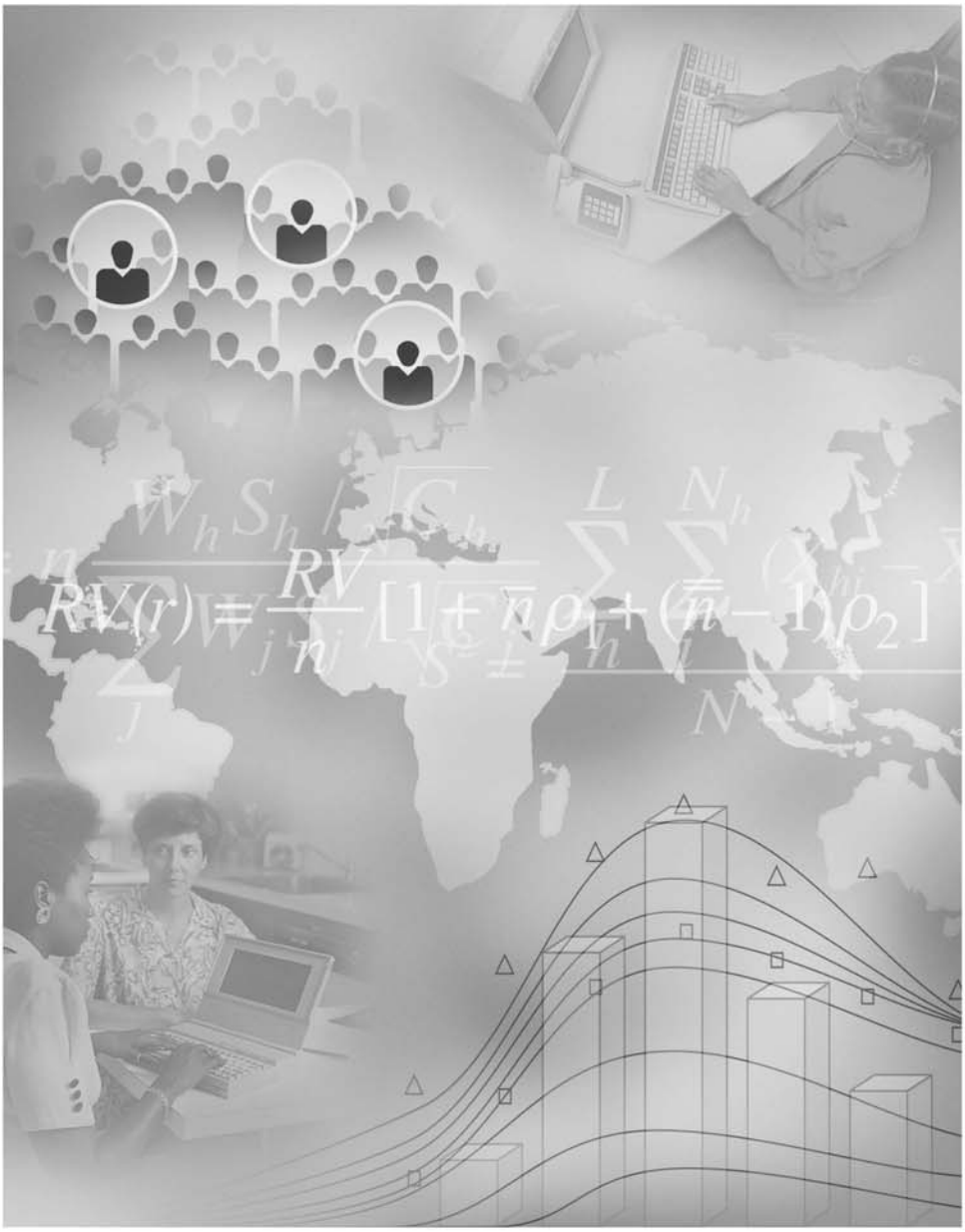
the Survey Statistician

The Newsletter of the International Association of Survey Statisticians

No. 62 **July 2010**



INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

International Statistical Institute

ASSOCIATION INTERNATIONALE DES STATISTICIENS D'ENQUETES

Institut International de Statistique

the Survey Statistician
The Newsletter of the International Association of Survey Statisticians
No. 62 July 2010

Editors
Natalie Shlomo and Frank Yu

Section Editors
Pierre Lavallée— Country Reports
John Eltinge — Book/Software Review
Robert Clark —Ask the Experts

Circulation
Évelyne Coutant
Carolyn Cruden
Tracey Wickens

The Survey Statistician is published twice a year in English and French by the International Association of Survey Statisticians and distributed to all its members. Information for membership in the Association or change of address for current members should be addressed to:

Secrétariat de l'AISE/IASS
c/o INSEE-CEFIL
Att. Mme Évelyne Coutant
3, rue de la Cité
33500 Libourne - FRANCE
E-mail: evelyne.coutant@insee.fr

Comments on the contents or suggestions for articles in *The Survey Statistician* should be sent via e-mail to frank.yu@abs.gov.au or mailed to:

Frank Yu
Australian Bureau of Statistics
Locked Bag 10
Belconnen ACT 2616
Australia

In This Issue

No. 62, July 2010

Letter from the President

Dear Colleagues

With just over a year to go, preparations for the 58th ISI World Statistics Congress in Dublin, 21 - 26 August 2011 are well underway. The programme of invited paper meetings sponsored by IASS is progressing well, thanks to Geoff Lee's work, and further information is provided later in this issue. Eric Rancourt (Eric.Rancourt@statcan.gc.ca) is Chair of the IASS Programme Subcommittee for the 59th ISI World Statistics Congress to be held in Hong Kong, and now is a great time to make suggestions for topics and potential participants for that programme.

Another very significant IASS activity connected with the ISI World Statistics Congress is the programme of IASS short courses to be offered in Dublin. This has been developed by the Scientific Secretary Leyla Mohadjer and is also progressing very well. Details of the courses are covered in her report below. As was the case in 2009, the IASS short courses will be part of a wider programme of short courses to be offered by ISI, and Steve Heeringa, who organised the IASS short courses in Durban has agreed to share his experience by co-ordinating the wider ISI program of courses. The short courses provide our members with substantial development opportunities, and I encourage you to have a look at the proposed programme and start thinking about your plans for Dublin.

As well as its involvement in the ISI congresses, the IASS supports its objective to promote the study and development of the theory and practice of sample surveys and censuses, by providing modest financial support to relevant international conferences on this topic. This year IASS Council are considering 5 requests for support related to such conferences being held in the period July 2010- June 2011.

The Council agreed last year to provide financial support to the 6th Francophone Conference on Surveys held in March 2010. In addition to the modest support provided from the IASS budget, the IASS was able to direct World Bank funding allocated to the ISI, to support the attendance of 16 participants from developing countries at the Conference.

The Council considered a report on services to members at its meeting in Durban last year, and some progress is being made on implementing its recommendations. A new team, jointly headed by Natalie Shlomo (n.shlomo@soton.ac.uk) and Frank Yu (frank.yu@abs.gov.au), is now editing *The Survey Statistician*, and will be looking to implement some fresh ideas. This is the first issue for the new team, and I would like to thank them for taking on this role. We have decided to move to electronic distribution for those members where this is possible. In the first instance this will simply be the electronic distribution of a publication designed for hard copy, but in time we hope to move onto an electronic design.

We now also have access to an electronic mail out facility to members. This facilitates electronic distribution of the Survey Statistician, but also provides the opportunity for other mail out deliveries of material that may be of interest to members. Council will develop a policy on the use of this mail out facility. Clearly only members who have provided their email address can be contacted electronically, so please provide these details to the IASS Office if you have not already done so, and remember to update them if they change.

To help us further progress improvements to the functionality of the website, a small group of IASS members has agreed to form a web team focusing on maintaining and developing the IASS web site. The team consists of Monica Pratesi (leader and co-ordinator), Mick Couper, Eric Rancourt and Krishna Palipudi (webmaster). An important early step for the team will be to consider how the IASS web site should fit with the redeveloped ISI web site. The team will report to the IASS council on current functionality, and proposed easy improvements in functionality, as well as longer term directions. I am very grateful to Monica, Mick, Eric and Krishna for agreeing to take on this job, as the web is a central means of providing services to those members who are not able to attend the biennial ISI meetings. If you are interested in being involved in this work, please contact Monica Pratesi (m.pratesi@ec.unipi.it) or myself.

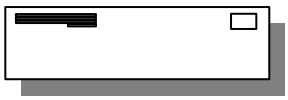
An ongoing challenge for the IASS is to keep up a network of Country Representatives, and encourage the sharing of information across countries through the use of Country Reports. A number of countries are missing a Country Representative and if you would like to become involved in this activity, please contact Linda Hewitt (linhew@trinidad.net).

The IASS is always keen to welcome new members both individual and institutional. The registration form for individual members can be found on the website. For those wishing to enquire about institutional membership, please contact the Executive Director, Catherine Meunier (catherine.meunier@insee.fr).

Finally I would like to note that Claude Olivier, who has been supporting the IASS Office for a number of years now, has recently retired. We thank her for all her help over the years and wish her well in her retirement. Her place will be taken by Évelyne Coutant and we warmly welcome Évelyne to the team.

Susan Linacre

President IASS



Letter from the Editors

The July 2010 IASS Newsletter, *The Survey Statistician*, is our first issue as editors. We would like to thank Victoria Leaver and Bianca Baldwin of the Australian Bureau of Statistics for their invaluable assistance in putting together this newsletter. We also would like to thank the previous editor, Dan Hedlin, for all of his hard work and efforts the last few years and for ensuring this smooth transition. Finally, we would like to thank the IASS President, Susan Linacre, for her support, the IASS team: Catherine Meunier and Évelyne Coutant, and Anne Patry of Statistics Canada for the translations into French.

The Survey Statistician IASS newsletter will continue with our regular sections and articles, including the President's address, all relevant information to our membership, and news updates for the Dublin ISI 2011 Conference, including important messages on IASS short course planning by the Scientific Secretary. We continue to provide advertising for upcoming conferences, workshops and job advertisements (depending on the correct timing of the issue to coincide with the closing date). The table of contents of journals are also featured with a broader selection of journals.

The Country Report Section will continue to be a central feature of the IASS newsletter. We ask all country representatives to submit articles and share information regarding current activities and practices in survey methods and research. All articles should be sent to the editor of this section, Pierre Lavallée at Pierre.Lavallee@statcan.gc.ca. We encourage all countries to participate in this section of the newsletter.

What we hope to add to the current structure of the IASS newsletter is more scientific content. We have introduced a new section called 'New and Emerging Methods' which will be edited by Leyla Mohadjer and Andrea Piesse. This issue of *The Survey Statistician* includes an article by Danny Pfeffermann titled: 'Small Area Estimation, Basic Concepts, Models and Ongoing Research'. We wish to thank Danny very much for his important contribution to this issue. Please let Leyla know if you would like to contribute to this new section in future IASS newsletters, starting with July 2011, at MohadJL1@westat.com.

We are also happy to recommence the 'Ask the Expert' column edited by Robert Clark at rclark@uow.edu.au. This issue of *The Survey Statistician* includes an article by David Steel and Robert Clark titled: 'Combining Survey Results'. We wish to thank David and Robert very much for their important contribution to this issue. Please send your questions to Robert on topics related to survey methods and research and we will aim to provide the answer in the next IASS newsletter.

Finally, we are also hoping to re-activate the section on 'Books and Software Review' which will be edited by John Eltinge: Eltinge.John@bls.gov. With many recent books published on survey methods recently, we feel it is of great importance to provide our membership with a review of books. We ask for volunteers to get in touch with John and submit book reviews. As well, we would also like to have a review of software for survey research, and in particular the use of the R-package. Recent European projects have put a lot of emphasis on R codes for survey methods and we are keen to make these new R libraries accessible to our membership.

In the future, we aim to coordinate the content of the newsletter with that of the website, and provide online dissemination which would enable more timely job advertisements and conference announcements.

So enjoy the new edition of *The Survey Statistician* and please send us your feedback and comment on how we can make improvements. Remember, the IASS newsletter depends on your commitment and volunteering to contribute to the contents. The more contribution we could get from you, the more the newsletter will interest all our membership. Keep the e-mails coming in and in particular, provide us with your input for the next, January 2011 issue.

Report from the Scientific Secretary

April 1 was the deadline for submission of proposals for the 2011 Short Courses in Dublin. I like to thank all members who submitted proposals for their contribution and for the time and effort they took to prepare the material for the submission. We received a rather large number of proposals. The Executive Committee reviewed the proposals and provided feedback regarding the appropriateness of the proposals as IASS Short Courses. The intent was to create a set of courses that are relevant to IASS, as balanced as possible across topics, appealing to a rather large number of IASS and ISI members usually attending these courses, and at the technical level that is appropriate for these short courses. In addition, we consulted with the Local Committee to seek their advice on the choice of course topics. We had to limit the number of selected courses to the maximum number allowable given the operational feasibility of carrying out these courses in Dublin. As a result, we were unable to accept a number of very good proposals. Proposals (not selected for IASS) that had relevance to official statistics were forwarded to the IAOS section.

The following is a description of the Short Courses that IASS will offer prior to the start of the 58th session of the ISI in Dublin.

1. Introduction to Survey Sampling

Instructors: Steven Heeringa, University of Michigan Institute for Social Research

Draft Description: The workshop is intended to provide an overview of principles of sample design selection and estimation. It will start from basic principles of sample design and selection beginning with simple random sampling, and build up to complex stratified multi-stage sample designs. It will cover the main sampling techniques and also discuss such issues as sampling frames and weighting. An introduction to variance estimation for complex sample designs will be presented at the end of the workshop.

Duration: Two days.

2. Analysis of Complex Sample Survey Data

Instructors: Jay Breidt, Colorado State University; Kirk Wolter, NORC.

Draft Description: Estimation procedures appropriate for data collected under complex survey designs will be discussed. The first part of the course will cover estimation and variance estimation for standard statistics, such as means, ratios, domain totals, and the entries in two-way tables. The use of survey data for the estimation of the parameters of statistical models is the focus of the second part of the course. Emphasis will be placed on efficient estimation of the parameters of regression models.

Duration: Two days.

3. Introduction to Survey Quality

Instructors: Paul Biemer, RTI International and the University of North Carolina—Chapel Hill; Lars Lyberg, Statistics Sweden.

Draft Description: The course spans a range of topics dealing with the quality of survey data. It begins with a discussion of dimensions of survey quality which include accuracy, relevance, timeliness, accessibility, and comparability. We describe an approach to maximizing survey quality in which total survey error is minimized subject to constraints on costs after accommodating the other quality dimensions. The major components of total survey error include: nonresponse error, frame error, measurement error, specification error, data processing error, and file preparation error. We describe a set of principles for evaluating these major error sources and for deploying survey resources optimally to reduce their effects on survey estimates – an approach embodied in the so-called total survey error paradigm. The TSE paradigm considers the origins of each error source (i.e., its root causes) and applies the most effective methods for reducing or controlling the errors under costs and quality constraints. Methods for evaluating survey error such as cognitive interviewing, pretesting, behavior coding, re-interview surveys and administrative records checks are also covered. The course reviews well-established as well as recently developed methods and concepts in the field. It also examines important issues that are still unresolved today and which are being actively pursued in the current survey methods literature. The course concludes with a discussion of the practical advice for designing and conducting surveys that consistently achieve near optimal levels of survey quality.

Duration: Two days.

4. Web Survey Design

Instructor: Mick P. Couper, University of Michigan Institute for Social Research, and the Joint Program in Survey Methodology at the University of Maryland.

Draft Description: The course will focus on the design of web survey instruments and procedures based on theories of human-computer interaction, interface design, and research on self-administered questionnaires and computer assisted interviewing. The course will cover various aspects of instrument design for Web surveys, including the appropriate use of widgets (e.g., radio buttons, check boxes)

for Web surveys, general formatting and layout issues, movement through the instrument (action buttons, navigation, error messages), and so on. The course will draw on empirical results from experiments on alternative design approaches as well as practical experience in the design and implementation of web surveys. The course will not address the technical aspects of web survey implementation, such as hardware, software or programming. The course will also not focus on question wording or sampling issues for Web surveys. The course will have a strong practical emphasis, examining many different examples of good and bad design, with recommendations for best practice.

Duration: One day.

5. Methods for Longitudinal Surveys

Instructor: Peter Lynn, Institute for Social and Economic Research (ISER) University of Essex

Draft Description: Introduce participants to issues which are specific to longitudinal and panel surveys, including topics in sample design, survey design, instrument design, weighting and imputation. The course would give an overview of the strengths and weaknesses of longitudinal surveys and an outline of key considerations in the design and implementation of such surveys.

Duration: one/ two days.

6. Workshop on Editing and Imputation of Survey Data

Instructors: John Kovar, Eric Rancourt, and Jean-Francois Beaumont, Statistics Canada.

Draft Description: Surveys and censuses conducted by national statistical agencies, research institutes and other survey organizations suffer from various degrees of nonresponse even under ideal conditions. In order to try to alleviate the problems caused by nonresponse, editing and imputation methods are usually applied. Since the process of editing and imputation is time and resource intensive, care must be exercised in controlling the efficiency as well as the effectiveness of the methods. The aim of this short course is to introduce the students to methods of nonresponse prevention and the treatment of suspicious, inconsistent and missing responses. Evaluation of such methods and their impact on the survey outputs will be highlighted. Examples will be provided to illustrate the material presented.

Duration: One and half days.

7. Business Survey Methods

Instructors: Mike Hidiroglou, and Wesley Yung, Statistics Canada

Draft Description: Business surveys are important sources of information for producing key economic indicators that monitor the economy over time and for constructing official statistics such as national accounts. While business surveys typically use simple sample designs they are not without their methodological challenges such as highly skewed and unstable populations, the quality of frame information and auxiliary data used in stratification, editing, imputation and estimation. This workshop will describe methods for designing business surveys and

will cover topics such as building and maintaining a Business Register, sample design, data collection, outlier detection and treatment, imputing for total and/or partial non-response, weighting and estimation and use of administrative data.

Duration: Two days.

The next step is to work with the Local Organizers in Dublin to help find a suitable venue for the Short Courses, and establish a reasonable course fee and a course schedule.

Participation of members from developing countries is very dependent on financial assistance. We will start approaching a number of organizations to seek sponsorship. Please contact me if you have any suggestions about potential sponsors or related contacts.

For additional information or any questions, please contact Leyla Mohadjer at leylamohadjer@westat.com.

IASS Satellite Conference on Small Area Estimation

I am very pleased to announce that IASS has made plans to sponsor a satellite conference on Small Area Estimation that will take place in Trier, Germany, August 11-13, 2011. The plans for the conference are currently under development. The details will be announced as soon as they become available.

There are tentative plans to offer the following short course in conjunction with the conference in Trier.

Small Area Estimation: Methods, Applications and Practical Demonstration

Instructors: J. N. K. Rao, Carleton University; Isabel Molina, Universidad Carlos III de Madrid.

Draft Description: This workshop will provide an introduction to model-based small area estimation. A variety of recent applications to real socio-economic data and practical demonstration of the methods will also be presented. The course is primary aimed at methodologists in government statistical bureaus or survey organizations, statistics graduate students and faculty from universities and users with adequate background in linear regression models and survey sampling theory and methods.

Duration: Two days.

.

Cochran-Hansen Prize 2011

Competition for Young Survey Statisticians from Developing and Transitional Countries

In celebration of its 25th anniversary, the International Association of Survey Statisticians established the Cochran-Hansen Prize to be awarded every two years to the best paper on survey research methods submitted by a young statistician from a developing or transitional country.

Participation in the competition for the Prize is open to nationals of developing or transitional countries who are living in such countries and who were born in 1967 or later.

Papers submitted must be unpublished original works. They may include materials from the participant's university thesis. They should be in either English or French. Papers for consideration should be submitted to the IASS Secretariat at the address below to arrive by December, 2010. Each submission should be accompanied by a cover letter that gives the participant's year of birth, nationality, and country of residence. The cover letter must also indicate if the work submitted is the result of a PhD thesis and, in the case of joint papers, the prize candidate must state clearly what his/her contribution to the paper is.
The papers submitted will be examined by the Cochran-Hansen Prize Committee appointed by the IASS. The decision of the Committee is final.

The winner of the Prize will be invited to present his/her paper at the 58th Session of the International Statistical Institute to be held in Dublin, Ireland, August 21-26, 2011, and the name of the winner will be announced at the ISI General Assembly in Durban.

The author of the winning paper will receive the Cochran-Hansen Prize in the form of books and journal subscriptions to the value of about € 500, and will have reasonable travel and living expenses paid in order to present the paper at the ISI Session in Durban.

For further information, please contact:

Madame Évelyne Coutant
IASS Secretariat
INSEE/CEFIL, 3 rue de la Cité, 33500 Libourne, France
Tel: +33 5 5755 5607
Fax: +33 5 5755 5620
E-mail: evelyne.coutant@insee.fr

Best Regards,
Leyla Mohadjer
leylamohadjer@westat.com

Combining Survey Results

Question: "What is the best way of combining the results from two surveys, which ask some common questions, possibly on overlapping samples, to improve the variance of the estimates for the common items?"

Discussion: David G Steel and Robert G Clark

1 Introduction

We will call the two surveys A and B, and assume that they are selected from sampling frames that cover the same population. If not, then dual frame methods are appropriate. It will be assumed for now that the aim is to estimate totals for the common questions, although it's possible that there are some non-common questions which are on A or B but not both.

We will assume throughout that the idea is to estimate totals, although the weights that we obtain could be used to calculate other weighted statistics, including means, ratios, quantiles and regression coefficients.

Section 2 will discuss some examples of this question. There are two main strategies for combining A and B: combining estimates and pooling samples. These will be discussed in Sections 3 and 4 respectively. Section 5 will cover a few other issues: use of auxiliary variables; dual frame methods; and estimation for non-common variables. A selective bibliography will also be given.

2 Some Examples

To set the scene, consider a couple of examples:

(a) Two opinion polls are conducted independently, at around the same time, with some questions in common but not others. The aim would then be to combine the two surveys to get the best estimates for the common key questions.

(b) A national statistics office might conduct many social surveys on different topics. Some common questions could be asked in each, for example household income (at least in a range), employment status, household composition or ethnicity. These surveys could be combined to get better estimates of the common items.

(c) A survey may consist of a core and a booster sample, where the booster sample is designed to increase the number of members of a subpopulation of interest. The question is then how to combine the core and booster sample to get best estimates for the subpopulation and for the overall population. This design has been used in New Zealand to over-sample the Maori population in health surveys (Wells, 1998). The core and booster samples are usually constructed to be non-overlapping.

(d) A so-called "split questionnaire design" could be used (Chipperfield and Steel, 2009). In this design, the sample is divided up into random subsamples, and different but overlapping question sets are asked for each subset. This design could be applied to exploit correlations between data items, while reducing respondent load by not asking any respondent the full set of question. The main aim would be to improve estimates for the non-common data items.
We do not consider the following examples to be special cases of the question, although they may appear so at first sight:

(e) A monthly (or quarterly) rotating panel survey may measure labour force status, as in many countries. The surveys from two different months measure the same concept, but with different reference periods, and so cannot be considered to have common items.

(f) A sample of the general population is taken from an area frame. Another sample is taken from a partial list of a sub-population of interest (e.g. recent migrants from a government list, or sufferers of a medical condition from hospital clinic lists). This is an example of a dual frame problem.

3 Combining Estimates

The idea here is that we obtain separate estimates, $\hat{Y}_A$ and $\hat{Y}_B$, of the total of a variable of interest from survey A and survey B. We can calculate a combined estimate using:

$$\hat{Y}_{combined} = \theta\hat{Y}_A + (1-\theta)\hat{Y}_B$$

where θ is between 0 and 1. In other words we take a weighted average of the estimates from the two surveys. It is assumed that $\hat{Y}_A$ and $\hat{Y}_B$ have the same expectation, i.e. that measurement biases are the same in both. In practice this may not be the case, in which case varying θ will vary the expectation of $\hat{Y}_{combined}$.
O'Muircheartaigh (2002) suggested

$$(1) \qquad \theta = \frac{n_A / deff_A}{n_A / deff_A + n_B / deff_B}$$

where $deff_A$ and $deff_B$ are the design effects for estimates of Y from surveys A and B, respectively. If $\hat{Y}_A$ and $\hat{Y}_B$ are independent, this is the optimal choice of θ. If the design effects are equal, then this implies just weighting the estimates according to their relative sample sizes. However, if the surveys have very different designs, the deffs may differ substantially.

Suppose that $\hat{Y}_A$ and $\hat{Y}_B$ are both weighted estimates, with $\hat{Y}_A = \sum_{s_A} w_{Ai} y_i$ and $\hat{Y}_B = \sum_{s_B} w_{Bi} y_i$, where s_A and s_B are the samples from surveys A and B, respectively. Let s_P be the pooled sample from combining s_A and s_B, with units in both samples being counted twice. Let $s_{A \cup B}$ be the pooled sample where units in both samples are counted only once. We can write the combined estimate in weighted form as

$$\hat{Y}_{combined} = \sum_{s_P} w_{combined,i} y_i$$

where

$$w_{combined,i} = \begin{cases} \theta w_{Ai} & \text{if} \quad i \in s_A \\ (1-\theta) w_{Bi} & \text{if} \quad i \in s_B \end{cases}. \tag{2}$$

However, the design effects depend, in general, on the variable of interest, so that θ and hence $w_{combined,i}$ will differ from variable to variable. So we are in the unfortunate possibility of needing a different set of weights for each variable. In practice, a single set of weights would be used, for example based on the design effects for the most important variable (in which case we would have to accept some inefficiency for other variables). In some cases, the design effect can be approximated by 1 plus the relative variance of the survey weights, which does not depend on the variable of interest, so we avoid the problem in this case.

This method can be applied when the samples for A and B may overlap, without knowledge of which units are in both samples.

4 Pooling Samples

If we are able to identify which respondents are in both samples, then we can define $s_{A \cup B}$ to be our sample, and can calculate probabilities of selection using:

$$\begin{aligned} \pi_{Pi} &= P\left[i \in s_{A \cup B}\right] = P\left[i \in s_A\right] + P\left[i \in s_B\right] - P\left[i \in s_B \cap s_B\right] \\ &= \pi_{iA} + \pi_{iB} - P\left[i \in s_B \cap s_B\right] \end{aligned}$$

If the two samples are independent, then $P\left[i \in s_B \cap s_B\right] = \pi_{iA}\pi_{iB}$ and so

$$\pi_{Pi} = P\left[i \in s_{A \cup B}\right] = \pi_{iA} + \pi_{iB} - \pi_{iA}\pi_{iB}.$$

Alternatively, the two samples may have been selected to force them to be non-overlapping, in which case $P\left[i \in s_B \cap s_B\right] = 0$. In practice it won't make much difference if probabilities of selection are small, as $\pi_{iA}\pi_{iB}$ will be negligible anyway.

Suppose that we are going to use the inverses of the probabilities of selection as weights. In this case, our estimator will be a weighted sum over the pooled sample:

$$\hat{Y}_{pooled} = \sum_{s_{A \cup B}} w_{pooled,i} y_i$$

where

$$w_{pooled,i} = \left(\pi_{iA} + \pi_{iB} - \pi_{iA}\pi_{iB}\right)^{-1}$$
$$= \left(w_{iA}^{-1} + w_{iB}^{-1} - w_{iA}^{-1}w_{iB}^{-1}\right)^{-1} \approx \left(w_{iA}^{-1} + w_{iB}^{-1}\right)^{-1} \quad (3)$$

Note that the weights do not depend on the particular variable of interest being considered, which is convenient for surveys collecting many variables. If we are not able to identify which units overlap, we would redefine π_{pi} to be the expected number of times that a unit appears in s_P (remembering that units appear twice, if they are selected in both surveys), and calculate π_{pi} using:

$$\pi_{Pi} = \pi_{iA} + \pi_{iB}.$$

Our estimate would then be given by $\sum_{s_P}\left(\pi_{iA} + \pi_{iB}\right)^{-1} y_i$.

In practice it may make little difference whether we use $s_{A\cup B}$ (where common units are counted only once) or s_P (where common units are counted twice), since there would often be few or no units selected in both A and B. However, if sampling fractions are large, then π_{pi} may be non-negligible.

Which is Better: Combining Estimates or Pooling Samples?

O'Muircheartaigh (2002) showed that pooling samples results in less variable weights than combining estimates. It's easy to see why. If surveys A and B have quite different probabilities of selection, this will mean that the combined weights in (2) will vary quite a lot. An appropriate choice of θ will reduce but not remove the problem that the weight may be quite different depending on whether a unit was selected in survey A or B. In contrast, the weights in (3) are based on the probabilities of selection in both surveys, regardless of whether the respondent happened to have been recruited via survey A or B.

It might seem as if the combining estimates method must perform better if θ is calculated by (1), since this is the optimal linear combination of $\hat{Y}_A$ and $\hat{Y}_B$. However, $\hat{Y}_{pooled}$ is not restricted to a linear combination of $\hat{Y}_A$ and $\hat{Y}_B$, and so it can (and apparently, usually does) do better than $\hat{Y}_{combined}$. This is the case even if there is only one variable of interest, and the case for $\hat{Y}_{pooled}$ is even stronger when there are multiple variables of interest, since the optimal θ cannot be used for every variable.

One possible barrier to pooling the samples is that we need to be able to calculate $P[i \in A]$ and $P[i \in B]$ for all respondents in both samples. In other words, we need to know $P[i \in A]$ for the respondents to survey B and vice-versa. This is sometimes a problem in example (c), because it requires that the screening questions be applied to both the core and booster samples (Wells, 1998). Lohr and Rao (2000; section 2.3) also pointed out this difficulty in a dual frame context.

5 Related Problems

5.1 Auxiliary Variables

So far we have assumed that weights will be based on the probabilities of selection. If some auxiliary variables are available, these could be used to improve estimation. One possibility would be to use the weights (2) or (3) as the initial weights in a calibration to population benchmarks.

Renssen et al. (1997) considered weighted combinations of $\hat{Y}_A$ and $\hat{Y}_B$. When there are multiple common variables, $\hat{Y}_A$ and $\hat{Y}_B$ were defined to be vector-values, and the combined estimator was given by

$$\hat{Y}_{combined} = M\hat{Y}_A + (I - M)\hat{Y}_B$$

where M is a matrix. They considered setting M to be proportional to the identity matrix, and also derived the optimal M, which is not, in general, a diagonal matrix.

Merkouris(2004) compares a number of alternative methods for calculating generalized regression estimates (GREGs) of common variables from two independent surveys A and B. One method is to calculate a GREG for A and another for B, and combine the two using

$$\hat{Y}_{combined} = \theta\hat{Y}_A + (1 - \theta)\hat{Y}_B$$

similar to the combining estimates method in Section 3, but using GREG estimators $\hat{Y}_A$ and $\hat{Y}_B$ rather than inverse probability estimators. An alternative method is to calculate a GREG estimator using the pooled sample. Merkouris also considered the case where different auxiliary variables are available for the two surveys.

5.2 The Dual Frame Problem

This is the closely related problem, where we have two surveys A and B with some common data items, where A and B are selected from different frames with different coverage. It is usually assumed that the overlap between surveys A and B cannot be identified. Lohr and Rao (2000) partition the population of interest into three subsets: *a* (in scope of A but not B); *b* (in scope of B and not A); and *ab* (in scope of both surveys). In this case, the population total for *a* must be estimated from survey A only, while that for *b* must be estimated from survey B only. The total for *ab* can be estimated using both surveys, and this problem is essentially the same as in Section 3 of our discussion.

They considered a number of possible strategies. The dual frame problem can be attacked by combining estimates, or by pooling samples, as for our case when *a* and *b* are empty. One method of combining estimates which performed well was the pseudo-maximum-likelihood method of Skinner and Rao (1996) – this approach has the advantage that the compositing factor does not depend on the variable of interest. Lohr and Rao found that pooling samples was generally not the best method, in apparent contradiction to O'Muircheartaigh (2002), although the situation is not identical when a and b are not empty.

5.3 Non-Common Variables

Suppose that there are some variables collected in A but not B. In this case, these non-common variables are observed just in s_A, whereas the common variables are observed for the larger sample $s_{A\cup B}$. If the non-common variables are correlated with the common variables, then we can use the data on the common variables to improve estimation of the totals of the non-common variables. One way of doing this is treating $s_{A\cup B}$ as the first phase sample and s_A as the second phase sample, and then using two-phase estimation methods such as the two-phase ratio or regression estimator.

Renssen et al. (1997) and Merkouris (2004) both considered estimation of non-common variables from A or B. They used the common variables as auxiliary variables, and benchmarked the estimates for the non-common variables using estimated totals of the common variables calculated using both A and B.

Chipperfield and Steel(2009) discuss this split questionnaire designs, where different groups of respondents are given different question sets, in order to improve efficiency while managing respondent burden.

6 Selective Bibliography

Chipperfield, James O. and Steel, David G. (2009). Design and estimation for split questionnaire surveys. *Journal of Official Statistics*, 25, 227-244

Lohr, Sharon L. and Rao, J. N. K. (2000). Inference from dual frame surveys. *Journal of the American Statistical Association*, 95, 271-280

Merkouris, Takis (2004). Combining independent regression estimators from multiple surveys. *Journal of the American Statistical Association*, 99, 1131-1139.

O'Muircheartaigh, Colm and Pedlow, Steven (2002). Combining samples vs. cumulating cases: a comparison of two weighting strategies in NLSY97. *ASA Proceedings of the Joint Statistical Meetings*, pp. 2557-2562. American Statistical Association (Alexandria, VA)

Renssen, Robbert H. and Nieuwenbroek, Nico J. (1997). Aligning estimates for common variables in two or more sample surveys. *Journal of the American Statistical Association*, 92, 368-374

Skinner, C. J. and Rao, J. N. K. (1996). Estimation in dual frame surveys with complex designs. *Journal of the American Statistical Association*, 91, 349-356

Wells, J. Elisabeth (1998). Oversampling through households or other clusters: Comparisons of methods for weighting the oversampled elements. *Australian & New Zealand Journal of Statistics*, 40, 269-277

Ask the Experts - Call for Questions

If you'd like to ask the experts a question, please contact Robert Clark rclark@uow.edu.au.

Australia

SOCIAL DISORDER

The Australian National Crime Victimisation Survey (NCVS) has been run as a modular, annual telephone-based survey since 2008-09, which allows alterations to the content of the survey to respond to user needs as they emerge without undue disturbance of existing time series for other offences and topics covered. User consultations highlighted the increased attention in policy, research and the media on what was variously termed 'incivility', 'anti-social behaviour' or 'social disorder' within the community. A number of jurisdictions were attempting to develop performance measures and indicators around these concepts, but found they were not well covered in the existing NCVS or other sources. The existing NCVS content covered only simple information on problems in the neighbourhood, which provided insufficient coverage of the topic. Consequently, the Australian Bureau of Statistics (ABS) commenced development on a new survey module to collect data on social disorder.

Social disorder information needs were understood to be for information about the physical and/or social signs of disorder in the community, with a focus on aspects most likely to lead to a fear of crime and a reduction in the quality of life of residents. The assumptions underpinning this approach were that obvious manifestations of social disorder can potentially have a greater negative effect on community members than actual criminal events. The research literature suggests that social disorder can relate to both criminal and non-criminal events. ABS reviewed international directions in collecting information about social disorder and anti-social behaviour, with a particular focus on the work done by the Home Office in recent years for the British Crime Survey (BCS). Elements which seemed most relevant to the Australian situation and which also reflected policy-maker concerns were adopted, and edited to fit within the constraints of the small survey module.

The first social-disorder module is currently in the field as part of the 2009-10 NCVS, and covers aspects such as respondent experience of noisy or dangerous driving, intimidation and harassment in the street, public drunkenness, rowdy and offensive behaviour, graffiti, people hanging around in groups and drug use in the local area. The module identifies the main problem noted, the perceived size of the problem, source of the perception of the problem, and location where the social disorder occurred. ABS may further expand this module in the future to gain greater insight into manifestations of social disorder and its effects on the community.

For further information please contact Sam McNally, Australian Bureau of Statistics Branch on +61 3 9615 7748 or email: sam.mcnally@abs.gov.au

Brazil

Marcel de Toledo Vieira

Planning for the **2010 Brazilian Population Census** started in 2007 and data collection will start on August 1st, 2010, and the first results will be released by the end of the same year by the **Brazilian Institute of Geography and Statistics (IBGE)**. More detailed results will be presented during the year 2011. The Census operation will therefore take place all over the country, with the objective of finding out who, how many and where the Brazilians are and also how they live, that is, it will provide a full-sized portrait of Brazil.

Municipal, state and federal **planning authorities** depend to a great extent on censuses to define policies which require updated information about the population in areas under their jurisdiction. Nevertheless, the demand for disaggregated information also comes from other segments, for example, the non-governmental and the private sectors.

The use of **GPS** in data collection equipment will allow the georeferencing of several physical elements of the land, expanding not only the range of possible results, but also their later use by the government and by society.

The **census operation** will involve about 230 thousand persons all over the country, including those hired on a temporary basis and servants from the permanent staff of IBGE. In order to fulfill this enormous demand, professionals who will take part in the 2010 Census have had specific training for each function.

A **Census** is an enterprise that needs a significant amount of resources in any country, mainly in one of continental dimensions such as Brazil, with about 8 million km^2 of a heterogeneous and sometimes inaccessible territory. The expected **budget** of the 2010 Census is approximately 1.7 Billion Brazilian Reals (equivalent to approximately 700 Million Euros), a figure which refers to census related activities in the period 2009-2011.

IBGE representatives have been working in the sphere of the United Nations Organization (UNO), the Statistical Office of the European Communities (Eurostat) and the Organisation for Economic Co-operation and Development (OECD), with partners of the Southern Common Market (Mercosur) and representatives of official bureaus of statistics of several countries. All these efforts were aimed at the exchange and acquisition of experiences concerning new demands for information, methodologies and technologies.

Making the whole **Brazilian society** aware of the importance of welcoming enumerators and correctly answering the questionnaire has been one of the greatest challenges of IBGE in the Census. Because of that, the institution has started to mobilize the population long before the collection phase.

Most information presented in this report have been extracted from a document prepared by IBGE, which is entitled "**CENSO 2010 – Summary of Survey Steps**" (IBGE, 2010). For further information regarding the 2010 Brazilian Census, please contact Sonia Albieri at salbieri@ibge.gov.br, or Antônio José Ribeiro Dias at vermelho@ibge.gov.br, or visit www.ibge.gov.br/english/ .

Canada

Advisory Committee on Statistical Methods

In April 2010, Statistics Canada will hold the 50th meeting of its Advisory Committee on Statistical Methods. The Committee was created in 1985 to advise the Chief Statistician on matters relating to the utilization of efficient statistical methods in the agency's program and on its program of research and development in statistical methods. Members come from Canadian and foreign universities, other statistical agencies and research agencies in the survey field. Since 1991, the Committee has been chaired by Prof. Wayne Fuller of Iowa State University. The Committee provides a forum for the review of methodology issues and directions. Its recommendations are addressed by the Statistical Methods Program, and follow-up actions are periodically reviewed. The Committee contributes to maintaining the quality of the Program by providing input to strategic directions relating to methodology issues that arise at the project planning and development stages, and by reviewing the methodologies proposed to ensure that they meet the challenges the agency is facing and that they are statistically appropriate.

Over its 25 years, the Committee has reviewed the main statistical programs and current topics at Statistics Canada. Every two to three years, the Committee has examined the methodological challenges and directions of the Census of Population and, more recently, of the Unified Enterprise Survey Program, the Household Survey Strategy and of individual business and household surveys, notably health surveys. It has also provided input to the development of policies, guidelines and management frameworks relating to quality. Meetings are held twice per year and minutes are distributed to the Provincial and Territorial Statistical Focal Points.

For more information, please contact Jean-Louis Tambay, jean-louis.tambay@statcan.gc.ca.

Developments in Time Series

SAS procedures developed by Statistics Canada for processing time series have been available in production version under the Forillon (V1.03.001) software since February 2, 2010. This system includes two procedures: *proc benchmarking*, which can benchmark a sub-annual series to control totals, and *proc tsraking*, which can restore additivity rules in a set of time series. The software is available for SAS 9.1.3 and SAS 9.2 under Windows and UNIX. A trial version and additional details can be obtained by contacting the development team at Forillon@statcan.gc.ca.

For the last few years, Statistics Canada has been modernizing its seasonal adjustment systems by migrating to the X-12-ARIMA Method. Seasonally adjusted data for several surveys, including the Monthly Wholesale and Retail Trade Survey, the Monthly Survey of Manufacturing and the International Travel Survey are prepared with the U.S. Census Bureau's X-12-ARIMA version. A system using the SAS 9.2 PROC X12 has been developed, and was first used for the publication of seasonally adjusted data from the January 2010 Labour Force Survey. This time series processing system also uses Forillon procedures. For additional information, please contact Michel Ferland at Michel.Ferland@statcan.gc.ca.

Germany

Ralf Münnich

National Educational Panel Study

In modern knowledge societies, education is the decisive precondition for not only participation in democracy but also economic growth and prosperity. The increasingly rapid changes in a globalized world are making it necessary to cope with new challenges in both private and working life. The National Educational Panel Study (NEPS) is currently being set up to face these challenges by finding out more about the acquisition of education in Germany, plotting the consequences of education for individual biographies, and describing as well as analyzing central education processes and trajectories across the entire lifespan. The interdisciplinary NEPS consortium combines the specific expertise of a number of research institutes, researcher groups, and research personalities under the management of Professor Hans-Peter Blossfeld at the Otto-Friedrich-University Bamberg. The project has been evaluated scientifically by the German Research Foundation, and is receiving both federal and state government support. It is being financed by the German Federal Ministry for Education and Research.

The NEPS distinguishes eight stages of education that are integrated and coordinated through a theoretical concentration on five interlinked dimensions. The dimensions are: (1) competence development in the life course; (2) education processes in learning environments; (3) social inequality and education decisions in the life course; (4) education acquisition with migration background in the life course; (5) returns to education in the life course. While these five dimensions form the main structure of the NEPS, their content will be examined throughout the lifespan with a particular focus on the following eight educational stages: (1) from birth to early childcare; (2) from kindergarten to elementary school; (3) from elementary school to lower secondary school; (4) from lower to upper secondary school or vocational training; (5) from upper secondary school to higher education, vocational training, or the labor market; (6) from the vocational education and training system to working life; (7) from higher education to the labor market; (8) adult education and lifelong learning. These theoretical dimensions and educational stages form the framing concept for the NEPS. In addition, there are method groups dealing with central questions like sampling, non-response, and mode effects.

Methodologically, the NEPS is based on a multicohort sequence design. Six starting cohorts – newborns, kindergarten children, 5th graders, 9th graders, university freshmen and adults – will be recruited between 2009 and 2012. These will contain a total of more than 60,000 participants who will be surveyed regularly over an extended period of time. Their competencies will also be assessed at set intervals. To document and analyze historical changes in the way people pass various transitions into, within, and out of the education system, new starting cohorts will also be recruited and integrated into the study in later years (creating a succession of cohorts). The data collected for the NEPS will be subjected to prompt and strict quality controls before being processed and documented in a user-friendly way. While complying strictly with personal data privacy requirements, this will grant researchers in Germany and abroad the opportunity to analyze the data as exhaustively as possible, thereby contributing to the greatest possible progress in education research.

The NEPS will deliver innovative impulses for basic research on developmental processes and trajectories. In the mid-term, the NEPS will also permit the study of

political reforms and their effects on, for example, the acquisition of competencies or equal opportunity in the education system. In sum, the NEPS is expected to decisively improve the framing conditions for empirical education research in Germany, provide an empirical basis for advising policymakers, make a major contribution to promoting the careers of young scientists, and lead to a marked improvement in the international standing of German education research.

For more information on the NEPS please visit our homepage at http://www.bildungspanel.de. Further information can also be obtained from the project team at contact.neps@uni-bamberg.de.

Hans-Peter Blossfeld, Jutta von Maurice, Thorsten Schneider & Susanne Rässler

Latvia

Martins Liberts

The refreshing of the design of Latvian Labour Force Survey (LFS) 2010 has been done. LFS is organised as continuous survey since 2002 in Latvia. The two stage sampling design is used for the survey. Census counting areas are selected as primary sampling units (PSU) with stratified systematic sampling with probabilities proportional to PSU size. Dwellings are selected as secondary sampling units with simple random sampling in each selected PSU. Households and individuals eligible to the LFS are selected in each sampled dwelling.

One of the main reasons of refreshing the design was necessity to update the sampling frame of primary sampling units. During the period 2002 – 2007 the 1st stage sampling frame was not updated. The over-coverage and under-coverage errors for census counting areas in sampling frame have emerged during this period. The size of census counting areas was measured as number of households from population census 2000. The size of PSUs has changed since 2000 because of the migration of population. Several improvements were done for the sampling frame of PSUs.

There were dwellings in sampling frame with no assignment to any census counting area. Typically these were dwellings built after the last population census. These dwellings contributed to the under-coverage of the frame. All those dwellings were assigned to the closest census counting area. The closest counting area was detected by measuring distance between a dwelling and all centres of census counting areas in stratum. Geographical coordinates of dwellings were used to compute the distances between dwellings and centres of counting areas. It was possible to use the geographic coordinates because of the matching of two administrative data sources – the population register and the register of buildings. Geographic coordinates were used for the first time to assign dwellings to census counting areas.

The size of all PSUs was updated using the latest information from the population register. Census counting areas without any occupied dwelling were deleted from the PSU frame. Dwellings with small number of private dwellings were merged to neighbour census counting areas. Some of the census counting areas have significantly increased by size. There were two counting areas each split in three smaller PSUs.

The PSUs were re-ordered in “serpentine” order in each stratum according to the latest reform of administrative territories in Latvia. Ordering of PSUs in “serpentine” order and using systematic sampling results in implicit stratification of sampling units by geographical breakdown.
New sample of PSUs was drawn using the updated sampling frame. The self-rotating sampling design was used. The goals set for the sampling design are:

- Rotation of PSUs according to the specific rotation pattern. The PSU selected in sample for the first time are kept in a sample for eight quarters – so each PSU is sampled eight times with intervals of 13 weeks. It is possible to draw different samples of dwellings in one PSU.
- Uniform distribution of areas over space. It provides possibility to get estimates by different geographical breakdowns.
- Uniform distribution of areas over time. It provides possibility to get estimates for different time periods – quarters, half-years and years. It is possible to shift an estimation period by any number of weeks. For example it is possible to estimate yearly statistics for time period starting in October and ending in September (shifted year estimates are common in some surveys).
- Easy management of PSUs in a sample. The sample of PSU can be prepared for several years in advance. It allows timely planning of the work for interviewers.
- The possibility to coordinate different continuous household samples. Currently there is a coordination of LFS, Household Budget Survey (HBS) and Survey of Domestic Travellers (SDT). The PSU samples of HBS and SDT are sub-samples of LFS PSU sample. One interviewer can manage to do all three surveys in a PSU with low travel times. The coordination allows to keep the total costs of three surveys low.
- The possibility to use the PSU sample for other ad-hock sample surveys.
- The possibility to use simple approximation methods for variance estimation (re-sampling techniques).

The rotating panel is used in LFS. The scheme 2-(2)-2 is used in Latvia (“Labour Force Survey in the EU, Candidate and EFTA countries - Main characteristics of the 2008 national surveys”, Eurostat, 2010). It means the part of the sample for year 2010 are already selected in 2008 and 2009. The total sample for 2010 are combination of sample selected before 2010 (old sample) and sample selected from 2010 (new sample). This was a challenge to select the new sample to minimize the overlap between old and new sample.

In a result we have made several improvements for Latvian Labour Force Survey and other coordinated surveys so that we can expect lower coverage errors and smaller sampling errors.

For further information please contact Martins Liberts, Central Statistical Bureau of Latvia (martins.liberts@csb.gov.lv or martins.liberts@gmail.com).

Malaysia

Aziz Mohammad

The Population and Housing Census of Malaysia 2010 will be conducted in July 2010. Its implementation was approved by Cabinet on April 22, 2009 with a total allocation of RM200.2 millions.

The enumeration period for the census is from 6 July to 22 August 2010, covering a period of two months. Census 2010 is set on Tuesday, July 6, 2010 which is assumed to have the minimum mobility of population.

This is the fifth census to be conducted since the formation of Malaysia in 1963, the earlier censuses were conducted in 1970, 1980, 1991 and 2000. Past censuses in Malaysia have traditionally employed the *de facto* approach in the enumeration of population. An attempt was made in the 1980 and 1991 Census to obtain a *de jure* count of the population in addition to the *de facto* approach. As a result, the *de jure* count yielded a higher figure than the *de facto* count that is about 2 per cent in the 1980 Census and 5% in 1991 Census. Similar to the 2000 Census, the *de jure* approach will be used for the 2010 Population and Housing Census.

Unlike previous censuses, the Department of Statistics Malaysia will implement a multi-modal approach in data collection, which comprises of face to face interview, self enumeration (drop-off and pick-up), and e-Census.

The enumeration will be a complete coverage of Malaysia and of all persons including non-citizens who had stayed or intended to stay in Malaysia for six months or more in the year 2010.

The data to be collected by the 2010 Census will cover three main categories, that is population (geographical, social, demographic, education and economic characteristics), household, and living quarters (including construction material of outer walls, occupancy status, no. of rooms and bedrooms, ownership status, drinking water supply, electricity supply, toilets and garbage collection facility).

For further information, please contact Ms. Norazlin Muharam or Ms. Hafizah Abdul Ghaffur at banci2010@stats.gov.my

New Zealand

Statistics New Zealand has developed a new process for designing enumeration areas for the New Zealand Census of Population and Dwellings. Approximately 6 500 enumeration areas are required to enable census collectors to deliver forms to each dwelling in the country. It is estimated that there will be 1.7 million dwellings and 4.5 million people in New Zealand at the time of the 2011 quinquennial Census.

In the past, the process to develop the enumeration areas has been clerical and resource intensive. There are approximately 45 000 units (meshblocks) at the lowest geographical level which need to be combined to form suitable enumeration areas for each census collector. Each enumeration area was the result of combining estimates of dwellings with an assessment of paper maps. Since changing one

enumeration area could have a large consequential effect on the other enumeration areas the process was difficult and produced many enumeration areas with disparate workloads and poor physical contiguity. Resource constraints limited the number of enumeration areas, which could be reviewed for changes, such as growth.

The new approach combines several sources of data and operational research techniques to produce a set of enumeration areas for the whole country. This required quantifying collector workload, specifying an ideal range of workload and developing measures of closeness for the smallest geographical units. A two-stage solution process was used to create the collection areas by region. A similar approach was used to create district areas, which are the next level of aggregation from collector areas.

The much shorter solution times allowed a range of options to be assessed and sensitivity analysis of the results. Collector workloads were kept within a narrower range, as well as producing collector areas, which were safer and more practical.

For further information, please contact John Créquer at John.Crequer@stats.govt.nz

Philippines

Gervacio G. Selda, Jr.

2010 Census of Population and Housing

The Philippine Statistical System (PSS) gears up preparations for the 2010 Census of Population and Housing (CPH) in the country and as part of the United Nation's 2010 World Population and Housing Census Programme. The National Statistical Coordination Board (NSCB), the policy-making body on all statistical matters in the country, urges utmost cooperation and support for the success of the 2010 Census of Population and Housing (CPH), which will be conducted by the National Statistics Office (NSO) from May 17 to June 11 2010, with May 1, 2010 as the reference date.

In accordance with Philippine laws on censuses, the decennial activity aims to get the inventory of total population and housing units by area in the country and collect information about their characteristics as basis for the formulation of rational plans and programs for national and local development. The 2010 CPH has been cleared by the National Statistical Coordination Board (NSCB) under its Statistical Survey Review and Clearance System (SSRCS), a mechanism instituted to: (1) ensure sound design for data collection; (2) minimize the burden placed upon respondents; (3) effect economy in statistical data collection; (4) achieve better coordination of government statistical activities, and (5) inform the private sector and the public in general that a government survey operation has passed clearance and enjoin their cooperation in the conduct of the survey.

The 2010 census will be the 13th census of population and the 6th census of housing since 1903. The following are the data/information to be generated: (1) size, distribution and composition of the population in terms of age, sex and marital status; (2) birth registration, disability, functional difficulty, literacy, school attendance, place of school, highest educational attainment, residence five years ago, overseas employment, usual occupation, kind of business or industry, class of worker, place of work, fertility, religion, citizenship, ethnicity and functional difficulty; (3) household characteristics, such as language/dialect generally spoken at home, presence of

household conveniences/availability of information and communications technology devices, and internet access; (4) housing units in the country and characteristics, such as type of building/ house, construction materials of the roof and outer walls, tenure status of the house and lot, fuel used for lighting and cooking, source of water supply, toilet facility and manner of garbage disposal; and (5) characteristics of the barangay, the smallest unit of government in the Philippines, which will be used as basis for urban-rural classification.

About 58,000 enumerators, 11,500 team supervisors, 3,300 census area supervisors and 2,800 assistant census area supervisors will be involved in the conduct of the forthcoming census. Final population counts by province, city, municipality and barangay are set to be released before December 31, 2010, and will be rendered official upon the proclamation of the President of the Philippines. The details on demographic and housing characteristics, on the other hand, are targeted to be released in July 2011.

For further information, please contact, Administrator Carmelita N. Ericta of the Philippine National Statistics Office at C.Ericta@census.gov.ph

Alternative Sampling Design and Survey Strategies Developed for the Conduct of the Census of Agriculture and Fisheries (CAF)

Three (3) agencies of the Philippine Statistical System - the Statistical Research and Training Center (SRTC), National Statistics Office (NSO) and Bureau of Agricultural Statistics (BAS) are undertaking a collaborative research effort on "Development of Alternative Sampling Design and Survey Strategies for the Conduct of the Census of Agriculture and Fisheries (CAF)." The study aims to come up with an optimum design and strategy for the conduct of CAF census in 2012.

The series of workshops and consultations enabled the group to decide that the upcoming agricultural census be conducted on a modular approach: a core census module and census supplementary modules.

The core census module will be conducted either on a complete enumeration basis or on a large sample basis, depending on budget allocation, with restricted range of key structural items of importance to national policy-making and construction of sampling frames. On the other hand, the census supplementary modules will be conducted on a sample basis at the same time as, or immediately after, the core census module - to provide more detailed structural data. The sample for the census supplementary modules will be selected based on sampling frames from the core census module.

The Steering Committee and the Technical Working Group of the project are working closely with census specialists/consultants in assisting the Philippine National Statistics Office in the preparatory activities of the census. The project team is now tasked to determine the data items and questions to be asked for these modules. There will be several activities to be undertaken for the modular census that include conduct of researches useful in verifying quality, comparability, and comprehensiveness of agriculture database.

For further information, please contact, Executive Director Gervacio G. Selda, Jr. at e-mail: ggseldajr@srtc.gov.ph

Housing Preference and Affordability Survey

The Philippine National Statistics Office, Housing and Urban Development Coordinating Council and Statistical Research and Training Center recently conducted the Housing Preference and Affordability Survey (HPAS) to determine the demand and preferences of the Filipino households for housing in selected highly urbanized areas in the Philippines. The survey was a rider to the first round of 2009 Family Income and Expenditure Survey (FIES). Limited budget confined the survey to only one region and three provinces - National Capital Region, Laguna, Cebu and Davao del Sur. In the HPAS, households that have plans to buy a house, lot or both were interviewed. Total sample covered was 8,404 households. Results of the survey will be used to update the housing plans and programs of the government.

For further information, please contact, Executive Director Gervacio G. Selda, Jr. at e-mail: ggseldajr@srtc.gov.ph

Turkey

Öztas Ayhan

In this issue, I would like to provide information on several important methodological projects which was recently initiated in Turkey.

Small Area Estimation for Labour Force, Employment and Unemployment Ratios in Provinces Using Household Labour Force Survey. Household Labour Force Survey (LFS) has been conducted by Turkish Statistical Institute (TurkStat) periodically since 1988 with some improvements in the estimation levels. The design of the survey has been revised in 2004 and it was aimed to produce quarterly estimates for urban and rural areas of Turkey, annual estimates for total in NUTS–2 level and for urban and rural in NUTS–1 level. Following the establishment of Address Based Population Registry System by TurkStat, the design of the survey has been modified in order to use the updated frame in 2009 by considering the same estimation level. Although the modifications in the estimation level met most of the user needs, some methodological alternatives have been considered for supplying the demands in province levels.

The study was implemented mainly in three stages. Direct estimation and related standard errors at province level by using 2008 LFS data, determination of the variables for the model by using external data sources, and calculation of the composite estimates. For the estimation of labour force, employment and unemployment ratios and their standard errors in province level, one of the model based small area estimation methods called “Empirical Best Linear Unbiased Predictor” (EBLUP) was applied by using 2008 LFS data. The *Ratio Approach* was aimed to provide the consistency between the estimates gathered from small area and direct estimates from LFS in NUTS–2 level. Therefore, in place of estimating the absolute value of the statistic (ph), the ratio of the statistic at province to its estimate at NUTS–2 level was used. In this approach, the covariates were also expressed in terms of ratios. TurkStat published the basic labour force indicators in province level for 2008, depending on the indirect estimations on labour force indicators obtained from modeling studies for the first time on December 2009. More information about the study is available from the web site www.turkstat.gov.tr and please contact Mr Enver Tasti at Enver.tasti@tuik.gov.tr

Regional Gross Value Added. Regional Gross Value Added (GVA) has been estimated using national accounts definitions and concepts, by the TurkStat. The 1998 based GDP series have been compiled according to the European System of Accounts (ESA-95) which is comprehensive and integrated set of accounts and the regional GVA figures are totally in line with the national figures. Regional GVA at current basic prices measured by using production approach and published at statistical region level 2 (26 regions) by kind of economic activity namely agriculture, industry and services for the period 2004-2006. For this study, most of all survey data and administrative sources have been analyzed. The main data sources are agriculture statistics, annual industry and services statistics and household labor force survey. In addition, other statistics such as consumer price indices, regional price level indices and administrative data sources have been analyzed.

Regional wages and salaries from annual industry and service statistics were used as a distribution key to regionalize gross value added. Labor input method was applied by using regional household labor force statistics to include non-observed economy. TurkStat published the figures for regional gross value added on January 2010. More information about the study is available from the web site www.turkstat.gov.tr and please contact Mr Mehmet Aktas at: Mehmet.aktas@tuik.gov.tr.

The Seasonal and Calendar Adjustment of Short Term Economic Indicators. To enable users and decision makers compare data pertaining to consecutive periods, Turkish Statistical Institute (TurkStat) and Central Bank of the Republic of Turkey have started their joint work to establish the seasonal and calendar adjustment of short-term economic indicators applying the methodology recommended by Eurostat. Seasonality can be defined as a pattern of a time series, which repeats at regular intervals every year. The identification of the seasonal patterns of economic variables plays a crucial role in the analysis of current economic developments.

TurkStat produced the seasonal and calendar adjustment of short term economic indicators by applying the methodology of ARIMA (Autoregressive Integrated Moving Average) model based on TRAMO-SEATS (Gomez and Maravall, 1996).

Seasonal and calendar adjusted series for the short term economic indicators started with the industrial production index and gross domestic product and these series are published for the first time in the press release namely Seasonal and Calendar Adjusted Indicators. The indicators will be extended with other short term indicators. More information is available from the web site www.turkstat.gov.tr and please contact Mr Mehmet Aktas at Mehmet.aktas@tuik.gov.tr.

Official Statistics Programme. Based on the Statistics Law (No 5429) of Turkey, in coordination with the responsible institutions and organizations, Turkstat prepared the Official Statistics Programme for a 5-year-period. With the implementation of the Programme, basic principles and standards dealing with the production and dissemination of official statistics have been determined, the production of reliable, timely, transparent and impartial data required at national and international level have been maintained. A programmed period for the Turkish Statistics System has been launched with the Official Statistics Programme 2007-2011. The Programme prevents repetitions in the production and dissemination of official statistics, decreases the burden on respondents, saves on labour force and resources and enhances the confidence on official statistics. Furthermore, official statistics are standardized, responsible and related institutions are defined, data compilation

methodology and the publication periodicity/schedule of official statistics are specified. More information is available from the web site www.turkstat.gov.tr.

New and Emerging Methods

Small Area Estimation:

Basic Concepts, Models and Ongoing Research

Danny Pfeffermann

msdanny@huji.ac.il

1. INTRODUCTION

Over the past three decades there has been a growing demand in many countries for reliable estimates of small area parameters such as means, counts, proportions or quantiles. Typical examples include the estimation of unemployment rates, proportions of people under poverty, disease incidence and use of illicit drugs. These estimates are used for fund allocation, new social or health programs, and more generally, for short and long term planning. Small area estimation techniques are also used in many countries to adjust the counts obtained from censuses. Although commonly known as "small area estimation" (SAE), the domains of studies may actually consist of socio-demographic subgroups as defined, for example, by gender, age and race, or the intersection of such domains with geographical location. Closely related terms in common use are 'poverty mapping' or 'disease mapping', whereby small area estimates of poverty measures or disease incidence are computed and then presented on a map, with different colours defining different values (categories) of the estimators.

The problem of SAE is that the sample sizes in at least some of the domains of study are very small, and often there are no samples available for many or even most of these domains. As a result, the direct estimates obtained from the survey are unreliable, and no direct survey estimates can be computed for areas with no samples. SAE methodology addresses therefore the following two major problems:

1. How to obtain reliable point estimates,
2. How to assess the error of the estimators (MSE, confidence intervals, etc.)

The common feature to all small area estimation problems is that point estimators and measures of error are required for each area separately, and not simply as averages over all the areas under consideration. Notice in this regard that even if direct survey estimates can be used for areas with samples, no design-based methodology exists for estimating the quantities of interest in areas with no samples, requiring instead the use of statistical models. The term "design-based inference"

refers to inference based on the randomization distribution over all the samples possibly selected from the finite population under study, with the population values considered as fixed numbers. Note also that the sample sizes in the various areas are random, unless if some of the areas are defined as strata and samples of fixed sizes are taken in these areas.

SAE methods can therefore be divided broadly into 'design-based' methods and 'model-based' methods. The latter methods use either the frequentist approach or a fully Bayesian methodology, and in some cases combine the two, which is known in the SAE literature as 'empirical Bayes'. Design-based methods often use a model for the construction of the estimators (known as 'model assisted'), but the bias, variance and other properties of the estimators are evaluated under the randomization (design-based) distribution over all possible samples. Model-based methods on the other hand generally condition on the selected sample, and the inference is with respect to the underlying model.

A common feature to design-based and model-based SAE is the use of covariate (auxiliary) information collected in the survey and in other large surveys or administrative records, such as censuses or registers. The use of auxiliary information for SAE is crucial because with the small sample sizes often encountered in practice, even the most sophisticated model can be of little help if it does not involve a set of covariates that provide good predictions of the small area quantities of interest. Some estimators require knowledge of the covariates at the area level only, some require knowledge of the covariates for the sampled units and the true area means of these covariates, and some estimators require knowledge of the covariates for every unit in the population.

In what follows I describe briefly some of the basic methods used for SAE, assuming for convenience that the sample is selected by simple random sampling. More advanced methods and related theory with many examples and references can be found in Rao (2003) and the review papers by Ghosh and Rao (1994), Rao (1999), Pfeffermann (2002) and Rao (2005). See also the chapters by Lehtonen and Veijanen (2009) and Datta (2009) in the new Handbook of Statistics, 29B. In the last section of this article I list some of the new developments in this area.

2. DESIGN-BASED METHODS

Let Y define the characteristic of interest and denote by y_{ij} the outcome value for unit j belonging to area i, $i = 1,...,M$; $j = 1...N_i$, where N_i is the area size. Let $s = s_1 \cup ... \cup s_m$ denote the sample, where s_i of size n_i is the sample observed for area i. Suppose that it is required to estimate the true area mean $\bar{Y}_i = \sum_{j=1}^{N_i} y_{ij} / N_i$. If no auxiliary information is available, the *direct* design unbiased estimator and its variance over the randomization distribution for given sample size n_i are given by,

$$\hat{\bar{Y}}_i = \sum_{j=1}^{n_i} y_{ij} / n_i \;\; ; \;\; Var_D[\hat{\bar{Y}}_i \mid n_i] = (S_i^2 / n_i)[1-(n_i / N_i)] = S_i^{*2}, \qquad (1)$$

where $S_i^2 = \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_i)^2 / (N_i - 1)$. Clearly, for small n_i the variance will be large, unless the variability of the *y-values* is sufficiently small. Suppose, however, that values x_{ij} of *p* concomitant variables $\mathrm{x}_1,...,\mathrm{x}_\mathrm{p}$ are measured for each of the sampled units and that the area means $\bar{X}_i = \sum_{j=1}^{N_i} \mathrm{x}_{ij} / N_i$ are likewise known. Assuming $\mathrm{x}_{1ij} = 1$ for all (i, j), a more efficient design-based estimator in this case is the regression estimator,

$$\hat{\bar{Y}}_i^{\mathrm{Re}g} = \bar{X}_i'\hat{\beta}_i \quad ; \; Var(\hat{\bar{Y}}_i^{REG} \,|\, n_i) \cong S_i^{*2}(1-\rho_i^2)\,, \tag{2}$$

where $\hat{\beta}_i = [\sum_{j=1}^{n_i} \mathrm{x}_{ij}\mathrm{x}_{ij}']^{-1} \sum_{j=1}^{n_i} \mathrm{x}_{ij}\mathrm{y}_{ij}$ is the ordinary least square estimator and ρ_i is the multiple correlation coefficient between Y and $\mathrm{x}_1,...,\mathrm{x}_\mathrm{p}$ in area i. The variance approximation in (2) assumes large n_i. Thus, by use of the concomitant variables, the variance is reduced by the factor $(1-\rho_i^2)$, illustrating the importance of using auxiliary information with good prediction power (large R^2) for SAE.

Although the regression estimator (2) usually has smaller variance than the simple sample mean, its variance may still be large for small sample size, unless the multiple correlation is very large. However, if the regression relationships between y and x are 'similar' across the areas, a more stable estimator is the *synthetic regression* estimator,

$$\hat{\bar{Y}}_i^{Syn} = \sum_{j=1}^{N_i} \hat{y}_{ij} \,/\, N_i = \bar{X}_i'\hat{B}\,, \tag{3}$$

where $\hat{y}_{ij} = \mathrm{x}_{ij}'\hat{B}$ and $\hat{B}$ may be computed as $\hat{B} = [\sum_{i,j\in s} \mathrm{x}_{ij}\mathrm{x}_{ij}']^{-1} \sum_{i,j\in s} \mathrm{x}_{ij} y_{ij}$. The prominent advantage of synthetic estimation is the substantial reduction in variance, because the estimator $\hat{B}$ uses all the sample data, but it can lead to severe biases if the regression relationships actually differ between the areas.

In order to correct for the possible large bias of the synthetic estimator, an approximately design-unbiased estimator in common use is the GREG estimator,

$$\hat{\bar{Y}}_i^{Greg} = \sum_{j=1}^{N_i} \hat{y}_{ij} \,/\, N_i + \sum_{k\in s_i} (y_{ik} - \hat{y}_{ik}) \,/\, n_i\,. \tag{4}$$

However, this estimator may again be unstable in small samples since the second expression in the right hand side of (4) is an area sample mean. Thus, the choice between the synthetic and GREG estimators is a trade off between bias and variance. A compromise is achieved by using a composite estimator of the form,

$$\hat{\bar{Y}}_i^{Com} = \alpha_i \hat{\bar{Y}}_i^{Greg} + (1-\alpha_i)\hat{\bar{Y}}_i^{Syn}; \quad 0 \le \alpha_i \le 1\,. \tag{5}$$

It is common to choose the coefficient α_i (different coefficients in different areas) as some function of the achieved sample size n_i, but such choices account only partly for the relative MSE of the two estimators, as reflected also by the fact that the coefficients would be the same irrespective of the target variable of interest.

Design-based estimators are generally viewed as 'model free', but the requirement for at least approximate design-unbiasedness can result in using estimators with large variances due to the small sample sizes. Recent research in design-based SAE focuses therefore on the use of calibration methods as a means of reducing the variance. See Lehtonen and Veijanen (2009) for review. Even so, the construction of confidence intervals generally requires large sample normality assumptions, which cannot really be assumed in SAE problems. No design-based theory exists for estimation in areas with no samples since the use of the randomization distribution does not extend to prediction problems.

3. MODEL-DEPENDENT ESTIMATORS

In view of the problems underlying the use of design-based methods, it is common practice in many applications to use instead statistical models that define how to "borrow strength" from other areas and/or over time in the case of repeated surveys. Clearly, any model used for SAE has to be tested very carefully before it can be applied. Let θ_k define the parameter of interest in area k, $k = 1,...,M$, and let y_i, x_i denote the data observed for sampled area i, $i = 1,...,m$, where m denotes the number of areas with data on the outcome variable. When the only available information is at the area level, y_i is commonly the direct estimator of θ_i and x_i is a vector of area-level covariates. When unit level information is available, y_i is a vector of individual outcomes and x_i is the corresponding matrix of individual covariate information.

A typical small area model consists of two parts: the first part models the distribution of $y_i \mid \theta_i; \psi_{(1)}$. The second part models the distribution of $\theta_i \mid \mathrm{x}_i; \psi_{(2)}$, linking θ_i to the parameters in other areas (or at different times) and to the covariates. The (vector) parameters $\psi_{(1)}$ and $\psi_{(2)}$ are usually unknown and are estimated from all the available data $D(s) = \{y_i, \mathrm{x}_i; i = 1...m\}$. In what follows I define and discuss briefly three models in common use.

3.1 *'Unit level random effects model'*

The model, employed originally by Battese *et al.* (1988) assumes,

$$y_{ij} = \mathrm{x}'_{ij}\beta + u_i + \varepsilon_{ij}, \tag{6}$$

where u_i and ε_{ij} are mutually independent error terms with zero means and variances σ_u^2 and σ_ε^2 respectively. The 'random effect' u_i represents the joint effect of area characteristics not accounted for by the covariates. Under the model, the true small area means are $\overline{Y}_i = \overline{X}_i'\beta + u_i + \overline{\varepsilon}_i$, but since $\overline{\varepsilon}_i = \sum_{j=1}^{N_i} \varepsilon_{ij} / N_i \cong 0$ for large N_i, the target parameters are often defined as $\theta_i = \overline{X}_i'\beta + u_i = E(\overline{Y}_i \mid u_i)$. For known variances $(\sigma_u^2, \sigma_\varepsilon^2)$, the best linear unbiased predictor (BLUP) of θ_i is,

$$\hat{\theta}_i = \gamma_i[\overline{y}_i + (\overline{X}_i - \overline{\mathrm{x}}_i)'\hat{\beta}_{GLS}] + (1-\gamma_i)\overline{X}_i'\hat{\beta}_{GLS}, \tag{7}$$

where $\hat{\beta}_{GLS}$ is the generalized least square estimator of β computed from all the observed data and $\gamma_i = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2 / n_i)$. For areas l with no sample, $\hat{\theta}_l = \overline{X}_l'\hat{\beta}_{GLS}$. The estimator $\hat{\theta}_i$ is again a composite estimator but now the "shrinkage factor", γ_i, is chosen optimally under the model. Notice that the synthetic estimator, $\overline{X}_i'\hat{\beta}_{GLS}$, is

unbiased for θ_i under the model in the sense that $E(\bar{X}_i'\hat{\beta}_{GLS} - \theta_i) = 0$. The BLUP $\hat{\theta}_i$ is also the Bayesian predictor (posterior mean) under normality of the error terms and a diffuse prior for β. In practice, however, the variances σ_u^2 and σ_ε^2 are seldom known. A Bayesian solution to this problem is to set prior distributions for the unknown variances and then compute the corresponding posterior mean and variance of $\theta_i \mid D(s)$ by aid of Markov Chain Monte Carlo (MCMC) simulations. The common procedure under the frequentist approach is to replace the unknown variances in the BLUP formula by standard variance components estimators such as Maximum Likelihood (MLE), Restricted MLE (REML) or Analysis of Variance (ANOVA) estimators. The resulting predictors are known as the Empirical BLUP (EBLUP). Substituting the true variances by sample estimates in the Bayesian predictors yields what is known as empirical Bayes (EB) predictors. See the references listed in the introduction for estimation of the MSE of the EBLUP and EB under different methods of variance estimation.

3.2 *'Area level random effects model'*

This model is in broad use when the concomitant covariate information is only at the area level. It was used originally by Fay and Herriot (1979) for predicting the mean per capita income in geographical areas of less than 500 inhabitants. Denote by $\tilde{\theta}_i$ the direct sample estimator of θ_i. The model assumes that,

$$\tilde{\theta}_i = \theta_i + e_i \;\; ; \;\; \theta_i = \mathrm{x}_i'\beta + u_i, \tag{8}$$

such that e_i represents the sampling error, assumed to have zero mean and known design variance $Var_D(e_i) = \sigma_{Di}^2$. The model integrates therefore a model dependent random effect u_i, and a sampling error e_i with the two errors being independent. The BLUP under this model is,

$$\hat{\theta}_i = \gamma_i\tilde{\theta}_i + (1-\gamma_i)\mathrm{x}_i'\hat{\beta}_{GLS} = \mathrm{x}_i'\hat{\beta}_{GLS} + \gamma_i(\tilde{\theta}_i - \mathrm{x}_i'\hat{\beta}_{GLS}), \tag{9}$$

which again is a composite estimator with coefficient $\gamma_i = \sigma_u^2/(\sigma_{Di}^2 + \sigma_u^2)$. As with the unit level model, the variance σ_u^2 is usually unknown and is either assigned a prior distribution under the Bayesian approach, or replaced by a sample estimate in (9), yielding the corresponding EBLUP (or the EB) predictor.

3.3 *Unit level random effects model for binary data*

The previous two models assume continuous outcomes. Suppose now that y_{ij} is a binary variable taking the values 0 or 1. For example, $y_{ij} = 1$ if individual j in area i is unemployed (or suffers from a certain disease) and $y_{ij} = 0$ otherwise, such that $P_i = N_i^{-1}\sum_{j=1}^{N_i} y_{ij}$ is the true area unemployment rate (true area disease prevalence). The following model is often used for predicting the proportions P_i:

$$\begin{aligned} &y_{ij} \mid p_{ij} \overset{indep.}{\sim} Bernoulli(p_{ij}) \\ &\mathrm{logit}(p_{ij}) = \log[p_{ij}/(1-p_{ij})] = \mathrm{x}_{ij}'\beta + u_i \,;\, u_i \overset{indep.}{\sim} N\left(0, \sigma_u^2\right) \end{aligned}, \tag{10}$$

where, as in (6), x_{ij} is a vector of covariates, β is a vector of fixed regression coefficients and u_i is a random effect representing the unexplained variability of the individual probabilities.

For this model there is no explicit expression for the predictor $\hat{P}_i$. Writing $P_i = N_i^{-1}[\sum_{j\in s_i} y_{ij} + \sum_{l\notin s_i} y_{il}]$, predicting P_i by its best predictor is equivalent to the prediction of the sum $\sum_{l\notin s_i} y_{il}$ of the missing observations. This is most conveniently implemented by use of the Bayesian methodology, which also produces simple measures of error, but frequentist solutions are also available. See Jiang *et al*. (2002) for the computation of the empirical best predictor and estimation of its MSE.

4. SOME RECENT DEVELOPMENTS

When I wrote my review in 2002, research on SAE was flourishing but my feeling at that time was that the topic had been more or less exhausted in terms of new developments. As the past 8 years show, this perception was completely wrong and research in SAE is accelerating. The diversity of new problems investigated is overwhelming, and the solutions proposed are not only very elegant and innovative but are also very practical. Below I list some of the new developments in recent years. I am in the (slow) process of writing a new review paper that will elaborate on these and other developments in more detail, with appropriate references.

Calibration by use of instrumental variables; Derivation of model dependent predictors that are design-consistent; Sampling designs for efficient SAE; Resampling methods for MSE estimation; Choice of matching priors in Bayesian applications; Benchmarking to design-based estimators in large areas; Treatment of outliers and errors in covariates; M-quantile SAE; Semi-parametric SAE; Prediction of ordered area means; SAE under informative sampling; Accounting for spatial correlations between the area means; Predictive fence methods for model selection.

5.REFERENCES

Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28-36.

Datta, G.S. (2009). Model-based approach to small area estimation. In: Handbook of Statistics 29B; *Sample Surveys*: *Inference and Analysis*. Eds. D. Pfeffermann and C.R. Rao. North Holland. pp. 251-288.

Fay, R.E., and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269-277.

Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: an appraisal (with discussion). *Statistical Science*, **9**, 65-93.

Jiang, J., Lahiri, P., and Wan, S.M. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *The Annals of Statistics*, **30**, 1782–1810.

Lehtonen, R., and Veijanen, A. (2009). Design-based methods of estimation for domains and small areas. In: Handbook of Statistics 29B; *Sample Surveys*: *Inference and Analysis*. Eds. D. Pfeffermann and C.R. Rao. North Holland. pp. 219-249.

Pfeffermann, D. (2002). Small area estimation - new developments and directions. *International Statistical Review*, **70**, 125-143.

Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, **25**, 175-186.

Rao, J.N.K. (2003), *Small Area Estimation*. New York: Wiley.

Rao, J.N.K. (2005). Inferential issues in small area estimation: some new developments. *Statistics in Transition*, 7, 513-526

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 Leyla Mohadjer at MOHADJL1@WESTAT.com.

Announcements

2010 International Methodology Symposium

Statistics Canada
October 26-29, 2010

Ottawa, ON, Canada

Social Statistics: The Interplay among Censuses, Surveys and Administrative Data

Statistics Canada's 2010 International Methodology Symposium will take place at the Crowne Plaza Hotel, located in the heart of downtown Ottawa, from October 26-29, 2010.

The Symposium will be titled **"Social Statistics: The Interplay among Censuses, Surveys and Administrative Data"**. Members of the statistical community, such as those from private organizations, governments, or universities, are invited to attend, particularly if they have a special interest in statistical or methodological issues resulting from the use of multiple sources of data (censuses, sample surveys or administrative data).

The first day will consist of workshops, while the following days will consist of both plenary and parallel sessions covering a variety of topics. Additional research and results may be presented via poster sessions.

The presentations will be related to the methodological aspects of using multiple sources of data. Topics may include:

- Sampling Frames and Sample Design
- Coordinating Samples
- Content and Questionnaire Design
- Data Collection Methods and Acquisition of Administrative Data
- Supplementing Survey Data with Administrative Data
- Administrative Data for Direct Estimation
- Statistical Databases from Administrative Data (e.g., Population Registers)
- Imputation
- Weighting and Estimation
- Dissemination and Data Access
- Record Linkage Techniques
- Record Linkage Software
- Measurement Errors
- Response Burden
- Treatment of Nonresponse
- Confidentiality, Privacy and Ethical Issues
- Small Area Estimation

Visit our Internet site regularly to obtain further details about the program, workshops, registration, accommodation, tourism information and more at
http://www.statcan.gc.ca/conferences/symposium2010/index-eng.htm

RSS 2010 Brighton, 13–17 September
International Conference

A forum for presentation and discussion of **methodological developments** and **areas of application** for statisticians and users of statistics.

An **opportunity** for statisticians, analysts, researchers and other users of statistics from **all sectors** to share current research and insights.

The social side

- Welcome reception, Monday 13 September, Brighton Sealife Centre
- Fish and chips on Brighton Pier with access to the rides on Tuesday evening.
- Conference dinner, Holiday Inn Hotel on Brighton seafront

Topics include

- climate change modelling
- trust in official statistics
- 2011 census
- measuring progress
- adaptive clinical trials
- performance indicators
- composite likelihood
- MCMC
- data capture
- risk
- statistical literacy

Key dates

31 March
Deadline for abstract submissions

24 May
Deadline for grant and bursary applications

7 June
Deadline for 'double discount' registration

9 August
Deadline for discounted registration

8 September
Deadline for pre-event registration

Confirmed speakers

- Tim Davis (Jaguar Land Rover)
- Peter Donnelly (University of Oxford)
- Nancy Reid (University of Toronto)

www.rss.org.uk/rss2010

FOURTH INTERNATIONAL CONFERENCE ON ESTABLISHMENT SURVEYS (ICES IV)

Planning is underway for the Fourth International Conference on Establishment Surveys (ICES IV) which is jointly sponsored by the American Statistical Association (ASA), the International Association of Survey Statisticians, the Statistical Society of Canada, the ASA Section on Survey Research Methods, and the ASA Section on Government Statistics. The conference will be held June 11-14, 2012 at the Sheraton Centre Montreal in Quebec, Canada.

The ICES series of conferences serves an important role in the world of survey methodology. Few other conferences focus on methods and applications for establishment surveys. The first three, held at seven-year intervals from 1993 to 2007, were very successful with over 400 delegates per conference. The conferences covered a broad spectrum of survey methods for businesses, farms, and institutions. The fourth conference will continue this. Examples of potential topics include:

- Efficient Use of Administrative Data in Business Surveys
- Advances in Disclosure Protection
- Usage of Linearization Variance Estimators for Survey Estimates
- The New Direction of Business Surveys
- Collecting Data Electronically from Businesses
- Updating Business Registers
- Sample Design Challenges
- Factors that Affect Establishment Survey Participation
- Generalized Survey Processing Systems – An Update
- Measuring Nonresponse Bias
- Issues of Multi-Mode Data Collection

The Conference will include:

- Short courses
- A keynote speaker
- Poster sessions, software demonstrations, invited and contributed paper sessions

On-line solicitation for invited papers will occur December 1, 2010 – March 1, 2011. Dates for on-line solicitation for contributed papers are forthcoming.
Inquiries may be directed to ices4@amstat.org.

Update on the IASS programme at ISI 2011

The programme of Invited Paper Meetings for the ISI meeting in 2011 has now been finalised. Last year, suggestions for topics were collected by the IASS programme subcommittee (consisting of Geoff Lee, Wendy Barboza, Walter Davis, George Djolov, Dave Dolson, Gabriele Durrant, Michael Larsen, Dalisay Maligalig, Colm O'Muircheartaigh, Yves Tillé, Marcel Vieira and Li-Chun Zhang). These suggestions were discussed by the Programme Co-ordinating Committee at the ISI meeting in Durban.

The IASS Invited Paper Meetings for the 2011 ISI meeting will be:

IPS54: "Complex analysis for longitudinal data"
IPS55: "Inference for linked data"
IPS56: "Estimation and inference in complex survey sampling"
IPS57: "Data mining and machine learning in statistics organisations"
IPS58: "Best practice methodologies for mixed mode surveys"
IPS59: "Record linking and imputation in administrative data"
IPS60: "Applying and measuring confidentiality methods"
IPS61: "New approaches for agricultural statistics"
IPS62: "Small area estimation for complex surveys"
IPS63: "Managing the effects of changing measurement technologies on environmental longitudinal surveys"
IPS64: "Census strategies"

For further information about the IASS programme, please contact Geoff Lee (geoff.lee@abs.gov.au).

Book and Software Review

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 John Eltinge at Eltinge.John@bls.gov.

In Other Journals

Survey Practice

Practical Information for Survey Researchers

About Submit Articles Editorial Review Guidelines for Authors Editorial Board

Survey Practice April 2010

Articles

More on the Extent of Undercoverage in RDD Telephone Surveys Due to the Omission of 0-Banks
Martin Barron, Jenny Kelly, Robert Montgomery, James Singleton, Hee-Choon Shin, Benjamin Skalland, Xian Tao, and Kirk Wolter

Testing Prepaid Incentives and a Mail Questionnaire to Increase Response in a Multi-Wave Telephone Survey
Anne Kenyon, Lynn Newman, Suzanne Triplett, Anne-Marie Knokey Kathryn Valdes, and Helen Smith

Does Prefilling Questions in a Longitudinal Survey Encourage Participation?
Geraldine M. Mooney, Melissa Krakowiecki, and Deborah Trunzo

Predicting response rate: A natural experiment
Kay W. Axhausen, Claude Weis

Making Good Use of Survey Paradata
Peter Lynn, University of Essex and Gerry Nicolaas

Summary of Address Based Sampling papers from 2009 AAPOR Conference
Jamie Paul and Sarah Grady

Recent Books in Public Opinion, Survey Methods, and Survey Statistics
Mario Callegaro

The Editors

John Kennedy Diane O'Rourke
David Moore Andy Peytchev

survprac@indiana.edu

Survey Research Methods
surveymethods.org

HOME ABOUT LOG IN REGISTER SEARCH CURRENT ARCHIVES

Home > **Vol 3, No 3 (2009)**

Survey Research Methods

The journal is edited by Peter Lynn of the University of Essex, UK, and Rainer Schnell of the University of Duisburg-Essen, Germany.

This peer-reviewed journal aims to be a high quality scientific publication that will be of interest to researchers in all disciplines involved in the design, implementation and analysis of surveys. The journal will be published electronically with free and open access via the internet. We aim for a fast review and publication process.

Vol 3, No 3 (2009)

Table of Contents

Articles

ISSN: 1864-3361

JOS
Journal of Official Statistics

Home
Current Issue
Archive
Search
Announcement
Journal Information

Current Issue: March 2010, Vol. 26 No. 1
Published Date : 31-March-2010

Statistics Canada Statistique Canada

Home > Publications >

Survey Methodology

Issue notes

Product:	**Survey Methodology**	
Issue:	December 2009, vol. 35 no. 2	Vie
Catalogue no.:	12-001-XIE2009002	
Format:	PDF	

Chronological index

Table of contents

Journal of Privacy and Confidentiality

Welcome to the new home of the Journal of Privacy and Confidentiality!

Our inaugural issue is up (2009/1/1), as well as our latest Special Issue.

We are currently accepting submissions for our next and subsequent issues. Please refer to Submission Guidelines [1] for detailed information regarding manuscript submissions to the JPC.

Current Issue: Volume 1, Issue 2 (2009)

Introduction

We are pleased to present Vol. 1, Issue 2 of the Journal of Privacy and Confidentiality. This special issue, compiled and guest edited by Alan F. Karr, Satkartar Kinney, and Joe Fred Gonzalez, draws from papers presented at a workshop in May 2008, "Data Confidentiality: The Next Five Years", co-sponsored by the National Center for Health Statistics/CDC (NCHS) and the National Institute of Statistical Sciences (NISS). We invite you to learn more about the the workshop and it's generation of this issue in the guest authors' summary.

Articles

TABLE OF CONTENTS

Volume 3, Issue 1, April 2010

This issue is still in press.

Communication-Efficient Privacy-Preserving Clustering
Geetha Jagannathan, Krishnan Pillaipakkamnatt, Rebecca N. Wright, Daryl Umano
Transactions on Data Privacy 3:1 (2010) 1 - 25

Abstract, PDF

Random Forests for Generating Partially Synthetic, Categorical Data
Gregory Caiola, Jerome P. Reiter
Transactions on Data Privacy 3:1 (2010) 27 - 42

Abstract, PDF

Design and Hardware Implementation of QoSS - AES Processor for Multimedia applications
Zeghid Medien, Mohsen Machhout, Belgacem Bouallegue, Lazhar Khriji, Adel Baganne, Rached Tourki
Transactions on Data Privacy 3:1 (2010) 43 - 64

Abstract, PDF

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

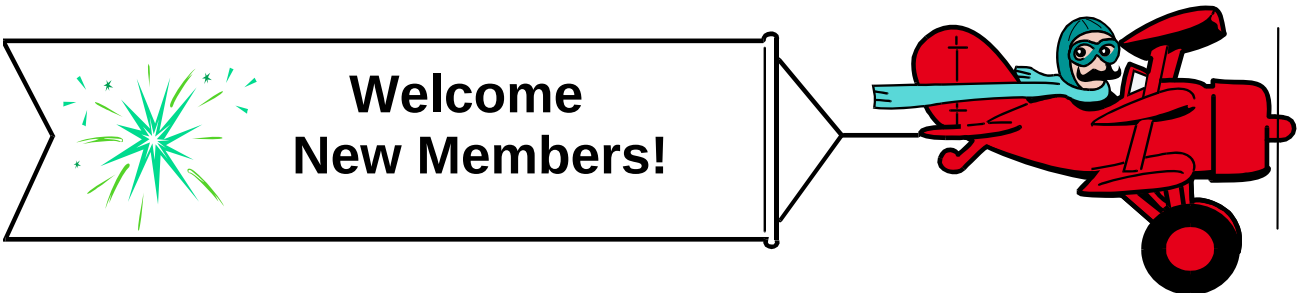
Journal of Statistical Planning and Inference

CONTENTS

Volume 121, Number 2

BIOMETRIKA VOL. 97, NO. 1

CONTENTS



We are very pleased to welcome the following new members:

Country	First Name	Last Name
Belgium	Machteld	Hoskens
Brazil	Luna	Hidalgo
Burma	Htun Htun	Oo
Burundi	Léonard	Ntakirutimana
Canada	Charles H.	Goldsmith
Central African Republic	Bienvenu Gustave	Semngai
Finland	Ville	Vertanen
Finland	Janika	Konnu
Finland	Outi	Ahti-Miettinen
France	Catherine	Meunier
Ghana	Nicholas Nicodemus	Nsowah-Nuamah
India	Hukum	Chandra
Norway	Li-Chun	Zhang
Switzerland	Esperanza	Magpantay
United Kingdom	Jose Manuel	Roche
United Kingdom	Antony	Fielding
United Kingdom	Emilio	Lopez Escobar
United Kingdom	Gabriele	Durrant
United States	Michael Lee	Cohen

IASS Officers and Council Members

Executive Officers		
President (2009-2011):	Susan Linacre (Australia)	susan.linacre@abs.gov.au
President-elect:	Ray Chambers (Australia)	
Vice-Presidents (2009-2011):	Peter Lynn (UK)	plynn@essex.ac.uk
	Linda Hewitt (Republic of Trinidad and Tobago)	linhew@trinidad.net
Scientific Secretary (2007-2009):	Leyla Mohadjer (USA)	LeylaMohadjer@Westat.com
Council Members serving 2009-2013	Mike Hidiroglou (Canada)	Mike.Hidiroglou@statcan.gc.ca
	Edith D. de Leeuw (The Netherlands)	e.d.deleeuw@uu.nl
	Monica Pratesi (Italy)	m.pratesi@ec.unipi.it
	Mick P. Couper (USA)	mcouper@umich.edu
	Eva Elvers (Sweden)	eva.elvers@scb.se
	Yves Tille (Belgium)	yves.tille@unine.ch
Members serving 2007-2011	Jairo Arrow (South Africa)	jairo@statssa.gov.za
	Louise Bourque (Canada)	louise.bourque@stat.gov.qc.ca
	Kathryn Inglis-Clark (Australia)	kate@inglis-clark.com.au
	Dalisay Maligalig (Philippines)	dmaligalig@abd.org
	Paul-André Salamin (Switzerland)	paulandre.salamin@bfs.admin.ch
	Shyam Upadhaya (Nepal)	s.upadhyaya@unido.org
Committee Chairs		
2011 Programme Committee:	Geoffrey Lee (Australia)	geoff.lee@abs.gov.au
The Secretariat		
Executive Director:	Catherine Meunier (France)	catherine.meunier@insee.fr
Treasurer:	Daniel Malaquin (France)	daniel.malaquin@insee.fr
Secretary:	Évelyne Countant (France)	evelyne.coutant@insee.fr

Institutional Members

4 International Organizations

AFRISTAT
CICRED
EUROSTAT
UNITED NATIONS STATISTICAL DIVISION

31 Bureaus of Statistics

ARGENTINA - INSTITUTO NACIONAL DE ESTADISTICA Y CENSOS - INDEC
AUSTRALIA - AUSTRALIAN BUREAU OF STATISTICS
BELGIUM - INSTITUT NATIONAL DE STATISTIQUE
BRAZIL - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA - IBGE
CANADA - STATISTICS CANADA
CHINA - GOVERNO DE MACAU
COTE D'IVOIRE - INSTITUT NATIONAL DE LA STATISTIQUE
CZECH REPUBLIC - CZECH STATISTICAL OFFICE
DENMARK - DANMARKS STATISTIK
FINLAND - STATISTICS FINLAND
FRANCE - INSTITUT NATIONAL DE STATISTIQUE ET D'ÉTUDES ÉCONOMIQUES - INSEE
GAMBIA - CENTRAL STATISTICS DEPARTMENT
GERMANY - STATISTICHE BUNDESAMT
GREECE - NATIONAL STATISTICAL SERVICE OF GREECE
IRAN - STATISTICAL CENTER OF IRAN
ITALY - INSTITUTO CENTRALE DI STATISTICA - ISTAT
MEXICO - INSTITUTO NACIONAL DE ESTADISTICA, GEOGRAFIA E INFORMATICA - INEGI
NETHERLANDS - CENTRAL BUREAU OF STATISTICS
NEW ZEALAND - STATISTICS NEW ZEALAND
NIGERIA - FEDERAL OFFICE OF STATISTICS
NORWAY - CENTRAL BUREAU OF STATISTICS
PORTUGAL - INSTITUTO NACIONAL DE ESTATISTICA - INE
REPUBLIC OF KOREA - NATIONAL STATISTICAL OFFICE - NSO
SPAIN - INSTITUTO NACIONAL DE ESTADISTICA
SWEDEN - STATISTICS SWEDEN
SWITZERLAND - OFFICE FEDERAL DE STATISTIQUE
TANZANIA - BUREAU OF STATISTICS
UNITED KINGDOM - OFFICE FOR NATIONAL STATISTICS
USA - BUREAU OF THE CENSUS
USA - DEPARTMENT OF EDUCATION
USA - DEPARTMENT OF HEALTH AND HUMAN SERVICES

6 Universities, Research Centers, Private Statistics Firms

ARGENTINA - UNIVERSITAD NACIONAL DE TRES DE FEBRERO
DENMARK - SFI
EUROPE - A.C. NIELSEN MANAGEMENT SERVICE
FRANCE - INSTITUT NATIONAL D'ÉTUDES DEMOGRAPHIQUES - INED
USA - RESEARCH TRIANGLE INSTITUTE
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
c/o INSEE-CEFIL
Att. Madame Évelyne Coutant
3, rue de la Cité
33500 Libourne – France

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.cbs.nl/iass/index.htm