problematic for some readers.

The book consists of 9 chapters, the first one providing a brief introduction and presenting the outline of the book. The other chapters are:

Chapter 2. Likelihood based approach
Chapter 3. Computation
Chapter 4. Imputation
Chapter 5. Propensity scoring approach
Chapter 6. Non ignorable missing data
Chapter 7. Longitudinal and clustered data
Chapter 8. Application to survey sampling
Chapter 9. Statistical matching

Chapters 2 and 3 deal with maximum likelihood estimation. Chapter 2 starts by reviewing maximum likelihood estimation for complete data. Then, the important concept of observed likelihood is introduced along with the concepts MCAR, MAR and NMAR. Chapter 3 describes several approaches for computing maximum likelihood estimators. This is a very nice chapter, tightly written, that would prove very useful to graduate students wanting to implement maximum likelihood estimation. It covers traditional methods such as the Newton-Raphson algorithm and the EM algorithm as well as Monte Carlo methods such as Monte Carlo EM and data augmentation. There is also an interesting section on the factoring likelihood approach, which is useful when the data are MAR and the nonresponse pattern is monotone.

Chapter 4 examines the theoretical properties of point and variance estimators under both single and multiple imputation. The concepts of nonresponse variance and imputation variance are clearly explained. Methods for reducing the imputation variance due to stochastic imputation, including the use of several imputed values to replace a missing value. Variance estimation, including replication methods, is a nice feature of the book. In the multiple imputation section, there is an interesting discussion about the congeniality condition which is required for the asymptotic unbiasedness of the multiple variance estimator of Rubin (1987). As this condition is often not satisfied in finite population sampling, multiple imputation generally leads to invalid inferences in this context. An alternative general tool for imputation is fractional imputation that has been developed, in the last decade, by the first author, J.K. Kim, and co-authors. It is presented in Section 4.6. Multiple examples are given so the discussion is relatively easy to follow.

Propensity score estimation is the topic of chapter 5. It starts by distinguishing the outcome regression model (also called the imputation model in a survey context) approach from the response probability model approach. The authors first assume that the response probabilities are known, which is, in my view, a pedagogical way to present this topic. Interesting links are established with well-known estimators (Horvitz-Thompson estimator and regression estimator) in the field of survey sampling. Then, the propensity score adjusted estimator based on estimated response probabilities is presented and its theoretical properties studied. Nonparametric estimation of the response probabilities is discussed in Section 5.7. Doubly robust estimation procedures, which have been widely treated in the literature, are nicely presented in Section 5.5.

Chapters 6-9 cover more advanced topics suited for more advanced readers doing research in the field of missing data. In these chapters, many recent articles are discussed, which is a nice feature of the book. In Chapters 2-5, the results were essentially developed under the MAR assumption. In practice, the probability of response may depend on variables that are not always observed, in which case the

data are NMAR. This is frequent in longitudinal and clustered data. Methods for dealing with this type of data are presented in Chapter 6, including the generalized method of moments approach, the pseudo-likelihood approach and the latent variable approach. Longitudinal and clustered data are often encountered in practice and are the subject of Chapter 7. Each section corresponds to a particular set of assumptions about the missing mechanism. Chapter 8 deals with inference in the presence of missing survey data. The authors revisit fractional imputation and propensity score adjusted estimation. Variance estimation through the so-called reverse approach is also presented. Important topics such as calibration weighting and two-phase sampling are also discussed. The text ends with a short chapter on statistical matching, which is a topic that is likely to attract some attention in the future.

I strongly recommend this text to statisticians who are interested in using state-of-the-art methods for dealing with missing data because it provides a unified, yet accessible treatment of the past and current work in the area of missing data. Anyone doing research in the area of missing data should definitely have this book on his/her bookshelf.

David Haziza,
Université de Montréal

We are interested in fostering review of books and software in the area of survey methods. This would include standard review of individual books or software packages. This may also include broader reviews of groups of text and monographs in specific sub-areas; or similarly broad reviews of available software. Of particular interest are some of the new R libraries that have been developed recently for survey methods. If you are able to write a) review for this section, please contact Natalie Shlomo (natalie.shlomo@manchester.ac.uk).

Australia

DataAnalyser – A new approach for disseminating microdata

The Australian Bureau of Statistics (ABS) is required by legislation to ensure that no statistical outputs are released in a manner that is likely to enable the identification of a particular person or organisation, but the ABS is also required to provide outputs from any data that is collected under legislation. Balancing both of these requirements can be a difficult task, especially for users seeking access to microdata. The ABS recognised that there was a demand for greater detail and quicker access to its microdata. In response to this demand, the ABS started to develop two new products: TableBuilder and DataAnalyser. TableBuilder is an online tool that allows users to create tables and graphs from ABS Microdata. DataAnalyser is also an online tool, with a menu-driven interface that enables clients to perform more complex statistical exploration and analysis of ABS microdata.

Within DataAnalyser, users have the ability to generate tables; create hex bin plots (confidentialised scatter plots); define new categorical variables by combining existing variables using AND / OR statements; and define new continuous variables by combining existing variables or constants with basic mathematical operations. Users also have the ability to drop records from the datasets using expressions involving existing variables on the dataset. DataAnalyser will allow users to perform regressions on the microdata, with linear, logit, probit, Poisson and multinomial regression models available in the system.

DataAnalyser confidentialises output on the fly, meaning that all confidentiality protections are applied automatically so that users see the results in seconds rather than hours or days. The main sources of protection in DataAnalyser are perturbation (which adds random noise to the estimates) and the dropping of records. Functionalities available in DataAnalyser employ different confidentiality protections to ensure that the confidentiality of individuals and/or businesses meets the requirements of the legislation.

A Beta release of DataAnalyser has recently been released. For more information about DataAnalyser, please contact Gareth Biggs (gareth.biggs@abs.gov.au).

Canada

G-Sam

The redevelopment of the Generalized Sampling System (G-Sam) started in April 2011 and was completed two years later. G-Sam is a Corporate Business Architecture (CBA) project that was conceived in 2010/2011 after a review of the generalized systems for sampling, estimation and imputation. Risks were assessed and, according to the CBA module, "... it was decided that the redevelopment of the Generalized Sampling System was the highest priority." G-Sam was one of the first tools to be redeveloped under the sponsorship of the CBA program.

The original Generalized Sampling System was conceived in the late 1980s. It was based on a now-outdated system architecture and it was becoming difficult to add functionality and to maintain the software. Redeveloping it required a team of methodologists and systems engineers. Laurie Reedman, the chief of Quality Assurance and Generalized Systems in Business Survey Methods Division (BSMD), headed up the methodologists. The science of sampling is constantly evolving, so not only did the methodologists look at existing methods, but they also developed new algorithms, taking advantage of current computing power.

The systems engineers were led by Yves Deguire, the chief of the SAS Technology Centre in System Engineering Division (SED). They researched software tools and programming techniques for building robust and efficient systems to develop software that would run reliably and quickly in production. Laurie Reedman stresses that "it was a true partnership between methodologists and systems engineers that resulted in a successful project."

Survey managers want to spend as little of their budget as necessary on sample units; therefore, the sample must be as representative as possible. G-Sam provides the functionality needed to draw a sample—that is, to stratify the population, allocate sample units to the strata, and draw the sample while controlling overlap with other surveys.

The Integrated Business Statistics Program (IBSP) is the first big client to use G-Sam. IBSP methodologists participated in the testing of modules as they were rolled out. An extra challenge came from the fact that IBSP was in development in parallel with G-Sam. Consequently, not all of the requirements were known in advance and some features of the sampling system had to be fine-tuned during development. The Generalized Systems Steering Committee was very helpful in determining priorities and controlling the scope of the project.

As with any software development, there are always more improvements to be made. Therefore, G-Sam maintenance activities have already begun. There is even interest in a sampling service to allow non-methodologists to use G-Sam. This is the motivation behind a pilot project with IBSP that is planned for the future.

For more information on G-Sam, contact Laurie Reedman at laurie.reedman@statcan.gc.ca.

India

Dr. Gayatri Vishwakarma

India celebrates National Statistics Day which is the birth anniversary of scientist and applied statistician P.C. Mahalanobis on June 29 every year. Taking into consideration the importance of statistics for planned economic development in an underdeveloped country under long colonial rule with poorly developed administrative structure and resources, he changed the system of National Sample Surveys as an efficient and cheap means of data collection. He is the founder of Indian Statistical Institute (ISI), a world class institute. Furthermore he started the Central Statistical Organization in the Government.

On this occasion, the Ministry of Statistics and Programme Implementation launched a knowledge-sharing platform for public in collaboration with NDSAP Programme Management Unit (PMU). India Statistics Community has been formed on Data Portal India in a bid to bring the data producers and users together and to create public awareness about the role of statistics in socio-economic planning and policy formulation.

India's National Informatics Centre (NIC) has launched a new version of its open data platform, 'Open Government Data Platform India v2.0', providing a better experience to both data providers and platform users with a unified interface. Some of the key improvements to the platform are responsive web design, better discoverability of resources, enhanced visualization tools, and an option to enable SMS and email alerts. The platform currently has more than 6000 open government datasets.

International Indian Statistical Association (IISA) will organize its 2014 conference on the advancements in the fields of Statistics, Biostatistics, Probability, and their application areas at Riverside, CA on July 11-13 2014. This conference is co-sponsored by the American Statistical Association.

Securities and Exchange Board of India (SEBI), in its silver jubilee year, has organized its First International Research Conference during January 27-28, 2014 in Mumbai. The theme of the Conference was "HFT, Algo Trading and Co-location". Academicians/market practitioners/ regulators, having experience in the field, from countries such as USA, Spain, Australia, Canada, Japan and India have participated in this conference. During the one and a half days of the conference, the participants discussed the issues related to impact of HFT on Market Quality, Financial Stability, Information asymmetry and retail investors, HFT in developing countries, regulatory mechanism and technology as an enabler to re-level the field.

Fiji

M.G.M. Khan

GDP Rebase

The Fijian economy, in constant terms, is currently measured based on its 2008 structure. The 2008 GDP rebase exercise was the latest conducted with the results published in October, 2013. A total of 19 industries were surveyed as per the Fiji Standard Industrial Classification (FSIC) 2010, and detailed reports were prepared. FSIC is a localised version of the ISIC Rev.4. The industry reports provide a wide range of information that enables the estimation of GDP using the Production, Expenditure and Income approach.

With the increase in demand for a recent base year, the Fiji Bureau of Statistics is currently working on a GDP rebase for the year 2011. A Supply and Use table for 2011 will also be compiled. The 2011 industry reports for the entire economy are being finalised for later use in the GDP rebasing exercise. Apart from the industry reports, the indicators and deflators used for GDP estimation like the Industrial Production Index, Consumer Price Index, Import & Export Price Index and Building Material Price Index will all be rebased to the year 2011.

These are important development works to update Fiji's GDP by expenditure and income approach which are currently available up to 2005. These statistics are published on our website www.statsfiji.gov.fj as well as on our quarterly publication titled Key Statistics.

The income and expenditure approach and Quarterly GDP will be reviewed by the Pacific Financial Technical Assistance Centre (PFTAC) consultants later this year after which the latest information will be published. Contact persons: Surveys – Ms Amelia Tungi (ameliat@statsfiji.gov.fj) and National Accounts – Mr. Bimlesh Krishna (bkrishna@statsfiji.gov.fj).

Improvement of External Sector Statistics

Fiji is one of the countries in the Pacific region engaged in a three-year IMF project funded by the Government of Japan to improve External Sector Statistics (ESS). The Japan Sub Account (JSA) project was launched in October 2012 and aims to improve the accuracy, availability, comparability, and timeliness of ESS in the beneficiary countries. Other Pacific Island Countries who are part of the Project are; Kiribati, Marshall Islands, Micronesia, Palau, Papua New Guinea, Samoa, Solomon Islands, Timor-Leste, Tonga, Tuvalu, and Vanuatu. A series of technical assistance missions and two topical workshops have been conducted in the region. Other missions and another workshop are planned for 2014. As part of this improvement Fiji will be releasing its first External Debt Statistics by mid-2014. For further information on the JSA ESS Project please contact:

Mr. Fernando Lemos, Manager for the Project (email: flemos@imf.org) and Mr. Borys Rolando CottoCobar, External Sector Statistics Advisor for the Project (email: bcottocobar@imf.org)

For further information on Fiji's External Debt Statistics Release contact: Ms Sashee Nath (Statistician Balance of Payments) on email: snath@statsfiji.gov.fj

Producer Price Index

Fiji's first ever Producer Price Index (PPI) is scheduled to be published by April 2014. The index would be compiled on a quarterly basis with its reference period being March Quarter 2011 = 100.0. Producer Price Indexes measure changes in the prices of domestic producer goods and services. The PPI is a better deflator for estimating Constant Price GDP. It will only focus on the 'Goods producing' industries while the 'Services' sector will be covered later on. A price index is a measure of the proportionate, or percentage, changes in a set of prices over time. Contact personnel are Mr Sitiveni Sikivou (ssikivou@statsfiji.gov.fj) and Ms Komal Devi (kdevi@statsfiji.gov.fj)

Household Income and Expenditure Survey (HIES)

The Fiji Bureau of Statistics is currently conducting the 2013-2014 HIES, the HIES is a year-long nationwide survey which gathers information on household income and expenditure from representative sample of households. The major uses of HIES data are as follows:

1. The reweighting of the Consumer Price Index (CPI).
2. The collection of Household Sector information which is an important input to the compilation of GDP. Benchmark estimates are derived when such surveys are carried out while indicators are used to provide non-survey year estimates.
3. Reveal the extent and nature of poverty in Fiji.
4. Informal sector studies.

The contact person is Mr. Tevita Vakalalabure, Acting Senior Statistician (Household Surveys) on email: tvakalalabure@statsfiji.gov.gov.fj

Poland

Tomasz Żądło

Data on economic, social, demographical and environmental situation of Poland available in English (http://www.stat.gov.pl/bdlen/app/strona.html?p_name=indeks) on the website of the Central Statistical Office (Local Data Bank) are being updated. Data including Agriculture Censuses (1996, 2002, 2010) and Population and Housing Censuses (1988, 2002, 2011) are presented for different NUTS levels.

The Small Area Estimation Conference 2014 is organized by Poznań University of Economics in cooperation with the Central Statistical Office and the Statistical Office in Poznań. It will take place in Poznań (Poland) 3-5 September. More information is available on the conference website: http://www.sae2014.ue.poznan.pl/

Palestine

Symposium in Celebration of the International Year of Statistics 2013
Tuesday, 09 April, 2013, Abu-Dies Campus, Palestine

The Department of Mathematics, Al-Quds University, Palestine, has conducted a symposium on 9 April 2013, in Celebration of the International Year of Statistics2013 "Statistics2013". The symposium was moderated by the associate professor of analysis of complex survey data under informative sampling design, Dr. Abdulhakeem Eideh. The lectures were delivered on different aspects of social statistics, for example, analysis of survey data, basic concepts of structural equation models, and what role economic statistics play in econometrics?

Palestinian Central Bureau of Statistics (PCBS), Palestine
An international conference on official statistics
Tuesday, 24/09/2013

The Palestinian Central Bureau of Statistics organized an international conference on official statistics on 24 September 2013. The conference commemorates a twenty years passage of the establishment of the Palestinian Central Bureau of statistics and coincides with the International Year of Statistics 2013, launching of the national strategy for official statistics 2014 –2018 and launching of the statistical monitoring system. This international conference includes many working papers submitted by Arab and international statistical institutions, international experts, national, regional and international research institutions. The papers were distributed on different topics concerning official statistics, for example, trends in information and communication technology, data dissemination and circulation, statistics and data for monitoring, and linking administrative and survey sources.

Switzerland

Construction of Full Time Equivalent for the Swiss Structural Business Statistics

In the past, a periodical Business Census played an important role for producing various statistics on the structure of the Swiss economy. The Business Census was held for the last time in 2008, but for the reference year 2011 it is replaced by a new system, called STATENT (Statistique Structurelle des Entreprises). STATENT is based on multiple sources like the business register, the social security register and complementary surveys, in particular the Quarterly Survey of Employment. The social security register, which contains information about gender, employment and wages (annual salary) at employee level, is linked to the business register at the enterprise level. Therefore, it is the main source of information about employment at the enterprise level while the business register provides information like NACE-code, legal form or enterprise structure (multi-establishments). Full-time equivalent (FTE) by gender, which is an important target variable, is known for the enterprises surveyed in the Quarterly Survey of Employment but is not available in the social security register for the other enterprises. This information is therefore estimated at the enterprise level by a linear model using explanatory variables from the business and the social security registers. The model is fitted on the linked data set combining

the social security register, the business register and the quarterly survey of employment.

The development of the model, its estimation and the assessment of its prediction quality are important methodological milestones within STATENT. On the other hand the integration of survey and administrative data revealed differences regarding the number of employees. The treatment of these differences, in order to obtain consistency between the survey FTE and the social security data, is another major methodological challenge. A deeper analysis showed that the simple ratio adjustment initially used for treating the inconsistencies was not optimal. Another approach is now used. It allows a better treatment of these differences by taking into account the type of employment (occupation rate, wage level). Note that with this new approach the same results are obtained as with the ratio adjustment when the missing or in excess employments can be considered as uniformly distributed on the whole range of occupation rates.

The definitive STATENT results for 2011 and the preliminary ones for 2012 are planned for August 2014.

For more information please contact Desislava Nedyalkova (Desislava.Nedyalkova@bfs.admin.ch) or Daniel Assoulin (Daniel.Assoulin@bfs.admin.ch)

Methodology of the coverage survey of the new census in Switzerland

The population census in Switzerland has been replaced by a system that is based on administrative data. Any population census may contain errors with regard to under- and over-coverage: some people may be forgotten by the statistics, while others may be incorrectly counted or even counted twice. A coverage survey was carried out following the last classic census in 2000. In order to evaluate the quality of its new census, the Federal Statistical Office (FSO) carried out a new coverage survey (EC2013) at the start of 2013. As in 2000, it is based on capture-recapture type methods. The EC2013 also additionally aims to evaluate the coverage of the Federal Register of Buildings and Dwellings.

A sample of geographical areas is chosen, then all the buildings, dwellings and inhabitants of the areas are enumerated. The probability that areas will be selected depends on their population density. Furthermore, the sample is balanced on the number of buildings, dwellings and persons in each area according to the FSO's data. If, as we can imagine, the survey's main variables of interest are quite similar to the balancing variables, this survey plan should allow greater precision.

One of the challenges of the weighting is to correct the non-response imbalance even though we are unaware of the number of units that a given area actually contains. Capture-recapture methods are also used during this stage. Once the field survey data have been weighted, we compare these to the FSO data in order to work out the under- and over-coverage rates.

The data are still being processed. This is expected to come to a close at the end of June 2014. However, the initial, provisional results give cause to believe that the new system is of very high quality.

Please contact Anne Massiani (anne.massiani@bfs.admin.ch) or Lionel Qualité (lionel.qualite@bfs.admin.ch) at the Federal Statistical Office for further information.

Upcoming Conferences and Workshops

60th ISI World Statistics Congress

Organized by: International Statistical Institute

Date: 26 – 31 Jul 2015

Venue Riocentro Rio de Janeiro, Brazil

Homepage: http://www.isi2015.org/

The congress will provide participants with an opportunity to meet and exchange ideas with members of the statistical community from more than one hundred countries, working in all fields of statistical sciences. The scientific programme will contain presentation and panel discussions by researchers, educators, officials from national and international organisations, and practitioners from industry. It will cover a broad range of topic dealing with cutting-edge research and practice, novel applications, new developments, and emerging challenges relevant to science and public policy. There will be also plenty of opportunities for discussion and exchange.

A rich and exciting Social Programme is also being developed, with plenty to see and enjoy for participants and their accompanying persons. We want to make sure that you trip to Rio and participation in ISI2015 is a truly unforgettable experience.

Enquiries/More information

For more information, contact wsc2015@ibge.gov.br

The 6th Conference of the European Survey Research Association
Organized by: European Survey Research Association (ESRA)

Date: 13 – 17 Jul 2015

Venue Reykjavik, Iceland

Homepage: http://www.europeansurveyresearch.org/conference

The 6th Conference of the European Survey Research Association (ESRA) will take place 13th-17th July 2015 in Reykjavik, Iceland. The scientific committee is now inviting researchers who are active in the field of survey research and survey methodology to submit proposals to organise sessions at the conference. Session proposals are invited in any area of survey methodology, or in substantive areas of survey research. We encourage proposals from researchers with a variety of backgrounds, including academic research, national statistics and market research. The following are examples of topics that are of particular interest:

- Sample designs, coverage, and sampling
- Fieldwork processes
- Unit and item nonresponse
- Weighting and imputation
- Questionnaire development, testing and piloting
- Interviewers and interviewer effects
- Mixing modes and mode effects
- Online survey methods
- Surveys using mobile devices
- Linking survey data to auxiliary data sources
- Using paradata to evaluate survey quality
- Methods for cross-national and cross-cultural surveys
- Longitudinal surveys and longitudinal analysis techniques
- Methodological considerations specific to certain survey modes: face-to-face, phone, web, mail, etc.
- Analysing, monitoring and reducing the Total Survey Error
- Data documentation, archiving and data access
- Survey analysis techniques
- Election polling and public opinion
- Social indicators
- Substantive applications of survey research

2014 Joint Statistical Meetings

Organized by: American Statistical Association, Institute of Mathematical Statistics, International Biometric Society (ENAR and WNAR), International Chinese Statistical Association, International Indian Statistical Association, International Society for Bayesian Analysis, Korean International Statistical Society, and Statistical Society of Canada

Date: 2 - 7 Aug 2014

Venue: Boston Convention and Exhibition Center, Boston, USA

Homepage: http://www.amstat.org/meetings/jsm/2014/

JSM (the Joint Statistical Meetings) is the largest gathering of statisticians held in North America. Attended by more than 6,000 people, meeting activities include oral presentations, panel sessions, poster presentations, continuing education courses, an exhibit hall (with state-of-the-art statistical products and opportunities), career placement services, society and section business meetings, committee meetings, social activities and networking opportunities.

For information, contact meetings@amstat.org.

Workshop of the Baltic-Nordic-Ukrainian Network on Survey Statistics

Organized by: Statistics Estonia, University of Tartu and the Estonian Statistical Society

Date: 25-28 Aug 2014

Venue Statistics Estonia, Tatari 51, Tallinn, Estonia.

Homepage: http://www.stat.ee/workshop-on-survey-statistics-theory-and-methodology-2014

This workshop is part of the series of events organized by the Baltic-Nordic-Ukrainian Network on Survey Statistics. The workshops, summer schools and international conferences on Survey Statistics are organized annually since 1997. Information about the past events can be found on the BNU Network webpage.

The aim of the network is to facilitate and encourage co-operation between researchers, students and practitioners in the field of survey statistics. The goal is to combine theory (academic institutions) with practice (national statistical offices). The workshop offers the possibility to share experience, present recent research, and learn from the experts in their field.

Enquiries/More information

Please contact Imbi Traat (invited speakers) imbi.traat@ut.ee, Maiki Ilves (practical matters) maiki.ilves@stat.ee, or Natalja Lepik (abstracts and papers) natalja.lepik@ut.ee.

ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

RSS 2014 International Conference

Organized by: Royal Statistical Society

Date: 1-4 Sept 2014

Venue: Sheffield City Hall, Sheffield, UK

Homepage: http://www.statslife.org.uk/events/annual-conference

Key Dates

The RSS Conference provides a unique opportunity in the UK for statisticians and users of statistics of all ages and professional backgrounds to gather and exchange knowledge and experiences, whether in the formal conference sessions or in the many opportunities for networking at refreshment breaks or at evening social events.

A strength of the conference is the breadth and variety of its programme of talks and workshops with sessions appealing to both theoretical and applied statisticians, those working in the areas of official, medical, environmental statistics (amongst many others), people working with data more generally or indeed those with a general interest in the topic.

The broad streams for the conference will be as follows:

- Bioinformatics, Genomics & Biostatistics
- Communication of statistical ideas (including data visualisation)
- Data science (including experimental design)
- Emerging topics (including big data)
- Environment and ecology
- Industry and commerce
- Medical, clinical trials and epidemiology
- Public sector and policy evaluation (including open data)
- Statistics in sport
- Statistical methods and theory

Enquiries/More information

For further information please contact Paul Gentry, p.gentry@rss.org.uk (RSS Meetings & Conferences Manager)

Small Area Estimation 2014

3-5 September, Poznan, Poland

Date: 3 – 5 Sept 2014

Venue Poznan University of Economics, Poland

Homepage: http://sae2014.ue.poznan.pl/index.html

SAE 2014 will be one of the most important events devoted to theoretical and methodological developments as well as practical applications of small area estimation methods.

The organizers of the SAE 2014 conference hope that it will provide a special opportunity and a platform for the exchange of ideas and experiences between representatives of the academic community, official statistics, research centers as well as other institutions involved in developing and applying small area estimation methods.

Plenary speakers:

• Prof. Malay Ghosh (University of Florida)

Empirical Bayes Small Area Estimation under Multiplicative Models

• Prof. J.N.K. Rao (Carleton University)

Inferential Issues in Model-based Small Area Estimation: Some New Developments

Enquiries/More information

For more information, contact sae2014@konf.ue.poznan.pl.

NISS National Institute of Statistical Sciences — The Statistics Community Serving the Nation

International Total Survey Error Workshop 2014

Organized by: National Institute of Statistical Sciences

Date: 1– 3 Oct 2014

Venue: Bureau of Labor Statistics Conference Center Washington, DC, USA

Homepage: http://www.niss.org/events/itsew2014

Key Dates

The 8th International Total Survey Error Workshop (ITSEW 2014) will take place from Wednesday, October 1 to Friday, October 3, 2014 at the Bureau of Labor Statistics Conference Center near Union Station in Washington DC. The theme of the 2014 workshop is **Total Survey Error: Fundamentals and Frontiers**.

Like previous ITSEWs, this year's gathering seeks to foster an exchange of ideas and preliminary research findings toward a better understanding of total survey error. Participation in this year's ITSEW will be particularly valuable for researchers planning to present invited or contributed papers at next year's *2015 International Total Survey Error Conference (TSE2015)*. Participants can share ideas, receive valuable feedback, identify opportunities for collaboration, and plan sessions or joint presentations for TSE2015 to be held in September, 2015 in Baltimore, Maryland.

Enquiries/More information

For more information contact itsew2014@niss.org.

The 2014 IAOS Conference on Official Statistics

Organized by: General Statistics Office, Vietnam

Date: 8 – 10 Oct 2014

Venue Pullman Danang Beach Resort, Da Nang, Vietnam

Homepage: http://isi.cbs.nl/iaos/Conferences/2014Vietnam.htm

The International Association for Official Statistics (IAOS) and the General Statistics Office (GSO) of Vietnam would like to welcome you to the IAOS Conference on Official Statistics to be held in Da Nang, Vietnam from 8 to 10 October 2014.

The IAOS Conference has been held biennially at many different places in the world. Each conference focuses on relevant themes reflecting current interests.

They promote the exchange of new experiences and ideas for improving official statistics among staff of national statistical institutes, government agencies that produce official statistics, and research and educational organizations, as well as users of official statistics.

The theme of IAOS 2014 is “Meeting the demands of a changing world”. The conference will cover a broad range of subjects from using new data sources - such as “big data” – to innovative forms of data sharing. We look forward to receiving your papers and to sharing new approaches and insights with international colleagues producing and using official statistics.

Enquiries/More information

For further information contact iaos2014@gso.gov.vn.

Statistics Canada 2014 International Methodology Symposium

Organized by: Statistics Canada

Date: 29 – 31 Oct 2014

Venue Palais des congrès de Gatineau, Gatineau, Canada

Homepage: http://www.statcan.gc.ca/conferences/symposium2014/index-eng.htm

The Symposium will be titled “Beyond traditional survey taking: adapting to a changing world”. Members of the statistical community, such as those from private organizations, governments, or universities, are invited to attend, and particularly those who have an interest in methodological challenges resulting from the use of non-traditional survey methods.

The Symposium will consist of both plenary and parallel sessions covering a variety of topics. Additional research and results might be presented via poster sessions.

Topics for the sessions may include:

- Big data
- Record linkage
- Administrative data
- Web (panel) surveys
- Optimization of data collection
- Mode effects
- Electronic questionnaire
- Model-based approach
- Balanced sampling
- Microsimulations
- Measurement error
- Total survey error

Enquiries/More information

For more information, contact symposium2014@statcan.gc.ca

10th International Conference on Transport Survey Methods (ISCSTC10)

Organized by: The International Steering Committee Travel Survey Conference series

Date: 16 – 21 Nov 2014

Venue Fairmont Resort, Leura, Australia

Homepage: http://www.regodirect.com.au/isctsc10/home/

Rapidly evolving problems and policy contexts are compelling us to advance the state-of-the-art of survey methods, tools, strategies and protocols, while assuring the stability and coherence of the data from which trends can be tracked and understood. This year's conference will build on outputs from previous conferences and try to address these emerging research issues.

It will place particular emphasis on two related elements:

1. New technologies: how they challenge traditional survey methods, their potential contributions to transportation planning and policy decision-making and the way they impact upon travel decisions.
2. Decision and behavioural processes: use of qualitative and quantitative datasets, behaviour changes with regard to understanding new issues and contexts (climate change, ageing population, social wellbeing, socio-technical transitions).

Key Dates

- 31 July 2014: Close of Early-bird Registration
- November 16-21, 2014: ISCTSC 10th International Conference on Transport Survey Methods

Enquiries/More information

For more information, contact helpline@vmsconferences.com.au.

Organized by: UNESCO Chair in Data Privacy

Date: 17 – 19 Sept 2014

Venue: Universitat de les Illes Balears, Eivissa, Balearic Island

Homepage: http://unescoprivacychair.urv.cat/psd2014/

Privacy in statistical databases is about finding trade-offs to the tension between the increasing societal and economical demand for accurate information and the legal and ethical obligation to protect the privacy of individuals and enterprises which are the respondents providing the statistical data. In the case of statistical databases, the motivation for respondent privacy is one of survival: statistical agencies or survey institutes cannot expect to collect accurate information from individual or corporate respondents unless these feel the privacy of their responses is guaranteed.

Beyond respondent privacy, there are two additional privacy dimensions to be considered: privacy for the data owners (organizations owning or gathering the data, who wouldn't like to share the data they have collected at great expense) and privacy for the users (those who submit queries to the database and would like their analyses to stay private).

"Privacy in Statistical Databases 2014" (PSD 2014) is a conference sponsored and organized by the UNESCO Chair in Data Privacy, with proceedings published by Springer-Verlag in Lecture Notes in Computer Science. It purpose is to attract world-wide, high-level research in statistical database privacy.

Like the preceding PSD conferences, PSD 2014 originates in Europe, but wishes to stay a worldwide event in database privacy and SDC. Thus, contributions and attendees from overseas are welcome.

Organized by: CMStatistics

Date: 6 – 8 Dec 2014

Venue: University of Pisa, Italy

Homepage: http://www.cmstatistics.org/ERCIM2014/index.php

The 7th International Conference of the ERCIM WG on Computational and Methodological Statistics (ERCIM 2014) will take place at the University of Pisa, Italy, 6-8 December 2014. Tutorials will be given on Friday 5th of December 2014.

This conference is organized by the ERCIM Working Group on Computational and Methodological Statistics (CMStatistics) and the University of Pisa. The journal Computational Statistics & Data Analysis publishes selected papers in special peer-reviewed, or regular issues.

The Conference will take place jointly with the 8th International Conference on Computational and Financial Econometrics (CFE 2014). The conference has a high reputation of quality presentations. The last editions of the joint conference CFE-ERCIM gathered over 1200 participants.

For further information please contact info@cmstatistics.org.

In Other Journals

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

VOL 30, NO 2 (2014)
http://iospress.metapress.com/content/x18176857331/?p=3f949982ec734fb0a4ed005d714a9ddbandpi=0

Editorial

Interview with Lars Thygesen

An evaluation of three disclosure limitation models
M. l Yang, M. Buso, S. Butani, D. Hiles, A. Mushtaq, S. Pramanik and F. Scheuren

Opening Statistics
M. A. Cameron

The reproducible research movement in statistics
Victoria Stodden

Meeting the challenges of data infrastructure for collaborative research
R. L. Sandland

Why data availability is such a hard problem
A. F. Karr

International support for data openness and transparency
M. V. Belkindas and E. V. Swanson

Synthetic establishment data: Origins and introduction to current research
J. M. Abowd

Expanding the role of synthetic data at the U.S. Census Bureau
R. S. Jarmin, T. A. Louis and J. Miranda

Journal of Survey Statistics and Methodology

VOL 2, ISSUE 2 (2014)

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

Design Effects for a Regression Slope in a Cluster Sample
S. L. Lohr

Evaluating Three Approaches to Statistically Adjust for Mode Effects
S. Kolenikov and C. Kennedy

Is the Collection of Interviewer Observations Worthwhile in an Economic Panel Survey? New Evidence from the German Labor Market and Social Security (PASS) Study
B.T. West, F. Kreuter and M. Trappmann

Efficient Use of Commercial Lists in U.S. Household Sampling
R. Valliant, F. Hubbard, S. Lee and C. Chang

Increasing Fieldwork Efficiency Through Prespecified Appointments
F. Kreuter, A. Mercer and W. Hicks

Survey Methodology

Statistics Canada Statistique Canada

DECEMBER 2013, VOL 39, NO 1

http://www.statcan.gc.ca/pub/12-001-x/12-001-x2013002-eng.htm

Three controversies in the history of survey sampling
K. Brewer

A Weighted composite likelihood approach to inference for two-level models from survey data
J. N. K. Rao, F. Verret and M. A. Hidiroglou

Comparison of different sample designs and construction of confidence bands to estimate the mean of functional data: An illustration on electricity consumption
H. Cardot, A. Dessertaine, C. Goga, E. Josserand and P. Lardin

Pseudo-likelihood-based Bayesian information criterion for variable selection in survey data
C. Xu, J. Chen and H. Mantel

Design-based analysis of factorial designs embedded in probability samples
J. A. van den Brakel

Estimation and replicate variance estimation of deciles for complex survey data from positively skewed populations
S. J. Kaputa and K. J. Thompson

Joint determination of optimal stratification and sample allocation using genetic algorithm
M. Ballin and G. Barcaroli

An appraisal-based generalized regression estimator of house price change
J. de Haan and R. Hendriks

Examining the effect of the welcome screen design on the response rate
R. Haer and N. Meidert

JOS

Journal of Official Statistics

June 2014, VOL 30, ISSUE 2
http://www.degruyter.com/view/j/jos.2014.30.issue-2/issue-files/jos.2014.30.issue-2.xml

Overview of the Special Issue on Surveying the Hard-to-Reach
G. B. Willis, T. W. Smith, S. Shariff-Marco and N. English

Potential Uses of Administrative Records for Triple System Modeling for Estimation of Census Coverage Error in 2020
R. A. Griffin

Technique for Remote, Hard-to-Reach, and Mobile Populations
K. Himelein, S. Eckman and S. Murray

Enumerating the Hidden Homeless: Strategies to Estimate the Homeless Gone Missing From a Point-in-Time Count
R. P. Agans, M.T. Jefferson, J. M. Bowling, D. Zeng, J. Yang and M. Silverbush

A Study of Assimilation Bias in Name-Based Sampling of Migrants
R. Schnell, M. Trappmann and T. Gramlich

Comparing Survey and Sampling Methods for Reaching Sexual Minority Individuals in Flanders

A. Dewaele, M. Caen and A. Buysse

A City-Based Design That Attempts to Improve National Representativeness of Asians
S. Pedlow

Recruiting an Internet Panel Using Respondent-Driven Sampling
M. Schonlau, B. Weidmer and A. Kapteyn

Locating Longitudinal Respondents After a 50-Year Hiatus
C. Stone, L. Scott, D. Battle and P. Maher,

Evaluating the Efficiency of Methods to Recruit Asian Research Participants
H. Park and M. M. Sha

Reaching Hard-to-Survey Populations: Mode Choice and Mode Preference
M. Haan, Y. P. Ongena and K. Aarts

VOL 7, NO 2 (2014)
www.surveypractice.org

Practical Guidelines for Dual-Frame RDD Survey METHODOLOGY (Now That the Dust is Settling)
M. Fahimi

Do In-person Interviews Reduce Bias in a Mixed-Mode Survey of Persons with Disabilities?
E. Grau

Predicting Bias in National House of Representatives Generic Ballot Questions
D. R Cassino

Measuring Propensity to Join the Military: Survey Data are Consistent Regardless of Response Option Order
J. R. Bergstrom, J. Hackenbracht and J. L Gibson

Undisclosed privacy. The effect of privacy rights design on response rates.
R. Haer and N. Meidert

Recent Books and Journals in Public Opinion, Survey Methods, and Survey Statistics
M. Callegaro

Survey Research Methods

VOL 8, NO 1 (2014)
https://ojs.ub.uni-konstanz.de/srm/issue/view/112

Age and Sex Effects in Anchoring Vignette Studies: Methodological and Empirical Contributions
H. Grol-Prokopczyk

Variation in Incentive Effects across Neighbourhoods
M. J. Hanly, G. M. Savva, I. Clifford and B. J. Whelan

On the Relative Advantage of Mixed-Mode versus Single-Mode Surveys
J. Vannieuwenhuyze

Measurement Error in Retrospective Work Histories
J. P. Sánchez, J. Koskinen and I. Plewis

Collecting Biomarkers Using Trained Interviewers. Lessons Learned from a Pilot Study
S. L. McFall, A. Conolly and J. Burton

International Statistical Review
Revue Internationale de Statistique

APRIL 2013, VOL 82, ISSUE 1
http://onlinelibrary.wiley.com/doi/10.1111/insr.v82.1/issuetoc

Soft Questions, Hard Answers: Jacob Bernoulli's Probability in Historical Context
S. Stigler

On the Bicentenary in St. Petersburg of Jacob Bernoulli's Theorem
E. Seneta

Tercentenary of Ars Conjectandi (1713): Jacob Bernoulli and the Founding of Mathematical Probability
E. Dudley Sylla

A Brief Survey of Modern Optimization for Statisticians
K. Lange, Eric C. Chi and H. Zhou

Latent Markov Models for Longitudinal Data by Francesco Bartolucci, Alessio Farcomeni, Fulvia Pennoni
S. Pynnönen

Current Topics in the Theory and Application of Latent Variable Models by Michael C. Edwards, Robert C. MacCallum (Editors)
S. Pynnönen

Multiple Imputation and its Application by James R. Carpenter, Michael G. Kenward
K. Nordhausen

Applied Categorical and Count Data Analysis by Wan Tang, Hua He, Xin M. Tu
T. Nummi

Advances in Machine Learning and Data Mining for Astronomy edited by Michael J. Way, Jeffrey D. Scargle, Kamal M. Ali, and Ashok N. Srivstava
K. Podgorski

Exercises in Probability: A Guided Tour from Measure Theory to Random Processes,via Conditioning by Loïc Chaumont and Marc Yor
K. Podgorski

Inference for Functional Data with Applications by Lajos Horváth and Piotr Kokoszka
G. A. Young

A Tiny Handbook of R by Mike Allerhand
A. Sengupta

An Introduction to Statistical Learning—with Applications in R by Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani
K. Nordhausen

Handling Missing Data in Ranked Set Sampling R by Carlos N. Bouza-Herrera
M. Ruiz Espejo

A First Course in Machine Learning by Simon Rogers and Mark Girolami
A. Sengupta

STATISTICS IN TRANSITION
new series

An International Journal of the Polish Statistical Association

VOL 15, NO 1, Winter 2014
http://pts.stat.gov.pl/en/journals/statistics-in-transition/

VOL 5, ISSUE 2 (2013)
http://repository.cmu.edu/jpc/

Improving User Choice Through Better Mobile Apps Transparency and Permissions Analysis
I. Liccardi, J. Pato, and D. J. Weitzner

Towards a Systematic Analysis of Privacy Definitions
B.-R. Lin and D. Kifer

Intruder Testing on the 2011 UK Census: Providing Practical Evidence for Disclosure Protection
C. Tudor, G. Cornish and K. Spicer

An Intuitive Formulation and Solution of the Exact Cell-Bounding Problem for Contingency Tables of Conditional Frequencies
S. E. Wright and B. J. Smucker

TRANSACTIONS ON
DATA PRIVACY

Foundations and Technologies
http://www.tdp.cat

April 2014, VOL 7, ISSUE 1
http://www.tdp.cat/issues11/vol07n01.php

Efficient Graph Based Approach to Large Scale Role Engineering
D. Zhang, K. Ramamohanarao, R. Zhang and S. Versteeg

IdentiDroid: Android can finally Wear its Anonymous Suit
B. Shebaro, O. Oluwatimi, D. Midi and E. Bertino

Combining Binary Classifiers for a Multiclass Problem with Differential Privacy
V. Sazonova and S. Matwin

Journal of the Royal Statistical Society SERIES A Statistics in Society

February 2014, VOL 177, ISSUE 2
http://onlinelibrary.wiley.com/doi/10.1111/rssa.2014.177.issue-2/issuetoc

Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics
J. Stoyanov

Exercises and Solutions in Biostatistical Theory
J. Stoyanov

R Graphics
S. K. Lhachimi

Journal of the American Statistical Association

VOL 109, ISSUE 506 (2014)
http://amstat.tandfonline.com/toc/uasa20/current

Flexible Marginal Structural Models for Estimating the Cumulative Effect of a Time-Dependent Treatment on the Hazard: Reassessing the Cardiovascular Risks of Didanosine Treatment in the Swiss HIV Cohort Study
Y. Xiao, M. Abrahamowicz, E. E. M. Moodie, R. Weber and J. Young

To Fuel or Not to Fuel? Is that the Question?
E. S. Ayra, D. R. Insua and J. Cano

A Bayesian Nonparametric Regression Model With Normalized Weights: A Study of Hippocampal Atrophy in Alzheimer's Disease
I. Antoniano-Villalobos, S. Wade and S. G. Walker

Automated Tsunami Source Modeling Using the Sweeping Window Positive Elastic Net
D. M. Percival, D. B. Percival, D. W. Denbo, E. Gica, P. Y. Huang, H. O. Mofjeld and M. C. Spillane

Bayesian Forecasting of Cohort Fertility
C. Schmertmann, E. Zagheni, J. R. Goldstein and M. Myrskylä

Estimating Risk With Time-to-Event Data: An Application to the Women's Health Initiative
D. Liu, Y. Zheng, R. L. Prentice and L. Hsu

Using Data Augmentation to Facilitate Conduct of Phase I–II Clinical Trials With Delayed Outcomes
I. H. Jin, S. Liu, P. F. Thall and Y. Yuan

A Bivariate Model for Simultaneous Testing in Bioinformatics Data
H. Y. Bar, J. G. Booth and M. T. Wells

Object-Oriented Data Analysis of Cell Images
X. Lu, J. S. Marron and P. Haaland

Source-Sink Reconstruction Through Regularized Multicomponent Regression Analysis—With Application to Assessing Whether North Sea Cod Larvae Contributed to Local Fjord Cod in Skagerrak
K. Chen, K.-S. Chan and N. Chr. Stenseth

Variable Selection in Nonparametric Classification Via Measurement Error Model Selection Likelihoods
L. A. Stefanski, Y. Wu and K. White

Group LASSO for Structural Break Time Series
N. H. Chan, C. Y. Yau and R.-M. Zhang

Testing Independence Among a Large Number of High-Dimensional Random Vectors
G. Pan, J. Gao and Y. Yang

Adaptive Multivariate Global Testing
G. Minas, J. A. D. Aston and N. Stallard

Adaptive Global Testing for Functional Linear Models
J. Lei

Local Empirical Likelihood Inference for Varying-Coefficient Density-Ratio Models Based on Case-Control Data
X. Liu, H. Jiang and Y. Zhou

Enriched Stick-Breaking Processes for Functional Data
B. Scarpa and D. B. Dunson

A Smooth Simultaneous Confidence Corridor for the Mean of Sparse Functional Data
S. Zheng, L. Yang and W. K. Härdle

Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules
R. Koenker and I. Mizera

A Generic Path Algorithm for Regularized Statistical Estimation
H. Zhou and Y. Wu

Sparse Additive Ordinary Differential Equations for Dynamic Gene Regulatory Network Modeling
H. Wu, T. Lu, H. Xue and H. Liang

Parametrically Assisted Nonparametric Estimation of a Density in the Deconvolution Problem
A. Delaigle and P. Hall

Estimation for General Birth-Death Processes
F. W. Crawford, V. N. Minin and M. A. Suchard

Nonparametric Regression for Spherical Data
M. Di Marzio, A. Panzera and C. C. Taylor

Spectral Density Ratio Models for Multivariate Extremes
M. de Carvalho and A. C. Davison

Self-Excited Threshold Poisson Autoregression
C. Wang, H. Liu, J.-F. Yao, R. A. Davis and W. K. Li

Selection of Mixed Copula Model via Penalized Likelihood
Z. Cai and X. Wang

A Class of Hazard Rate Mixtures for Combining Survival Data From Different Experiments
A. Lijoi and B. Nipoti

Fused Estimators of the Central Subspace in Sufficient Dimension Reduction
R. D. Cook and X. Zhang

EMVS: The EM Approach to Bayesian Variable Selection
V. Ročková and E. I. George

Simulated Stochastic Approximation Annealing for Global Optimization With a Square-Root Cooling Schedule
F. Liang, Y. Cheng and G. Lin

Book Reviews

Correction :

"Ensemble Subsampling for Imbalanced Multivariate Two-Sample Tests," Chen, L., Dou, W. W., and Qiao, Z. (2013), Journal of the American Statistical Association, 108, 1308–1323.
L. Chen, W. W. Dou and Z. Qiao

BIOMETRIKA

June 2013, VOL 100, ISSUE 4
http://biomet.oxfordjournals.org/content/current

Direct estimation of differential networks
S. D. Zhao, T. T. Cai and H. Li

Variance estimation in high-dimensional linear models
L. H. Dicker

Bayes and empirical Bayes: do they merge?
S. Petrone, J. Rousseau and C. Scricciolo

Bayesian monotone regression using Gaussian process projection
L. Lin and D. B. Dunson

Latin hypercube designs with controlled correlations and multi-dimensional stratification
J. Chen and P. Z. G. Qian

Permuting regular fractional factorial designs for screening quantitative factors Yu Tang and
H. Xu

Indicator functions and the algebra of the linear-quadratic parameterization
A. Sabbaghi, T. Dasgupta and C. F. J. Wu

Locally ϕp-optimal designs for generalized linear models with a single-variable quadratic polynomial predictor
H.-P. Wu and J. Stufken

Logistic regression for spatial Gibbs point processes
A. Baddeley, J.-F. Coeurjolly, E. Rubak and R. P. Waagepetersen

A combined estimating function approach for fitting stationary point process models
C. Deng, R. P. Waagepetersen and Y. Guan

Distances and inference for covariance operators
D. Pigoli, J. A. D. Aston, I. L. Dryden and P. Secchi

Measurement bias and effect restoration in causal inference
M. Kuroki and J. Pearl

Propensity score adjustment with several follow-ups
J. K. Kim and J. Im

Testing equality of a large number of densities
D. Zhan and J. D. Hart

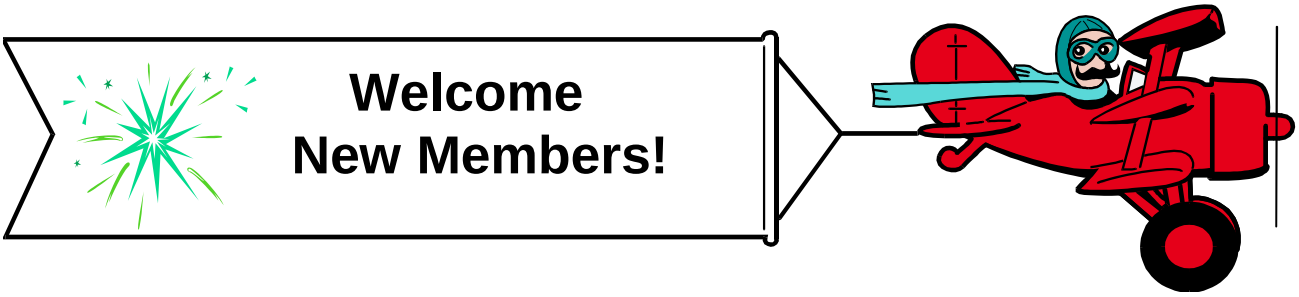
Bootstrap for the case-cohort design
Y. Huang

Hypothesis testing for band size detection of high-dimensional banded precision matrices
B. An, J. Guo, and Y. Liu

Convergence of sample eigenvalues, eigenvectors, and principal component scores for ultra-high dimensional data
S. Lee, F. Zou, and F. A. Wright

Sequential combination of weighted and nonparametric bagging for classification
M. Soleymani and S. M. S. Lee

Multiscale variance stabilization via maximum likelihood
G. P. Nason



We are very pleased to welcome the following new members!

Member		Country
Mr.	Arouna Anjuenneya Njoya	Burkina Faso
Mrs	Julia Aru	Estonia
Dr.	Signe Balina	Latvia
Mr.	Tom Caplan	Israel
Dr.	Nancy Chin	Italy
Dr.	Mahmoud Elkasabi	United States
Dr.	Parthasarathi Lahiri	United States
Dr.	Natalja Lepik	Estonia
Mr.	Kaur Lumiste	Estonia
Ms	Angela Luna Hernandez	United Kingdom
Mr.	Yusuf Sahin	Turkey
Dr.	Nikolaos Tzavidis	United Kingdom
Dr.	Berislav Zmuk	Croatia

IASS Officers and Council Members

Executive Officers

President:	Danny Pfeffermann (Israel)	msdanny@huji.ac.il
President-elect:	Steve Heeringa (USA)	sheering@isr.umich.edu
Vice-Presidents:	Jairo Mabil Oka Arrow (South Africa)	jairo.arrow@gmail.com
	Geoffrey Lee (Australia)	geoff.lee99@bigpond.com
Scientific Secretary:	Mick Couper (USA)	mcouper@umich.edu
Council Members (2011-2015):	Christine Bycroft (New Zealand)	christine.bycroft@stats.govt.nz
	Ka-Lin Karen Chan (China)	klchan@censtatd.gov.hk
	Olivier Dupriez (Belgium/USA)	odupriez@worldbank.org
	Natalie Shlomo (UK)	natalie.shlomo@manchester.ac.uk
	Marcel de Toledo Vieira (Brazil)	marcel.vieira@ufjf.edu.br
	Alvaro Gonzalez Villalobos (Argentina)	alvarun@gmail.com
Council Members: (2013-2017)	J. Michael Brick (USA)	mikebrick@westat.com
	Daniela Cocchi (Italy)	daniela.cocchi@unibo.it
	Jack Gambino (Canada)	gambino@statcan.ca
	Risto Lehtonen (Finland)	risto.lehtonen@helsinki.fi
	Ralf Münnich (Germany)	muennich@uni-trier.de
	Jean Opsomer (USA)	jopsomer@stat.colostate.edu

Committee Chairs

Chair of the Rio 2015 Programme Committee	Christine Bycroft (New Zealand)	christine.bycroft@stats.govt.nz
Chair of the committee for the Cochran Hansen prize award	Risto Lehtonen (Finland)	risto.lehtonen@helsinki.fi

Ex Officio Members

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

Institutional Members

2 International Organisations

AFRISTAT
EUROSTAT

15 Bureaus of Statistics

AUSTRALIA – AUSTRALIAN BUREAU OF STATISTICS
BRAZIL – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE)
CANADA – STATISTICS CANADA
CHINA – DIREÇCÃO DOS SERVIÇOS DE ESTATÍSTICA E CENSOS
DENMARK –DANMARKS STATISTIK
FINLAND – STATISTICS FINLAND
GERMANY –STATISTICHE BUNDESAMT
ITALY –INSTITUTO NAZIONALE DI STATISTICA (ISTAT)
KOREA, REPUBLIC OF – STATISTICS KOREA
MEXICO –INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (INEGI)
MAURITIUS – STATISTICS MAURITIUS
NEW ZEALAND – STATISTICS NEW ZEALAND
NORWAY – STATISTICS NORWAY
PORTUGAL –INSTITUTO NACIONAL DE ESTADÍSTICA (INE)
SWEDEN – STATISTISKA CENTRALBYRÅN

5 Universities, Research Centers, Private Statistics Firms

USA – CENTERS FOR DISEASE CONTROL AND PREVENTION
USA – RESEARCH TRIANGLE INSTITUTE
USA – SURVEY RESEARCH CENTER, UNIVERSITY OF MICHIGAN
USA – U.S. DEPARTMENT OF AGRICULTURE
USA – WESTAT

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/