Survey Methodology, December 2006
Vol. 32, No. 2, pp. 121-122
Statistics Canada, Catalogue No. 12-001

In This Issue

This issue of *Survey Methodology* opens with the sixth paper in the annual invited paper series in honour of Joseph Waksberg. It is with sadness that we note the passing of Joseph Waksberg in January of 2005. A short biography of Joseph Waksberg was given in the June 2001 issue of the journal, along with the first paper in the series. For more information about the life and work of Joseph Waksberg, see the Statistical Science article (Vol. 15, No 3) "A Conversation with Joseph Waksberg," by David Morganstein and David Marker available at http://projecteuclid.org/Dienst/UI/1.0/Home. I would like to thank the members of the selection committee – David Bellhouse, chair, Gordon Brackstone, Sharon Lohr and Wayne Fuller – for having selected Alastair Scott as the author of this year's Waksberg Award paper.

In his paper entitled "Population-Based Case Control Studies", Scott discusses the analysis of case control studies in which the controls are obtained from a complex sample survey. Using the example of logistic regression, he shows how the survey weighted estimates can be quite inefficient because of the relatively small weight given to the cases. Drawing on an analogy with maximum likelihood estimation, he then proposes a simple, much more efficient alternative that is, however, biased for the intercept term. Efficiency and robustness properties are illustrated through examples. Finally, he briefly discusses the problem of case-control family studies.

Kott considers the use of weight calibration to correct for nonresponse and coverage errors. He gives a general description of calibration estimation, and extends Estavao and Särndal's functional form approach to general calibration. He then discusses properties of this calibration method to correct for unit nonresponse and coverage errors under a quasi-randomization model. He concludes with an empirical example and discussion of some issues.

Reiter, Raghunathan and Kinney investigate through a simulation study the effect of ignoring sampling design variables when building imputation models in a multiple imputation context. They show that potential biases can be reduced by controlling for these design variables in the imputation model, either through a fixed-effect or mixed-effect model. They conclude that a useful prescription for imputers is to include as predictors all variables that are related to the variables being imputed, particularly sampling design variables, so as to make the usual assumption of ignorable non-response satisfied.

The article by Funaoka, Saigo, Sitter and Toida investigates the use of bootstrap variance estimators in stratified multi-stage sampling where the sampling fractions are large. They propose a Bernoulli-type bootstrap that provides consistent bootstrap variance estimates when simple random sampling without replacement is used at each stage. The proposed method is simple to implement and can be extended to any number of stages without much complication. The method is illustrated through a limited simulation study and using data from the 1997 Japanese National Survey of Prices.

In the Kozak and Verma paper, the geometric approach to stratification proposed by Gunning and Horgan (2004) is compared with two optimization approaches; the Lavallée-Hidiroglou algorithm (Lavallée and Hidiroglou 1988) and an optimization algorithm proposed by Kozak (2004). Using five artificial populations of various sizes, the three methods are compared under two scenarios; comparison of the resulting CV under a fixed sample size and comparison of the resulting sample sizes under a fixed level of precision.

Deville and Lavallée present general theoretical foundations for the weight share method in indirect sampling. They define the important concept of a link matrix in indirect sampling, which specifies how the elements of the sampled population are linked to the target population and gives weights to these links that permit unbiased estimation. They discuss important properties of the link matrix, and derive necessary and sufficient conditions for an optimal link matrix to exist. The theory is illustrated with some interesting examples.

Deville and Maumy-Bertrand study the determination of a sampling design and an estimation method for a tourist survey. The main issue that this type of survey has to address is the absence of a sampling frame that can be used to directly reach tourists. To get around this problem, authors suggest to sample services for tourists. This is thus a situation of indirect sampling for which the generalized weight-share method is used to obtain estimates of parameters of interest. Some extensions to the method become necessary. The authors focus more specifically on one of them and describe it in greater detail.

Félix-Medina and Monjardin consider a variant of link-tracing sampling. They use a Bayesian approach to construct estimators of population size, however in order to make inferences about the population size that are robust to erroneous specification of the assumed model, the authors make inferences under the frequentist design-based approach. Based on the results of the simulation study, the proposed estimators perform better than the maximum likelihood estimators that are currently used.

The paper by Dorfman, Lent, Leaver, and Wegman presents a comparison of the Consumer Price Index design methodologies of the United Kingdom and the United States employing the same "scanner" data. They conclude that in the population studied, the UK approach, which involves tighter stratification and, more importantly, more restrictive judgment sampling within strata than the probability sampling of the US approach, does better in estimating a target superlative index. This is shown to be the case, whichever low level price index estimator (the ratio of averages, the geometric mean, or the average of ratios) is employed.

In their paper, Thomas, Raghunathan, Schenker, Katzoff and Johnson use multiple imputation to analyze data with missing values caused by a matrix sampling design. In matrix sampling, only a subset of questions is administered to each respondent in order to reduce respondent burden. The authors develop a method for creating matrix sampling forms, each form containing a subset of questions to be administered to randomly selected respondents. The method is designed so that each form includes questions that are predictive of the excluded questions in order to recover some of the information about the latter. The proposed method and multiple imputation are evaluated using data from the National Health and Nutrition Examination Survey.

Harold Mantel, Deputy Editor