Resampling Variance Estimates for Complex Survey Designs
A Simulation Study

William H. Robb
Macro International Inc., 126 College St., Burlington, VT 05401

Key Words: Resampling, School Surveys, Cluster Sampling

I Introduction

Sampling designs for large scale, complex surveys often induce a non i.i.d. structure to the data through such techniques as sampling without replacement, stratification, and multi-stage or unequal probability selection. Although variance estimation techniques do exist for these designs, they are often prohibitively complex to implement, or do not extend to the more complex designs. The most commonly applied methods include balanced repeated replication, and the linearization, or taylor series method implemented in the commercially available SUDAAN package.

There is a growing body of work investigating the application of resampling methods to complex survey data. Early work on the problem can be found in [3,6,7]. These methods are attractive in that they build on existing estimation methods, substituting computer power for theory. These methods are relatively simple to implement, and increasingly practical with the advent of cheep and plentiful computing horsepower. Recent work [5,7,10,11] has extend the bootstrap to stratified, two stage cluster sampling.

The purpose of this study is apply one of these recently developed bootstrapping methods to a synthetic population, that although simplified, represents some of the complexities of a "real life" problem. It is hoped that this will provide a better understanding of the applicability of these methods. Secondly the performance of the bootstrap will be compared with the to the taylor series method of variance estimation.

The synthetic population created is intended to approximate in structure the national population of high school students in terms of PSU and school clustering. The sampling design preserves some of the main elements of the design employed in the Youth Risk Behaviors Survey conducted yearly by the CDC Division of Adolescent and School Health.

II Variance Estimation Methods

Bootstrap

Several extensions of the bootstrap to complex survey designs can be found in the literature. These are the Bootstrap Without Replacement (BWO) [1], the Rescaling method [8], and the Mirror-Match method [10, 11]. Note that the Mirror-Match method includes the Bootstrap With Replacement [7] as a special case. Each method attempts to balance the two factors. The first is the use of the original sampling mechanism (in terms of preserving the sampling fractions and dependency at each stage) for resampling. This allows the bootstrap to generate estimates of arbitrary parameters for an arbitrary design, giving the bootstrap it's power. The second is the drawing of a resample sized so that the estimate of variance is unbiased. Each method is discussed briefly below.

In the BWO, a pseudo population is created by replicating the sample data vector. This pseudo population is then resampled without replacement, taking the bootstrap sample size n^* to be n, the size of the sample. This method produces biased estimates. However this can be corrected for in simple cases.

The Rescaling bootstrap resamples a general sample size m^* from the sample with replacement. Sample values are then rescaled so that the resulting variance estimate matches the usual variance estimate in the linear case. This method has a computational disadvantage as each data point must be rescaled at each iteration. Further this method requires summary statistics by sub-populations for each estimate be calculated for use in computation of the rescaling factors.

The Mirror-Match method is the focus of this paper. This method preserves the features of the original sampling mechanism, in that the resampling is done without replacement using the original sampling fractions (mirror) at each stage. Variances are matched by then repeating the sub-sampling at each stage k times. Thus, a bootstrap sample at each stage is composed of k independent replicates of a sample that mimics the original sampling as completely as possible.

Taylor Series Linearization

The linearization method is a well know method for variance estimation. This study used the method as implemented in the SUDAAN software package. The reader is referred to the SUDAAN technical appendix [12] for details of the method and it's implementation.

III Simulation Study with Synthetic Population

Description of Study Population

The synthetic population used in this study is built to mimic a the sampling frame for the YRBS study. This frame is based of NCES enrollment data for a national population of approximately 21,000 high schools. In this frame groups of counties serve as PSU, schools as SSU and students as the final sampling unit. The frame is stratified at the PSU level.

Keeping the structure of strata/PSU/SSU/Student, a synthetic population was generated containing 12 psu in 2 strata. Each PSU contained on the average five SSU, and each SSU contained on the average 100 students. These figures approximate the SSU and student frame counts for the 12th grade portion of the NCES frame.

Generation of the response data

Each student was assigned a gender, with 50% male and 50% female within school. Responses were then randomly generated to three Yes/No questions. As particular intra-cluster correlation was of particular interest, question one was generated with a overall percentage "yes" of 50%, but with a moderately high intra-PSU correlation (rho = .2) The generated response to the question two was a uniform 10% across the population. Finally, the response to question three was generated with a uniform response of 20% for females and 60% for males.

Description of Methodology

Sample Design

The sample design under study is a stratified, three stage cluster design. PSU are stratified and sampled without replacement. Schools are selected without replacement as the second stage clusters. Students are selected without replacement as the final stage.

The sampling fractions used at the PSU level were 0.5 and 0.75. As in the YRBS study, the number of schools sampled in a PSU is fixed, and f allowed to vary. For this simulation, two SSU were selected from each PSU, with f averaging around 0.333 Students were selected with a sampling fraction of 0.4.

Statistics of Interest

Of primary interest is the variance of percentages within subpopulations. While generally expressed as a percentage, this is actually a ratio estimator. Six such estimators were used in this study; the percentage responding yes to the three generated questions within subpopulations defined by gender.

Study Design

The study consisted of drawing independent samples from the population described above. For each sample, variances were estimated for the statistics of interest using the mirror-match bootstrap, and SUDAAN. For the bootstrap estimates 100 resamples were used for each estimate. The MSE was estimated using a separate simulation of 1000 independent samples. Results are reported in terms of relative stability and bias of the estimates, computed as outlined in [10].

Results

Table 1 presents the results of the simulation, listing for the each question within gender the population percentage responding yes, the MSE as computed above, and the relative bias and instablilty of the variance estimates for each method. Figure 1 presents a plot of the variances estimated for each percentage estimate by method, with the method labeled MSE the population MSE.

The bootstrap method yielded estimates with consistently higher than SUDAAN estimates in all cases. Of note are the two cases in which SUDAAN yeilded a negative bias.

The SUDAAN method produced estimates were more stable that the bootstrap method for all but one case. The difference in stability was not consistent, being negligible for question one, and considerable for on question three. An examination of figure 1 reveals that the variance estimates are distributed in an approximately similar fashion for both methods.

IV Conclusion

The higher bias in the bootstrap method would suggest the use of the SUDAAN for estimating variance in for this particular sampling design. However, for designs that SUDAAN cannot handel, the bootstrap would be a viable alternative.

The differences in performance by each method across estimates suggests that the relative performance of both methods should be evaluated against a wider variety of response distributions.

The issue of stability in the bootstrap estimates could be addressed by using a higher number of bootstrap replications.

Discussion of this effort with Dr. Sitter indicates that the performance of the Mirror-Match method would improve for sampling designs employing larger sampling fractions, and that the rescaling methods might be better suited for this application. Either method might be adapted to the PPS sampling used in the YRBS study.

The author plans to pursue these lines of investigation, and welcomes inquires about future progress or collaborative efforts.

References

1. Bickel, P. J. ; Freedman, D. A. Asymptotic Normality and The Bootstrap in Stratified Sampling. The Annals of Statistics; 1984; 12(No 2): 470 - 482.

2. Dawson, A. C. Efficient Bootstrap Simulation. Biometrika; 73: 555 - 561.

3. Effron, B. The Jackknife, the Bootstrap, and Other Resampling Plans. Philadelphia: Society for Industrial and Applied Mathematics; 1982.

4. Johnson, E. G. Considerations and Techniques for Analysis of NAEP data. Journal of Education Statistics; 1989; 14.

5. Kovar, J. G., ; Rao, J. N. K., ; Wu, C. F. Bootstrap and Other Methods to Measure Errors in Survey Estimates. The Canadian Journal of Statistics; 1988; 16(Supplement).

6. McCarthy, P. J. Replication; An Approach to the Analysis of Data From Complex Surveys. Vital and Health Statistics; 1979; Series 2(No 14); ISSN: DHEW Publication No. (PHS) 79 - 1269.

7. McCarthy, P. J; Snowden, C. B. The Bootstrap and Finite Population Sampling. Vital Health Statistics; 1985; Series 2(No 95); ISSN: DHHS Pub No. (PHS) 85 - 1369.

8. Rao, J. N. K. ; Wu, C. F. J. Resampling Inference With Complex Survey Data. Journal of the American Statistical Association; 1988; 83(No 401).

9. Rust, K. F.; Johnson, E. G. Sampling and Weighting in the National Assessment. Journal of Education Statistics; 1992; 17: 111 - 129.

10. Sitter, R. R. Comparing Three Bootstrap Methods for Survey Data. The Canadian Journal of Statistics; 1992; 20(No 2).

11. Sitter, R. R. A Resampling Procedure for Complex Survey Data. Journal of the American Statistical Association; September 1992; 87(419).

12 Shaw, B.V, et al Statistical Methods and Mathematical Algorighms Used in SUDAAN; Research Triangle Institute, Research Triangle Park, NC

Table 1

Simulation Results							
Estimate		Population Values		Relative Bias		Relative Instability	
Q	Gender	P	MSE	SUDAAN	Bootstrap	SUDAAN	Bootstrap
1							
	Male	0.50	0.00416	-0.081	0.032	0.076	0.079
	Female	0.53	0.00395	0.112	0.202	0.105	0.112
2							
	Male	0.09	0.00050	-0.059	0.085	0.084	0.106
	Female	0.10	0.000573	0.072	0.162	0.122	0.127
3							
	Male	0.19	0.00082	0.153	0.337	0.138	0.157
	Female	0.59	0.00133	0.071	0.169	0.114	0.081

Figure 1

Variance Estimates by Question/Gender/Method