Survey Methodology

June 2002
How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1-800-263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line 1-800-263-1136
National telecommunications device for the hearing impaired 1-800-363-7629
Depository Services Program inquiries 1-800-700-1033
Fax line for Depository Services Program 1-800-889-9734
E-mail inquiries infostats@statcan.ca
Website www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Publications.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About us > Providing services to Canadians.
Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.
In This Issue

This issue of Survey Methodology contains the second in an annual invited paper series in honour of Joseph Waksberg. A brief description of the series and a short biography of Joseph Waksberg were given in the June 2001 issue of the journal. The author of the Waksberg Invited Paper for 2002 is Wayne Fuller. I would like to thank the members of the Committee, Graham Kalton (chair), Chris Skinner, David Binder and Paul Biemer, for having chosen such a distinguished statistician, who has made profound contributions to many areas of statistical theory and practice, as the author of the second paper in the Waksberg Invited Paper Series.

In his paper entitled “Regression Estimation for Survey Samples” Wayne Fuller presents a broad overview of historical and recent developments in the use of regression models in surveys for estimation, weight calibration and non-response adjustment. After a brief introduction and historical background, he discusses the use of regression models for estimation in complex surveys from a design based perspective. He follows this with an exploration of the model based perspective. Other topics discussed are the use of regression models for multinomial data, techniques available when auxiliary variables are available for every unit of the population, and regression to account for the effects of non-response in surveys. Finally, consideration of a few practical aspects of applications rounds out this insightful overview of an important area of inference from survey data to which Wayne Fuller himself has made many important contributions.

This issue also contains a special section “Remembering Leslie Kish” which includes four papers, one by Leslie Kish himself containing some of his last thoughts on the topics of combining samples and surveys. Two of the other papers discuss implementations of Leslie Kish’s idea of rolling censuses. These two papers were also presented at the Statistics Canada Symposium 2001 in a special session entitled “Remembering Leslie Kish”.

The first paper in the special section, by Graham Kalton, presents an inspiring overview of Kish’s contributions to many areas of statistics. Many of the problems that Kish worked on are put into historical perspective and their practical importance is emphasized.

The paper by Kish presents ideas that he was still working on at the time of his death in October 2000. I am grateful to Graham Kalton and Jack Gambino for making editorial corrections to the paper, but it is presented largely as it was at the time of Kish’s death. In this paper he argues that, just as statistics represented a new paradigm in the scientific method, and survey sampling required a new paradigm in statistics, so rolling samples and multi-population surveys require new paradigms in survey methods. We can only speculate as to what the final paper would have been like had Kish lived.

Alexander describes the American Community Survey, planned to be introduced by the U.S. Census Bureau in coming years as a replacement for the decennial census long form. This is a very large survey based very much on the idea of rolling samples and censuses that Kish introduced more than twenty years ago. This paper discusses the concepts, frame, sampling design, and cumulation of samples and weighting.

The final paper in the special section, by Durr and Dumais, describes the new rolling census being introduced in France to replace their more traditional census. In this rolling census, every small commune will be surveyed once within a five year period; larger communes will be divided into five rotation groups, each rotation group being surveyed in one of the five years. This paper describes objectives, design and estimation procedures for the rolling census.

In their article, Cahill and Chen develop an approach to exploit data from multiple surveys and epochs by benchmarking the parameter estimates of logit models of binary choice and semi-parametric survival models. Estimates obtained from a survey rich in explanatory variables are benchmarked to information from a survey with significant historical depth. Cahill and Chen demonstrate how the method can be applied, using the maternity leave module of the LifePaths dynamic microsimulation project at Statistics Canada.
Garren and Chang consider the problem of the non-telephone population in telephone surveys using random digit dialing. Using Public Use Microdata Samples, the propensity that a household owns a phone is estimated using generalized linear regression and is used during estimation. Asymptotic biases and variances are presented for both the non-poststratified and poststratified estimators incorporating and not incorporating the estimated propensity. These four estimators are further compared through a simulation study.

The article by Tillé develops an estimator that can be used to avoid the problem of empty post-strata that can occurs with the usual post-stratified estimator. The idea involves using a conditionally weighted estimator and conditioning on ranks in the population of an auxiliary variable known for all units of this population. In this way, the sizes of the post-strata are set in the sample and random in the population. The next step is to calculate the mean of the conditionally weighted estimators to obtain greater stability. The estimator obtained is calibrated on distribution, linear and exactly unbiased. A simulation study is used to show that the proposed estimator is more robust than the generalized regression estimator when the relation of the variable of interest and the auxiliary variable is not linear. Lastly, the article proposes an approximate estimator of the variance verified using simulations.

Shao and Butani consider the problem of estimating variances for imputed survey estimators. They show that the resulting variances can be estimated in two parts, the first of which can be estimated using a grouped half-sample method that incorporates adjustments to take imputation into account. As the estimation of the second part may entail many derivations, Shao and Butani propose an adjustment to the grouped half-sample method that leads to approximately unbiased variance estimates.

In his paper Cohen describes a method to implement Rao and Shao’s jackknife method of estimating variances to account for imputation using replicate weights. Rao and Shao’s method involves calculation, for each jackknife replicate, adjusted values of imputed data points. The method can be used with either mean imputation or hot deck imputation. Cohen’s method involves adding extra rows to the replicate weight file. For each imputed value, one extra row is added for each respondent in the same imputation class.

In the last paper of this issue, Valliant studies several variance estimators for the General Regression (GREG) estimator. The interest is in finding variance estimators that, under certain conditions, are approximately unbiased for both the design-variance and the model-variance even if the model that motivates the GREG has an incorrect variance parameter. A key feature of these robust estimators is the adjustment of squared residuals by factors analogous to the leverages used in standard regression analysis. It is shown that the delete-one jackknife implicitly includes the leverage adjustments and is a good choice from either the design-based or model-based perspective. A simulation study shows that these variance estimators have small bias and produce confidence intervals with near-nominal coverage rates.

M.P. Singh