

In This Issue

This issue of *Survey Methodology* includes papers on a variety of topics including overviews of small area statistics and data quality in statistical offices, survey nonresponse and imputation, survey design, data collection and estimation.

In the first paper of this issue, Brackstone identifies strategies and approaches for the development of small area statistics programs in national statistical offices. The topic of small area estimation will be covered by a number of papers in a special section in the June 2003 issue of *Survey Methodology*. The paper first considers the crucial role of censuses, and discusses issues related to their usefulness for small area statistics. Other potential sources of small area statistics include administrative files and sample surveys, either on their own or combined with census data to provide estimates for the intercensal period or for characteristics not directly covered by the census. Rolling censuses are also discussed, as well as the unique challenges in producing small area business and environmental statistics. Finally, issues of organization of national statistical offices for production and dissemination of small area statistics are considered.

Trewin reviews the practices and approaches used to maintain high quality of output from a national statistical office. Important ingredients include good relations with respondents, skilled and motivated staff, sound statistical and operational methods, and relevance of statistical programs. Current challenges include increasing the use of administrative data sources, effective use of the internet for both collection and dissemination, maintaining knowledge and skills as staff leave, and handling increasing user expectations. This paper is based on a talk presented as the keynote address at Statistics Canada's Symposium 2001.

Thibaudeau presents an innovative approach to the imputation of demographic characteristics in a large scale survey or Census. Instead of relying on the usual approach of either the closest complete record in the processing stream or constructing imputation groups, Thibaudeau proposes a compromise method which uses maximum likelihood estimation based on the conditional probabilities. This approach seeks to create groups that are close in order and in geography to the imputed record. He also presents an interesting Bayesian approach to evaluating the method.

Nandram, Han and Choi consider the problem of analyzing multinomial nonignorable non-response data from small areas in the framework of Bayesian inference. This paper extends some earlier work by Stasny by assuming a Dirichlet prior underlying the multinomial probabilities and using a prior distribution on the hyperparameters. The authors apply this model to Body Mass Index data from a complex survey design.

In the Stewart paper, the possible biases introduced by different contact strategies in telephone time-use surveys are investigated. Two contact strategies, convenient-day scheduling, where the designated reference day changes with the contact day, and designated-day scheduling, where the reference day remains fixed, are discussed and compared through simulation studies.

Bell and McCaffrey consider the problem of unbiasedly estimating the variance of coefficients of linear regressions from multi-stage survey data when only a small number of Primary Sampling Units (PSUs) are sampled. After investigating situations where the bias of the linearization variance estimator can be large, a bias reduced linearization variance estimator is proposed. In addition, a Satterthwaite approximation is used to determine the degrees of freedom to be used for tests and confidence intervals in conjunction with the bias reduced linearization variance estimator.

Sirken considers estimation of the volume of transactions that a population of establishments has with a population of households. An approach based on indirect sampling of establishments through the households that they have transactions with is compared to the more typical approach based on direct pps sampling of establishments. Estimators and expressions for the variances are derived and compared for the two methods. Situations where one approach or the other is preferable are explored.

Rivest considers the problem of identifying stratum boundaries. The commonly used Lavallée-Hidroglou algorithm assumes that the values of the study variable are available and are used in the determination of optimal stratum bounds. In his paper, Rivest relaxes this assumption and modifies the Lavallée-Hidroglou algorithm to account for a discrepancy between the stratification variable and the study variable through the use of models that link these two variables together. These models are then incorporated into the Lavallée-Hidroglou algorithm.

In the Lu and Sitter paper, the problem of the sample size being smaller or only slightly larger than the total number of strata is considered. Consequently, conventional methods of sample allocation to strata may not be applicable. One solution for this problem is to use a linear programming technique to minimize the expected lack of desirability of the samples subject to a constraint of expected proportional allocation (EPA). However, as the number of strata increases this solution rapidly becomes expensive in terms of magnitude of computation. In the proposed approach, the amount of computation is reduced substantially at the small cost of approximate EPA for strict EPA.

Renssen and Martinus explore the use of generalized inverse matrices in survey sampling. After reviewing the properties of generalized inverses, they consider the generalized regression estimator when the set of regressors is not of full rank, and they set out a regularity condition under which the estimator is invariant to the choice of generalized inverse. They then present an algorithm for calculating the regression weights, and briefly discuss weighting in the Dutch Labour Force Survey.

M.P. Singh