PROPERTIES OF THE SCHOOLS AND STAFFING SURVEY'S BOOTSTRAP VARIANCE ESTIMATOR

Steven Kaufman, National Center for Education Statistics
Room 422d, 555 New Jersey Ave. N.W., Washington, D.C. 20208

Key Words: Simulation, Half-Sample Replication, Bootstrapping, Variances

Introduction

The National Center for Education Statistics' (NCES) Schools and Staffing Survey (SASS) conducted by the Census Bureau has a complex sample design. Public schools are selected using a stratified systematic PPS (unequal selection probabilities) sample design. From this design, data are collected at the school and school district level. The school district is an aggregation unit (i.e., the district selection probability is computed by aggregating school selection probabilities containing the district across the school strata). The probability is nonlinear with respect to the school sample sizes. A bootstrap variance estimator (Kaufman,93) has been developed that provides better variance estimates than the balanced half-sample replication (BHR) variance estimator for the SASS public school district component. A bootstrap variance estimator for the other SASS components was presented in 1994 (Kaufman,94). The bootstrap variance estimators reflects the finite population correction associated with the SASS high sampling rates, without using the joint inclusion probabilities. **A set of bootstrap replicate weights are generated that work like BHR replicate weights, so that the bootstrap variances can be generated from any BHR variance software package.** It has also been shown that the bootstrap variance estimator performs better than BHR with other designs with high sampling rates (Kaufman, 94). This bootstrap variance estimator has been implemented into the 1994 SASS survey.

The goal of this paper is to provide results from simulation studies that demonstrate the bootstrap variance estimator (Kaufman, 94) works better than BHR with designs with low sampling rates. In addition, a balanced bootstrap will be presented that works better than the non-balanced bootstrap variance estimator.

First, a motivation why the bootstrap variances estimator may perform better than BHR is presented. Next, the balanced and non-balanced bootstrap variance estimators are described, as well as, the BHR estimators. The methodology presented here is the same as what is presented in (Kaufman, 94), except for the balancing of the bootstraps. A description of the designs being tested in this study follows. Finally, the results are presented showing the bootstrap variance estimator's superiority for the designs tested.

Motivation

BHR assumes 2 PSUs are selected with replacement within each stratum. To fit PPS systematic sampling of n PSUs into this model, sampled PSUs are paired by the order in which they were selected. Each pair is then treated as a stratum for variance estimation (variance stratum). If a systematic equal probability sample of size 10 is selected from a frame of 100 PSUs, the BHR model would have more than 10 trillion possible samples. In reality, there are only 10 possible systematic samples. Without, further homogeneity assumption, BHR can be a very large overestimate, even if the sampling rates are low. For this example, since the bootstrap selection is done systematically, approximately 10 possible bootstrap-samples can be selected from the bootstrap frame with each randomization of the bootstrap frame. Unlike the BHR estimator, a homogeneity assumption does not appear to be required; so the bootstrap estimator may get closer to the true variance.

Public and Private School-Bootstrap Frame

The idea behind the bootstrap samples is to use the sample weights (W_i) from the selected units to estimate the distribution of the school frame. From the estimated bootstrap-school frame, B bootstrap samples can be selected. The bootstrap-school frame is generated in the following manner:

For each selected school i, W_i bootstrap-schools (bi) are generated. If W_i has a noninteger component then a full school is generated with a reduced selection probability and weight. As shown in the bootstrap weighting section, the bootstrap expectation of the bootstrap weights (W_{bi}) equals the full-sample weight (W_i). The bi^{th} bootstrap-school has the following measure of size (m_{bi}):

$$m_{bi} = I_{bi} * 1/W_i,$$

$$I_{bi} = \begin{cases} 1 & \text{if bi is an integer component of } W_i \\ C_i & \text{if bi is a noninteger component of } W_i, \\ & C_i \text{ being the noninteger component} \end{cases}$$

Bootstrap Sample Size

The bootstrap sample size is usually chosen to provide unbiased variance estimates. When the original sample is a simple random sample of size n then Efron (1982) shows a bootstrap sample size should be n-1. Sitter (1990) has computed the bootstrap sample size for the Rao-Hartley-Cochran method for PPS sampling. A variation of this result is used in this simulation. Sitter's bootstrap sample size (n^*) is the sample size which makes the following quantity closest to 1:

$$\left(\sum_{g=1}^{n^*} (N_g^{*2}-N^*)\right)/\left(\sum_{g=1}^{n}(N_g^2-N)\right)*\left(N^2-\sum_{g=1}^{n} N_g^2\right)/\left(N^* * (N^*-1)\right)$$

n^*: is the bootstrap stratum sample size
g: represents a sampling interval in the stratum
N_g^*: is the number of bootstrap-schools in the g^{th} sampling interval, where the bootstrap-schools are in a random order
n: is the sample size in the stratum
N^*: is the number of bootstrap-schools in the stratum
N : is the number of schools in the stratum
N_g: is the number of schools in the g^{th} sampling interval, where the schools are in their original order; either a random order for the Rao-Hartley -Cochran method or the specific nonrandom order for the SASS method

n^* can not be calculated directly. The quantity above is computed for each n^* from n-20 to n. The n^* that is closest to one is used in the bootstrap selection.

The variation to Sitter's formulation is in the computation of N_g^* and N_g. Two modifications are made. The first occurs when I_{bi} is not equal to 1. Instead, of using 1, as Sitter does when counting units; I_{bi} is used to calculate N_g^*. The second modification is due to the fact that a school or bootstrap-school can be in two sampling intervals. When this happens, N_g and N_g^* are not increased by one. Instead, they are increased by the proportion of the unit that actually goes into the sampling interval. If I_{bi} does not equal 1, and the bootstrap-school is in two sampling intervals then N_g^* is increased by the product of the two modifications described above.

Determining the Bootstrap Sort Order

If the bootstrap variance estimate is to work correctly, it is important that the school-bootstrap frame be randomized in an appropriate manner. In one extreme, when the bootstrap frame is sorted by the order of selection from the original sample and n^*=n, the variance estimate will be zero. In the other extreme, when the bootstrap frame is sorted randomly, the variance estimate ignores the original ordering and may overestimate the variance. Bootstrap variances will be computed using a number of sort orderings for each of the simulation samples. Coverage rates are computed for each ordering. The coverage rates are compared with estimates of the true coverage rate. The ordering associated with the coverage rates closest to the true coverage rates is the ordering that is used for the bootstrap estimator. These comparisons are made at a level where the coverage rates should have some degree of stability given the number of simulations. For the designs in this study, the comparisons are made at the general school association/region level. The bootstrap sort orders are described below.

School Sort Method j

Selected schools within a stratum are sorted by order of selection. Next, schools are consecutively paired within each stratum. Each pair is assigned a random number. The bootstrap-schools generated within each pair of schools are assigned bootstrap-school random numbers. If $n-n^* \leq j$, for a stratum, the bootstrap-schools are sorted by bootstrap-school random number. If $n-n^* > j$, for a stratum, the bootstrap-schools are first sorted by the school pair random number; within each school pair the bootstrap-schools are sorted by the bootstrap-school random number. In other words, if the difference between the original and bootstrap sample sizes is small, as defined by j, then ignore the original sort ordering when randomizing the bootstrap-schools. Otherwise, randomize within pairs that reflect the original sort ordering.

The bootstrap program used in these simulations requires an initial bootstrap sort. Given this sort, the program searches for the sort that minimizes the maximum absolute bias in the average, total and ratio coverage rates. If the maximum absolute bias is less than or equal to 0.07 for a bootstrap sort, then that bootstrap sort will be used as the final sort for the association. Otherwise, the program tries other logical sorts. After the sort searching has finished, the coverage rate biases are reviewed and a final bootstrap sort is determined for each general association/region group.

Rationale for School Sort Method j

Sitter shows that if the number of schools in a sampling interval is constant across the intervals, then n^* will be close to n-1. If schools are sorted randomly, then the expected number of schools in the intervals is constant and n^* should be close to n-1. Therefore, if n^*=n-1, the assumption is that the sort ordering is effectively random, so that the school pairing should be ignored. Sort method j=1, sorts bootstrap schools randomly if n^*=n-1. The smaller n^* is relative to n-1, the more effective the ordering is (i.e., the ordering acts less like a random ordering) and the more important the school pairings are to the sort method. Again, this is the affect of sort method j, when j is small.

When the pairings are ignored, a bootstrap-school generated for a particular school is in more sampling intervals and therefore can be selected more often. All other things kept equal, this should increase the bootstrap variance estimate. One then expects the variance from sort method j to be ≥ the variance from sort method k, when j ≥ k. This rule can be used to determine which sort to use to improve the variance estimate. The rule, however, does not always work. This might be due to random error or to the implicit bootstrap-school joint inclusion probabilities that are generated. The coverage rate from a particular sort that matches the true coverage rate is implicitly: 1) matching the effective randomness of the original sort (sort method j=1), adding variability as necessary (sort method j > 1), as well as, 3) matching the bootstrap-school joint inclusion probabilities to the true school joint inclusion probabilities.

Bootstrap Sample Selection

Given the bootstrap frame, m_{bi} as the measures of size, stratum bootstrap sample sizes and bootstrap-school ordering, select the bootstrap sample using the same sampling scheme as in the original sample. The bootstrap frame is randomized with each sample selection. Bootstrap-schools, generated from noncertainty schools, with measures of size larger than the sampling interval are not removed from the sampling process. If a bootstrap-school is selected more than once, the bootstrap-school weight is multiplied by the number of times it is selected.

Balanced and Non-Balanced Bootstraps

Since systematic sampling gives good sample size control by values of the first sort variable, the variance estimate may be improved if the bootstrap samples have the same control (balance). This can be achieved by ordering the bootstrap frame by the first sort variable. Then, the bootstrap-schools can be randomized as described above within each of the values of the first sort variable. If the first sort variable is continuous there may not be enough bootstrap-schools within the sort variable's values to accurately estimate the variance using balanced bootstraps. In this situation, it might help to categorize the sort variable. Both balanced and non-balanced bootstrap variances will be presented.

Number of Replicates and Bootstraps

Since the old SASS BHR variances are based on 48 replicates, 48 bootstrap samples are computed for each simulation sample. Given the time it take to select a set of bootstrap samples, only 60 simulation samples are used.

Bootstrap Weights

The bootstrap-school weight, W_{bi}, is:

$$W_{bi} = I_{bi} * M_{bi}/p_{bi}$$

M_{bi}: is the number of times the bi^{th} bootstrap-school is selected

p_{bi}: is the bootstrap selection probability for the bi^{th} bootstrap-school

$$E_*(\sum_{bi} W_{bi})=\sum_{bi} I_{bi} =\sum_{i} W_i, \text{ as desired.}$$

E_*: is expectation over the bootstrap samples

Since the available data are defined by the schools selected in the original sample, a bootstrap-school weight indexed by i (BW_i) is required:

$$BW_i = \sum_{bi \in S_{iB}} W_{bi}$$

S_{iB}: is the set of all bi∈i selected in the B^{th} bootstrap sample.

Balanced Half-sample Replicates

The r^{th} school half-sample replicate is formed using the usual textbook methodology (Wolter, 1985) for establishment surveys with more than 2 units per stratum. Since the SASS half-sample variances are based on 48 replicates, the simulations will be based on 48 half-sample replicates.

Three BHR variance estimates will be presented based on the methodology described above. The first **(BHR no FPC)** is the variance estimates described above. This estimate does not make any type of Finite Population Correction (FPC) adjustments.

The other two make simple FPC adjustments. The second BHR variance estimate **(BHR Prob FPC)** adjusts the first variance estimator by $1\text{-}P_h$, where P_h is the average of the selection probabilities for the selected units within stratum h.

The third BHR variance estimate **(BHR SRS FPC)** adjusts the first variance estimator by $1\text{-}n_h/N_h$, where n_h is the number of sample units in stratum h and N_h is the number of units on the frame in stratum h.

Low Sampling Rate Design I and II

The sample frame is the list frame component of NCES's Private School Survey (PSS). The list frame is stratified by general School Association (4 groups), within Association by Census Region (4 levels), and within Region by school level (elementary, secondary and combined). The school sample is selected using a systematic probability proportionate to size sampling procedure. The design I uses square root teachers as the measure of size, while design II uses teachers. Before sample selection, the school frame is sorted by Urbanicity.

Sample Estimate

For each of the simulation samples, totals, averages and ratios are computed. The variables used are all on the sample frame. Two averages, one ratio and three totals are computed using estimated schools, teachers and students. For each of the 60 simulation samples, the sample estimates and respective sample variances are computed. An estimate of the true variance for the sample estimates can be obtained by computing the simple variance of the sample estimates across the 60 simulations. The bootstrap and BHR sample variance can now be compared with the estimate of the true variance.

When determining the bootstrap sort order estimates are computed within general school association/region. The estimates used in the tables are publishable estimates other than the ones used determining the sort order. To maintain stability given only 60 simulation are used, the samples used in the tables are the same as thoses used to determine the sort order.

The analysis statistics used to evaluate the variance estimates is described below.

Analysis Statistics

Coverage Rates

To measure the accuracy of the variance estimates, a one sigma two-tailed coverage rate is computed by determining what proportion of the time the population estimate is within the respective confidence interval. If the estimates are approximately normal then the coverage rates should be close to 0.68.

Coverage Rate Bias (**Bias**)

$$\text{Bias} = R_e - R_t$$

R_e: is the coverage rate based or either a bootstrap or BHR variance estimate

R_t: is an estimate of the true coverage rate, based on the simple variance of the simulation estimates

The distribtion of the coverage rate bias will be presented two ways. The first way, looks at the distribution of various publishable estimates implied by the sample design, this treats each publishable estimate equally. The second way, sums independent sets of the publishable total variances to produce a overall national variance for a estimated total. This method gives larger totals more weight than smaller totals and allows for variance underestimates to cancel out variance overestimates. In other words, the second way, provides a method of judging how well the variance procedures works when estimates are aggregated to produce new estimates.

Results - Coverage Rates by Publishable Estimates

Design 1 (Tables 1-3)

BHR No FPC variance estimates can be very large overestimates (**BIAS GE .14**); 35, 37 and 11 percent of the simulated estimates are in this category respectively for averages, totals and ratios. These coverage rates are closer to what one would expect from a two sigma coverage rate than from a one sigma coverage rate. Applying simple FPC adjustments helps somewhat, but the FPC adjusted BHR estimates still have between 6 to 29 percent in the very large overestimate category.

The tables shows the balanced bootstrap has 8, 4 and 0 percent in the very high overestimate category, respectively for averages, totals and ratios. This is much better than any of the BHR estimators. The not balanced bootstrap estimator has 16, 35 and 0 percent in this category respectively for average, totals, and ratios. The 35%, for the not balanced totals comes from estimates whose domain are not functions of the stratification variables (e.g., urbanicity and region/urbanicity). Once the bootstrap sample sizes are controlled on urbanicity (balanced bootstraps) the 35% drops to 4%.

For averages and totals, the BHR No FPC has the fewest number of estimates in the low bias category ([-.07,0.0) and [0.0, .07) categories); 30 and 14 percent, respectively for averages and totals. Applying simple FPC adjustments helps only marginally. Within the small bias category the bootstrap estimators perform better than BHR for averages and totals. The balanced bootstrap has 57 and 61 percent, in the low bias category, respectively for averages and totals. The not balanced bootstrap has 61 and 53 percent, in the low bias category, respectively for averages and totals.

For ratio estimates, the BHR estimators performs better than the bootstrap estimators. The BHR estimators have between 61 and 72 percent in the low bias category, while the bootstrap estimators have 53 and 58 percent in this category.

The bootstrap estimators are the only estimators that have a few estimates in the very large underestimate category (BIAS LT-.14). These coverage rates are closer to what one would expect from a .5 sigma coverage rate than from a one sigma coverage rate. The balanced bootstrap has 2, 2 and 6 percent in this category, respectively for averages, totals and ratios. The not balanced bootstrap has 10% in this category for ratios.

Design II (Tables 5-7)

Tables 5-7 don't give either BHR or bootstrap a strong advantage.

The only estimator that has a large problem in the very large overestimate category is the not balanced bootstrap, with 20% for totals. The 20% comes from estimates whose domain are not functions of the stratification variables (e.g., urbanicity and region/urbanicity). Once the bootstrap sample sizes are controlled on urbanicity (balanced bootstraps) the 20% drops to 0%. All estimators have some estimates in the very large overestimate category, but usually just a few percent.

For totals and ratios, the BHR No FPC has the fewest number of estimates in the low bias category ([-.07,0.0) and [0.0, .07) categories); 59 and 72 percent, respectively for totals and ratios. Applying simple FPC adjustments helps. For BHR SRS, 67 and 76 percent are in this category, respectively for totals and ratios. For BHR Prob, 74 percent are in this category, for both totals and ratios. Within the small bias category the bootstrap estimators performs about as well as the BHR Prob and SRS estimators, for totals and ratios. The balanced bootstrap has 74 and 76 percent, in the low bias category, respectively for totals and ratios. The not balanced bootstrap has 70 and 82 percent, in the low bias category, respectively for totals and ratios.

For averages, the BHR estimators performs better than the bootstrap estimators. The BHR estimators have between 55 and 65 percent in the low bias category, while the bootstrap estimators have 43 and 51 percent in this category.

Results - Coverage Rates for Overall Estimates

The results presented above treat each estimate equally. The results in this section, provide a measurement of how well the estimators work when estimates are aggregated. Table 4 shows that for designs I and II, the bootstrap coverage rate biases are much smaller than the BHR coverage rate biases. The bootstrap biases are between 0.8 and 2.6 percent, while the BHR biases are between 3.8 and 7.3 percent. The no balanced bootstrap biases are the lowest. This indicates, since the main purpose of the balancing is to improve variance estimate for domains defined by the first sort variable, that if the only variances required are where the domain is defined by the stratification then the no balanced bootstrap is better than the balanced bootstrap.

Conclusions

The overall conclusion is that the bootstrap methods are better than the BHR methods for the designs in this study. How much better one method is than another depends on the sample design and the estimates of interest.

Coverage Rates for Published Estimates

These coverage rates treat each estimate equally.

For design I, using square root teachers as the measure of size, all BHR procedures have serious problems overestimating the variance. The balanced bootstrap procedure has a much smaller problem overestimating the variance. The no balanced bootstrap procedure overestimate the variance a large percent of the time when estimating averages and totals. For totals, this overestimation is caused from domains that are not function of the sampling strata.

For design II, using teachers as the measure of size, **BHR Prob** and the balanced bootstrap are comparable. The no balanced bootstrap procedure overestimate the variance a large percent of the time when estimating totals, but the the overestimation is caused from domains that are not function of the sampling strata.

Coverage Rates for Overall Total Estimates

These estimates provide a measure of how well the different variances work when aggregaing estimates. For both designs, the no balanced bootstrap is better than the balanced bootstrap. This indicates, if the only variances required are where the domain is defined by the stratification then the no balanced bootstrap is better than the balanced bootstrap. Both bootstrap methods are superior to all of the BHR methods.

References

Efron, Bradley(1982). The Jackknife, the Bootstrap and Other Resampling Plans. SIAM No. 38, p62.

Kaufman, Steven(1993). A Bootstrap Variance Estimator for the Schools and Staffing Survey. ASA 1993 Survey Research Methods Proceedings.

Kaufman, Steven(1994). Properties of the Schools and Staffing Survey's Bootstrap Variance. ASA 1994 Survey Research Methods Proceedings.

Sitter, R.R.(1990). Comparing Three Bootstrap Methods for Survey Data. Technical Report Series of the Laboratory for Research in Statistics and Probability, No. 152, p9-10.

Wolter, K. M.(1985). Introduction to Variance Estimation. New York: Springer-Verlag, p110-145.

Table -- 1 Publishable Estimate Dist. of Coverage Rate Bias for Averages using Private Design I

Bias	Bootstrap		BHR Estimates		
Averages (% Freq.)	Bal	No Bal	Prob	SRS	No FPC
LT -.14	2	0	0	0	0
[-.14,-.07)	2	2	2	2	0
[-.07,0.0)	14	20	12	12	8
[0.0,.07)	43	41	29	25	22
[.07,.14)	31	21	39	43	35
GE .14	8	16	18	18	35

Table -- 2 Publishable Estimate Dist. of Coverage Rate Bias for Totals using Private Design I

Bias	Bootstrap		BHR Estimates		
Totals (% Freq.)	Bal	No Bal	Prob	SRS	No FPC
LT -.14	2	0	0	0	0
[-.14,-.07)	0	0	0	0	0
[-.07,0.0)	14	20	6	6	4
[0.0,.07)	47	33	26	22	10
[.07,.14)	33	12	41	43	49
GE .14	4	35[1]	27	29	37

Table -- 3 Publishable Estimate Dist. of Coverage Rate Bias for Ratios using Private Design I

Bias	Bootstrap		BHR Estimates		
Ratios (% Freq.)	Bal	No Bal	Prob	SRS	No FPC
LT -.14	6	10	0	0	0
[-.14,-.07)	29	22	10	10	8
[-.07,0.0)	27	27	33	31	20
[0.0,.07)	26	31	39	41	41
[.07,.14)	12	10	12	12	20
GE .14	0	0	6	6	11

Table -- 4 Average Coverage Rate Bias from Overall Estimates of Totals Generated from Independent Groups by Design and Variance Estimator

Percent	Bootstrap		BHR Estimates		
Private Design	**Bal**	**No Bal**	**Prob**	**SRS**	**NO FPC**
Design I	0.9	0 .8	4.0	4.1	6.1
Design II	2.6	1.9	3.8	5.4	7.3

Table -- 5 Publishable Estimate Dist. of Coverage Rate Bias for Averages using Private Design II

Bias	Bootstrap		BHR Estimates		
Averages (% Freq.)	Bal	No Bal	Prob	SRS	No FPC
LT -.14	4	6	4	0	0
[-.14,-.07)	18	14	18	21	8
[-.07,0.0)	14	10	18	12	24
[0.0,.07)	29	41	47	43	33
[.07,.14)	31	25	11	18	25
GE .14	4	4	2	6	10

Table -- 6 Publishable Estimate Dist. of Coverage Rate Bias for Totals using Private Design II

Bias	Bootstrap		BHR Estimates		
Totals (% Freq.)	Bal	No Bal	Prob	SRS	No FPC
LT -.14	4	0	2	0	0
[-.14,-.07)	10	8	12	8	4
[-.07,0.0)	31	35	25	24	14
[0.0,.07)	43	25	49	43	45
[.07,.14)	12	12	10	23	33
GE .14	0	20[1]	2	2	4

Table -- 7 Publishable Estimate Dist. of Coverage Rate Bias for Ratios using Private Design II

Bias	Bootstrap		BHR Estimates		
Ratios (% Freq.)	Bal	No Bal	Prob	SRS	No FPC
LT -.14	0	0	0	0	0
[-.14,-.07)	12	10	20	10	4
[-.07,0.0)	53	45	41	45	25
[0.0,.07)	23	37	33	31	47
[.07,.14)	8	6	4	12	20
GE .14	4	2	2	2	4

---

[1] The increase bias relative to the balanced bootstrap is due to the bias in estimates that are not functions of the stratification variables (i.e., urbanicity and region/urbanicity). Balancing the bootstrap samples by urbanicity reduces the bias.