Survey Methodology, December 2001
Vol. 27, No. 2, pp. 119-120
Statistics Canada, Catalogue No. 12-001

In This Issue

This issue of *Survey Methodology* contains papers on a variety of topics touching on coverage issues, nonresponse, imputation, survey designs, survey weighting and analysis of data from complex surveys.

In the first paper of this issue, Blenk and Stasny develop a weighting adjustment in order to reduce the coverage bias in telephone surveys while controlling the increase in variance due to weighting. The weighting adjustment is applied to *transient* households, which are households moving in and out of the telephone population during the year. It is assumed that the transient telephone population is representative of the non-telephone population. The weighting adjustment proposed is based on propensity scores for transience obtained using a logistic regression model. The proposed method and several alternatives are compared using data collected from a survey of distressed and non-distressed regions of Kentucky, Ohio, and West Virginia.

Mariano and Kadane use the information on the number of calls in a telephone survey as an indicator of how difficult an intended respondent is to reach. This permits a probabilistic division of the nonrespondents into those who will always refuse to respond and those who were not available to respond in a model of the nonresponse. It also permits an evaluation of whether the nonresponse is ignorable for inference about the dependent variable by incorporating the information on the number of calls into the model. These ideas are implemented on data from a survey in Metropolitan Toronto of attitudes toward smoking in the workplace. The results reveal that the nonresponse is not ignorable and those who do not respond are twice as likely to favor unrestricted smoking in the workplace as are those who do.

In his paper, Hidiroglou unifies the nested and non-nested cases found in the double sampling theory. The nested case, also known as two-phase sampling, corresponds to the traditional case in which a first-phase sample is initially taken so that additional information may be collected. This is followed by a second-phase sample taken within the first one, which contains the variables of interest. The non-nested case reflects a situation in which both samples are selected independently from the same frame or possibly from different frames. Using the generalized difference, an estimator is proposed for both cases, and an optimal estimator that minimizes variance is developed. Variance estimation is also discussed for both cases. Numerous examples of surveys conducted at Statistics Canada illustrate the unification of both cases.

Lavallée and Caron investigate the problem of producing estimates when using record linkage methods to link two populations together. In particular, they consider the problem of producing estimates for one of the populations using a sample from the other one, assuming the two populations have been linked together. The Generalized Weight Share method is adapted to take into account the linkage weights in three different ways: (1) all links where the linkage weight is non-zero; (2) all links where the linkage weights are greater than a given threshold; and (3) the links are randomly chosen. These proposed estimators are compared with the classical approach through a simulation study.

Merkouris considers the problem of producing cross-sectional estimates with data collected from multiple panel surveys. Coverage of the cross-sectional population maybe incomplete due to individuals leaving or entering the population after the selection of the panel. By recognizing that a repeating panel survey is a special type of multiple frame survey, Merkouris is able to propose weighting strategies suitable for various multiple panel surveys. These weighting procedures can be used to combine information from the multiple panels to produce cross-sectional estimates that take into account the dynamic character of the multiple panel design.

Marker investigates survey design strategies to improve the quality of direct small area estimators, thus reducing the need for indirect, model-based estimators. Factors considered include stratification and oversampling, combining data from repeated surveys, harmonizing across different surveys, supplemental samples, and improved estimation procedures.

In their paper, Saigo, Shao and Sitter address the important problem of variance estimation under imputation for missing data. In their paper, they propose a bootstrap method that works for both smooth and non-smooth statistics, even for the case where the number of sampled clusters is small. This improves on their previously proposed bootstrap method which could suffer from serous overestimation when the number of sampled clusters is small. In addition to a bootstrap method, Saigo, Shao and Sitter also propose a repeated Balanced Repeated Replication method that captures the imputation variance in the presense of random imputation. These methods are illustrated through a simulation study.

Bellhouse and Stafford consider nonparametric local polynomial regression as an exploratory data analysis tool for data from complex surveys. They consider a single continuous regressor variable x, which is binned into a finite number of possible values, which may correspond to the precision of measurement of x, but may also be chosen otherwise. Point estimates of the local regression function, and associated variance estimates, are developed. The method is illustrated with an analysis of body mass indices from the Ontario Health Survey, and the nonparametric estimates are compared to those obtained from a parametric model.

In the final paper of this issue, Silva and Smith use a state space approach for modelling of compositional time series using data from a repeated complex survey. A compositional time series is a multivariate time series of proportions constrained to add to one at each time point. They first transform the data using an additive logistic transformation, and then model the transformed series. Estimation methods based on the Kalman filter are developed and then applied to data from the Brazilian Labour Force Survey. The Kalman filter also provides model-based estimates of variance and confidence limits for the transformed series. Estimates of trends and seasonal effects are compared to those obtained using X-11 ARIMA, and found to be generally smoother since they explicitly account for sampling errors in the raw estimates of the series.

M.P. Singh