SIMULATION METHODS FOR EVALUATION OF THE STABILITY OF DESIGN-BASED VARIANCE ESTIMATORS

J.L. Eltinge, Texas A&M University and V.L. Parsons, NCHS
V.L. Parsons, NCHS, 6525 Belcrest Road, Room 915, Hyattsville, MD 20782

Key Words: Complex sample survey; Model based inference; Operating characteristic surface; Randomization based inference; Stratified multistage sample survey; Superpopulation model.

1. Introduction

1.1 Moments of point estimators and variance estimators

This paper examines the use of some simulation methods to evaluate the properties of variance estimators intended for use with complex survey data. Although these methods apply to a wide class of variance estimators, principal attention is devoted to design-based estimators.

Consider a finite population of N elements i with characteristics Y_i that are independently and identically distributed according to some superpopulation model ξ. Define the finite population total $\theta = \sum_{i=1}^{N} Y_i$; the main ideas considered below apply to other finite population parameters (e.g., a population mean, ratio or regression coefficient) but will not be considered further here. A sample s of n elements is selected according to a complex design p, and we use the sample observations Y_i, $i \in s$, to compute an estimator $\hat{\theta} = \sum_{i\in s} w_i Y_i$ of θ. Here, the weight w_i is defined to equal the inverse of the first-order selection probability for unit i.

Now consider estimation of the design variance of $\hat{\theta}$. To simplify notation, we restrict attention to the commonly encountered case of stratified multistage sampling with L strata and two primary sample units selected with replacement from each stratum. Standard arguments indicate that one may rewrite our point estimator $\hat{\theta} = \sum_{h=1}^{L} \hat{\theta}_h$ where $\hat{\theta}_h = \hat{\theta}_{h1} + \hat{\theta}_{h2}$, $\hat{\theta}_{hj} = \sum_{i\in s_{hj}} w_i Y_i$ and s_{hj} is the set of selected sample elements contained in selected primary unit j in stratum h. Then a commonly used design unbiased estimator of the design variance of $\hat{\theta}$ is

$$\hat{V} = \sum_{h=1}^{L} \hat{V}_h \tag{1.1}$$

where $\hat{V}_h = (\hat{\theta}_{h1} - \hat{\theta}_{h2})^2$.

Moments of $\hat{\theta}$ and $\hat{V}$ can be evaluated with respect to the random variability generated by the design, the superpopulation model, or both. Specifically, consider a given finite population of size N and a proposed design selecting n units. For each unit i in the finite population, consider the pair (Y_i, R_i), where Y_i is the unit characteristic of principal interest and R_i is a selection indicator, equal to one if unit i is selected by the design, and equal to zero otherwise. For a given selected sample s, let $\mathrm{E}_\xi(\cdot|s)$ and $\mathrm{V}_\xi(\cdot|s)$ denote moments evaluated with respect to the model, conditional on the specific selected sample s. In addition, let $\mathrm{E}_p(\cdot|U)$ and $\mathrm{V}_p(\cdot|U)$ denote moments evaluated with respect to the sample design, *conditional* on a given realization U of the finite population; let $\mathrm{E}_p(\cdot)$ and $\mathrm{V}_p(\cdot)$ denote moments evaluated with respect to the *unconditional* distribution of the R_i, and let $E_{p\xi}(\cdot)$ and $V_{p\xi}(\cdot)$ denote the expectation and variance evaluated with respect both the model and the design. Then standard double-expectation and variance results (e.g., Woodroofe, 1975, pp. 281-282) show that

$$E_{p\xi}(\hat{\theta}) = E_\xi\{E_p(\hat{\theta}|s)\} = E_p\{E_\xi(\hat{\theta}|s)\}; \tag{1.2}$$

$$V_{p\xi}(\hat{\theta}) = E_\xi\{V_p(\hat{\theta}|U)\} + V_\xi\{E_p(\hat{\theta}|U)\}; \tag{1.3}$$

and

$$V_{p\xi}(\hat{\theta}) = E_p\{V_\xi(\hat{\theta}|s)\} + V_p\{E_\xi(\hat{\theta}|s)\}. \tag{1.4}$$

Similarly, for the variance estimator $\hat{V}$,

$$E_{p\xi}(\hat{V}) = E_\xi\{E_p(\hat{V}|s)\} = E_p\{E_\xi(\hat{V}|s)\}; \tag{1.5}$$

$$V_{p\xi}(\hat{V}) = E_\xi\{V_p(\hat{V}|U)\} + V_\xi\{E_p(\hat{V}|U)\}; \tag{1.6}$$

and

$$V_{p\xi}(\hat{V}) = E_p\{V_\xi(\hat{V}|s)\} + V_p\{E_\xi(\hat{V}|s)\}. \tag{1.7}$$

1.2 Three types of simulation based performance evaluation criteria for complex sample analysis methods

The sample survey literature often uses simulation methods to evaluate some components of expressions (1.2) through (1.7). This simulation work tends to use one of the following approaches.

1.2.1 Design-based assessment with a known fixed finite population

One option is to restrict attention to a fixed finite population for which the Y_i values are known for all population units. In some cases (e.g., Rao and Bayless, 1969), the randomization moments can be evaluated directly from known population parameters. In other, more complicated cases, randomization moments (especially for small or moderate sample sizes) are less tractable, and generally are evaluated through simulated repeated sampling according to the specified design.

In some simulation work, the known finite population has been obtained from census data or is constructed by concatenation of several large sample datasets. In other cases (e.g., Hansen, Madow and Tepping, 1983, Section 2.2), the known finite population is generated as one set of N realizations of a specific superpopulation model ξ.

Under this option, the simulation work involves repeated sampling from the fixed finite population and thus leads to estimates of the design moments $E_p(\hat{\theta}|U)$, $V_p(\hat{\theta}|U)$, $E_p(\hat{V}|U)$ and $V_p(\hat{V}|U)$. These moments are all conditional on U. In that sense, they can be viewed as ξ-unbiased estimators of, respectively, $E_{p\xi}(\hat{\theta})$, $E_\xi\{V_p(\hat{\theta}|U)\}$, $E_{p\xi}(\hat{V})$ and $E_\xi\{V_p(\hat{V}|U)\}$.

1.2.2 Design-related assessment assuming distributional characteristics for lower-level estimators

A second option, which can be closely related to the first, is to assume that lower-level estimators follow a specified distribution. For example, due to central limit theorem arguments, in some cases it is plausible to assume that the primary sample unit estimators $\hat{\theta}_{hi}$ are approximately distributed as normal random variables with specified means and variances that may differ across h or i. Similar assumptions are sometimes used for the distribution of primary-unit-level variance estimators, or point estimators at a secondary-unit or finer level; see, e.g., Eltinge and Jang (1996, Section 5.2).

Under this approach, simulation work involves generating variates from the specified normal distributions and then assessing the resulting distributional characteristics of higher-level estimators like $\hat{\theta}$. This can be a tractable approach to evaluation of some properties of estimators $\hat{\theta}$ that are complicated nonlinear functions of the lower-level estimators. In a formal sense, if one considers the specified normal distribution of the primary-unit or secondary-unit level estimators to be induced solely by the sampling design, conditional on a given realization U of the superpopulation model ξ, then the simulation output is intended to describe the properties of the resulting distribution of $\hat{\theta}$, conditional on the realization U, e.g., $E_p(\hat{\theta}|U)$ or $V_p(\hat{\theta}|U)$. On the other hand, if the specified normal distributions of the primary-unit level estimators are considered to be induced by random variability associated with both the sample design and the superpopulation model, then the simulation output is intended to describe properties of the unconditional distribution of $\hat{\theta}$, e.g., $E_{p\xi}(\hat{\theta})$ or $V_{p\xi}(\hat{\theta})$.

1.2.3 Explicit use of a model for element-level characteristics

A third option is to restrict attention to one particular selected sample s and then generate simulated observations $\{Y_i, i \in s\}$ according to a conditional superpopulation model $\xi|s$. This leads to estimates of the moments $E_\xi(\hat{\theta}|s)$, $V_\xi(\hat{\theta}|s)$, $E_\xi(\hat{V}|s)$ and $V_\xi(\hat{V}|s)$, which in turn can be viewed as p-unbiased estimators of, respectively, $E_{p\xi}(\hat{\theta})$, $E_p\{V_\xi(\hat{\theta}|s)\}$, $E_{p\xi}(\hat{V})$ and $E_p\{V_\xi(\hat{V}|s)\}$. This third type of simulation work appears to be carried out fairly commonly in published studies of estimator performance, and in related work in statistical agencies.

1.3 Outline of main ideas

The remainder of this paper explores the framework of Section 1.2.3 in additional depth. Section 2 briefly reviews some ideas of design-based and model-based evaluation of complex survey data analysis methods. In this, we take a relatively detached approach to the "design based vs. model based" controversy, and view design-based, model-based and combined evaluation approaches as all offering potentially useful insights into the operating characteristics of a given analysis method.

Section 3 considers the assessment of variance estimator stability through estimation of the components of expressions (1.6) and (1.7). These components are estimated using data from one realization of the sample design (i.e., the one set of sample units actually selected for our survey); and from a large number of realizations of the model ξ for our set of selected sample units. Section 4 discusses some possible extensions and applications of the main ideas considered in this paper.

2. Operating Characteristic Surfaces Implied by Pure Randomization, Superpopulation and Mixed Approaches

2.1 Three classes of evaluation criteria

Evaluation of estimator performance with respect to distributions induced the randomized sample design p, the model ξ, or both, can be motivated in several related ways. First, in some cases an analyst has a high level of certainty that the observed data are consistent with a certain model or class of models ξ. For example, this prior modeling information may arise from previous studies, combined with careful model checking for the current data. For such cases, some statisticians prefer to evaluate estimator performance with respect to the model ξ alone. See, e.g., Royall and Herson (1973a, b), Royall and Cumberland (1978), and Scott et al. (1978) for discussion of these and related ideas. As is noted in some of this literature, a potential fundamental limitation of a purely model-based approach is that it may not provide information regarding the operating characteristics of a proposed estimation and inference method for cases in which the specified model ξ is not satisfied.

Second, at the other extreme an analyst may have little or no information about a plausible underlying model. In that case, a pure randomization approach provides a certain basic level of assurance regarding estimator properties (e.g., approximate p-unbiasedness of a point estimator), conditional only on the assumption that the randomization design was carried out as specified. See, e.g., Cochran (1953, Chapters 2 through 7) for a classical development of this idea. Also, some authors advocate careful examination of the randomization properties of estimation and inference methods, even if these methods were originally motivated by, or derived under, a specific model. See, e.g., Rubin (1987, pp. 117-118; 1996, p. 474); and related comments in Box (1980, Section 1.1).

In a sense, a pure randomization approach can generally be viewed as conservative, and this conservative characteristic can naturally be somewhat limiting. For example, randomization results may indicate that two point estimators or two variance estimators are each approximately unbiased; but with some exceptions, comparisons of their small-sample variances or general small-sample distributions require additional information. In particular, the relative magnitudes of two randomization-based variances may, in principle, depend heavily on special characteristics of the finite population in question.

Due to these limitations, some authors have chosen to evaluate estimator performance simultaneously with respect to *both* the randomization and the model ξ. For example, one may evaluate an expectation $\mathrm{E}_{p\xi}(\cdot)$, integrating with respect to both the randomization distibution and the ξ distribution. Similarly, one may evaluate the combined variance $V_{p\xi}(\cdot)$, with the associated components discussed in Section 1.1. For some general background on these and related ideas, see, e.g., Fuller (1975), Cassel et al. (1977), Särndal et al. (1992) and references cited therein.

2.2 Additional motivation for evaluation of $p\xi$ moments

In addition to the abovementioned formal mathematical motivation, the $p\xi$ approach is supported by the following informal arguments. First, viewed broadly, one may consider estimator operating characteristics (e.g., bias or variance) across a multidimensional space, with some dimensions determined by different realizations of different possible models ξ, and other dimensions determined by the particular set of sample units selected by the sample design. In essence, if we focus exclusively on design-based properties, we are conditioning on a specific realization of a given ξ model, and our evaluation focuses on "averages" evaluated across different possible sets of selected sample units. Similarly, evaluation of model-based properties amounts to conditioning on a given selected set of sample units, and examining "averages" evaluated over different possible realizations of the model ξ. In a sense, evaluation of $\mathrm{E}_{p\xi}(\cdot)$, $\mathrm{V}_{p\xi}(\cdot)$ and related $p\xi$-properties amounts to an attempt to obtain a somewhat less conditional characterization of estimator properties. In an informal sense, one might say that $p\xi-$ evaluation allows one to "average out" the "rough edges" (i.e., idiosyncratic characteristics) that can in principle arise in a given realization of the sample design and model ξ. However, this naturally entails some loss of information, e.g., problems with performance of a given estimator under extreme realizations of either the design p or of the model ξ. Similar concepts apply to performance characteristics of formal inference procedures, e.g., the size and power of hypothesis tests, or the mean width or coverage rate of confidence intervals.

Second, in practical applications (e.g. the health examination survey work that motivated this paper), one often collects a large number of variables (e.g., demographic characteristics; health knowledge, attitude and behavior; and anatomical and physiological measurements) for the persons selected through a single realization of a given complex sample design. Due to potentially large misspecification effects (e.g., Skinner, 1989), it is considered important to account for the design in the analysis of the data. However, due to the large number of analyses to be carried out, it is sometimes considered appropriate to, in essence, discuss estimator performance "averaged" over the large set of variables (or associated residual-type terms) of interest. If a given set of variables (e.g., several continuous anatomical and physiological measurements) are all believed to satisfy a given class of models ξ, then under mild conditions, "averaging over the variables" is closely related to integrating with respect to the ξ distribution. See, e.g., Cochran (1953, p. 169), and Des Raj (1958) for some related discussion. Also, similar ideas can be used to some degree to motivate the use of generalized variance functions (e.g., Wolter, 1985, Chapter 5) or "average design effects."

2.3 An analogy with time series signal processing

A review of the above-cited literature suggests that the simulation work described in Section 1.2.3 is loosely analogous to some time series analyses of complex signal-processing systems. Specifically, some signal-processing work introduces a "shock," with known characteristics, or a sequence of shocks, into the system. The subsequent system output is then used to make inferences regarding certain properties of the system itself.

A complex survey procedure (i.e., a sample design, along with associated point estimation, variance estimation and inference methods) can be analogous to a signal processing system in the following sense. First, conditional on s (and thus on the w_i and the specific forms of $\hat{\theta}$ and $\hat{V}(\cdot)$ viewed as functions of $\{Y_i, i \in s\}$), sample observations Y_i are generated according to the model $\xi|s$; the resulting observations $Y_i, i \in s$ are analogous to a specific system input signal. Second, the analysis results (e.g., realized values of the point estimators, variance estimators and confidence bounds) are analogous to signal-processing output. Third, consider the linkage of the conditional input distribution $\xi|s$ with the output distribution of analysis results (e.g., the distributions of $\hat{\theta}$, $\hat{V}$ or t statistics). For a given sample s, this linkage offers insight into one specific dimension of the operating characteristic surface of this survey procedure. This dimension is conceptually distinct from randomization (i.e., $p|U$) dimension of operating characteristics generally emphasized in design-based literature. Finally, under this loose analogy, one can view the sample design as generating a distribution of signal processing systems. Thus, $p\xi$ distributional characteristics (e.g., the $\mathrm{E}_{p\xi}(\cdot)$ and $\mathrm{V}_{p\xi}(\cdot)$ moments) can be interpreted as descriptions of this distribution of processing systems, now averaged with respect to the sample design.

2.4 Balanced interpretation of evaluation criteria

The remainder of the paper will take the relatively broad view that for both design- and model-derived analysis methods, it can be of serious interest to evaluate performance with respect to either the p or ξ or $p\xi$ distribution. In essence, one may consider these evaluation approaches to be largely complementary, with each offering insight into different dimensions of the overall operating characteristic surface for a proposed analysis method. One generally may seek to use methods that perform reasonably well under each criterion.

However, this view is often complicated by the fact that these evaluation criteria (especially when focused on variances and other efficiency measures) depend on the numerical values of assumed conditions, e.g., relevant superpopulation parameters or design features. This in turn highlights the importance of explicitly linking qualitative evaluation conclusions with assumed conditions. The practical information conveyed by this linkage naturally depends heavily on the size of the "neighborhood" of conditions relevant to a given survey analysis; and on the amount of curvature in these evaluation criteria within that neighborhood.

To interpret the resulting linkage, consider again the loose analogy to time series signal processing introduced in Section 2.3. In signal processing "high pass" and "low pass" filters are labeled and broadly understood to perform well on fundamentally different types of signal; and practical decisions regarding the choice of filtering method in a given application are accordingly conditional on the type of signal anticipated to be present. The present sample survey case is arguably more complicated due to qualitative differences between the p and ξ approaches. However, the signal-processing analogy may nonetheless be somewhat instructive, e.g., in its emphasis on conditioning ideas.

3. Properties of a Standard Variance Estimator

3.1 Additional notation and conditions

Note first that under mild regularity conditions, routine arguments show that the second component of expression (1.3) is negligible for large N. Consequently, if $\hat{V}(\hat{\theta}|U)$ is a design unbiased estimator of the design variance $V_p(\hat{\theta}|U)$, then $\hat{V}(\hat{\theta}|U)$ is approximately $p\xi$ unbiased for $\mathrm{V}_{p\xi}(\hat{\theta})$. Thus, the estimator $\hat{V}(\hat{\theta}|U)$ continues to be of intrinsic interest when we move attention from evaluation of design variances to evaluation of $p\xi$ variances.

Now consider the stability of the variance estimator $\hat{V}(\hat{\theta}|U)$. Note that for a given realization s of the sample design, one may use the model $\xi|s$ to generate R simulated sets $\{Y_{i(r)}, i \in s\}$ of sample observations. These observations in turn lead to R simulated point estimators $\hat{\theta}_{(r)}$ and variance estimators $\hat{V}_{(r)}$, say. Then the simulation mean $\tilde{E}_{R\xi}(\hat{V}_{(r)}|s) = R^{-1}\sum_{r=1}^{R}\hat{V}_{(r)}$ and variance $\tilde{V}_{R\xi}(\hat{V}_{(r)}|s) = (R-1)^{-1}\sum_{r=1}^{R}\{\hat{V}_{(r)} - \tilde{E}_{R\xi}(\hat{V}_{(r)}|s)\}^2$, say, converge with probability one to $E_{\xi}(\hat{V}|s)$ and $V_{\xi}(\hat{V}|s)$, respectively, as R increases.

Section 3.2 discusses ways to use $\tilde{E}_{R\xi}(\hat{V}_{(r)}|s\}$, $\tilde{V}_{R\xi}(\hat{V}_{(r)}|s\}$ and related quantities to construct estimators of the components of the variance decomposition (1.7). To simplify notation, we will restrict attention to the stratified two-per-stratum with-replacement design discussed in Section 1.1. In addition, we will assume that the sample design and superpopulation mechanism are independent, and that the survey items Y_i are independent and identically distributed with mean zero and constant variance σ^2.

3.2 Stability of a variance estimator

Now consider the two components of the variance decomposition (1.7), and note that the simulation variance

$$\tilde{V}_{R\xi}(\hat{V}_{(r)}|s) \tag{3.1}$$

is $p\xi$ unbiased for the first component, $E_p\{V_{\xi}(\hat{V}|s)\}$. In addition, conditional on s, note that the differences $\hat{Y}_{h1} - \hat{Y}_{h2}, h = 1, \ldots, L$, are distributed independently with common mean zero and variances $\sigma^2\sum_{j=1}^{2}\sum_{i\in s_{hj}} w_i^2$. Thus,

$$E_{\xi}(\hat{V}|s) = E_{\xi}\{\sum_{h=1}^{L}(\hat{Y}_{h1} - \hat{Y}_{h2})^2|s\} = \sigma^2\sum_{j=1}^{2}\sum_{i\in s_{hj}} w_i^2.$$

Routine arguments then indicate that under the design assumed in Section 1.1, a design unbiased estimator of $V_p\{E_{\xi}(\hat{V}|s)\}$ is

$$\sigma^4\sum_{h=1}^{L}(\hat{W}_{h1} - \hat{W}_{h2})^2 \tag{3.2}$$

where $\hat{W}_{hj} = \sum_{i\in s_{hj}} w_i^2$. Consequently, given known σ^2, R simulation replications of ξ and one realization of the sample design s, the sum of expressions (3.1) and (3.2) is a design unbiased estimator of $V_{p\xi}(\hat{V})$. Thus, for cases in which the abovementioned assumptions are reasonable, relatively simple simulation methods allow assessment of the stability of the variance estimator $\hat{V}$. In addition, separate computation of the two terms (3.1) and (3.2) provides an indication of the relative contributions of the design and model to the overall variability of $\hat{V}$.

Finally, consider the components of expression (1.6). Arguments similar to those at the start of Section 3.1 indicate that under the assumptions stated above and additional mild regularity conditions, the second term of expression (1.6), $V_{\xi}\{E_p(\hat{V}|U)\}$ is negligible relative to the magnitude of the first term. To evaluate that first term, assume that conditional on U, the difference $\hat{\theta}_{h1} - \hat{\theta}_{h2}$ is distributed as a normal random variable with mean zero and variance $V_h = V_p(\hat{\theta}_{h1} - \hat{\theta}_{h2}|U)$; cf. Section 1.2.2 above. Then routine arguments show that $V_p(\hat{V}|U) = 2\sum_{h=1}^{L} V_h^2$; and that a p-unbiased estimator of $V_p(\hat{V}|U)$ is

$$3^{-1}2\sum_{h=1}^{L}\hat{V}_h^2, \tag{3.3}$$

where $\hat{V}_h = (\hat{\theta}_{h1} - \hat{\theta}_{h2})^2$ as in Section 1.1. This leads to two results. First, given a single realization of the superpopulation model $\xi|s$, expression (3.3) gives a $p\xi$ unbiased estimator of the dominant term in the $p\xi$ variance of $\hat{V}$. Note especially that this can be computed directly from the true observations actually collected in an actual survey, without use of simulations. Second, suppose that for each of R simulation realizations of the model $\xi|s$, we compute the value $V_{(r)}^*(\hat{V}|U)$, say, of expression (3.3). Then the average $R^{-1}\sum_{r=1}^{R} V_{(r)}^*(\hat{V}|U)$ of these variance estimates gives a more stable estimate of the dominant term of expression (1.6). Thus, provided one is interested in evaluation of the dominant term in expression (1.6) rather than the conditional variance $V_p(\hat{V}|U)$, use of $\xi|s$ simulation work can help to improve the stability of the stability measure itself. This can be of serious interest because stability measures such as expression (3.3) are functions of fourth moments, and thus are themselves subject to stability problems, especially in applications involving a relatively small number L of strata, or involving severe heterogeneity of the stratum-level variances V_h.

4. Discussion

In closing, we note several possible extensions of the ideas considered here. First, the current paper has restricted attention to the simple case in which the element-level Y_i are independent and identically distributed random variables, independent of the design selection indicators R_i. However, extensions of Section 3 to cases involving dependence between the selection indicators R_i and the observations Y_i can offer insight into the links between the $p\xi$ properties of a variance estimator and the informativeness of a sample design. This is of special interest because nontrivial relationships between the design and the superpopulation model can have an important practical impact on the properties of variance estimators; see, e.g., Korn and Graubard (1997) and references cited therein.

Second, the present work has focused primarily on variance estimator performance. In many applications, variance estimation is of interest primarily as as intermediate step in construction of t-type test statistics or confidence intervals for θ. Consequently, in keeping with simulation work carried out by several previous authors, it is useful to extend Sections 3 through 5 to evaluate the approximate distribution of t statistics under the p, ξ and $p\xi$ approaches.

Third, in work not detailed here for reasons of space, we have applied some of the proposed methods to interview and examination data from the U.S. Third National Health and Nutrition Examination Survey (NHANES III). In particular, empirical results of this application work offered some useful insights into the relative performance of competing NHANES III variance estimators.

Acknowledgements

Work by J.L. Eltinge was supported in part by funding from the National Center for Health Statistics. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the National Center for Health Statistics.

References

Box, G.E.P. (1980). Sampling and Bayes' Inference in Scientific Modelling and Robustness (with Discussion). *Journal of the Royal Statistical Society, Series A* **143**, 383-430.

Cassel, C.M., Särndal, C.-E. and Wretman, J.H. (1977). *Foundations of Inference in Survey Sampling.* New York: Wiley.

Cochran, W.G. (1953). *Sampling Techniques.* (First Edition). New York: Wiley.

Des Raj (1958). On The Relative Accuracy of Some Sampling Techniques. *Journal of the American Statistical Association* **53**, 98-101.

Eltinge, J.L. and Jang, D.S. (1996). Stability Measures for Variance Component Estimators Under a Stratified Multistage Design. *Survey Methodology* **22**, 157-165.

Fuller, W.A. (1975). Regression Analysis for Sample Survey. *Sankhyā, Series C* **37**, 117-132.

Hansen, M.H., Madow, W.G. and Tepping, B.J. (1983). An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys (with discussion). *Journal of the American Statistical Association* **78**, 776-807.

Korn, E.L. and Graubard, B.I. (1997). Variance Estimation for Superpopulation Parameters. Unpublished manuscript, Biometric Research Branch, National Cancer Institute, Bethesda, Maryland.

Rao, J.N.K. and Bayless, D.L. (1969). An Empirical Study of the Stabilities of Estimators and Variance Estimators in Unequal Probability Sampling of Two Units Per Stratum. *Journal of the American Statistical Association* **64**, 540-569.

Royall, R.M. and Cumberland, W.G.. (1978). Variance Estimation in Finite Population Sampling. *Journal of the American Statistical Association* **73**, 351-357.

Royall, R.M. and Herson, J. (1973a). Robust Estimation in Finite Populations I. *Journal of the American Statistical Association* **68**, 880-889.

Royall, R.M. and Herson, J. (1973b). Robust Estimation in Finite Populations II: Stratification on a Size Variable. *Journal of the American Statistical Association* **68**, 890-893.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Rubin, D.B. (1996). Multiple Imputation After 18+ Years. *Journal of the American Statistical Association* **91**, 473-489.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling.* New York: Springer-Verlag.

Scott, A.J., Brewer, K.R.W. and Ho, E.W.H. (1978). Finite Population Sampling and Robust Estimation. *Journal of the American Statistical Association* **73**, 359-361.

Skinner, C.J. (1989). Introduction to Part A. Pp. 23-58 in *Analysis of Complex Surveys*, C.J. Skinner, D. Holt and T.M.F. Smith, eds. New York: Wiley.

Wolter, K.M. (1985). *Introduction to Variance Estimation.* New York: Springer-Verlag.

Woodroofe, M. (1975). *Probability with Applications.* New York: McGraw-Hill.