A BOOTSTRAP VARIANCE ESTIMATOR FOR THE SCHOOLS AND STAFFING SURVEY

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

Keywords: Half-Sample Replication, Bootstrapping

Introduction

The National Center for Education Statistics' (NCES) School and Staffing Survey (SASS) conducted by the Census Bureau has a complex sample design. 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. It has been demonstrated (Kaufman,1992) under the usual Balanced Half-Sample (BHR) sample design that the BHR variance estimator for these district estimates can overestimate the variance. The apparent reason for the bias in the BHR estimator is that the district variances decrease faster than the inverse of the sample size, which BHR assumes. Since the bootstrap variance estimator doesn't necessarily make this assumption, this simulation study investigates whether a bootstrap variance estimator can perform better than the BHR variance estimator.

Another aspect of this paper is to investigate whether the bootstrap variance estimator reflects the finite population correction generated from the SASS sample design without using the joint inclusion probabilities. If independent systematic samples are selected, using the original sample design, then the simple variance of the estimates produced for each of the samples will reflect the appropriate variance. In this situation, units with selection probabilities close to one will appear in each sample more often then units with smaller selection probabilities. Since the bootstrap variance estimator mimics this process better than the BHR variance estimator, it might provide a better variance estimate, when the sampling rates are large.

The goal of this paper is to investigate, using a simulation study, whether a bootstrap variance estimator: 1) provides better variance estimates than the BHR estimator when estimates are based on aggregation units (school districts); and 2) reflects the impact of large sampling rates better than BHR. **The proposed bootstrap variances can be computed using any BHR program without any modifications.**

The SASS sample design for schools and school districts will be used for the simulation. The SASS district sample design will be used to study goal 1. Since the SASS is designed to produce State estimates, the sampling rates in small States are high; therefore, the SASS is a good design to demonstrate goal 2. Since the SASS sample design sorts the frame in a specific nonrandom order, four methods of sorting the bootstrap frame will be tested in the simulations.

Design of SASS School and District Surveys

The school survey uses NCES's public school Common Core of Data file as the frame. The frame is stratified by State, and within State by school level (elementary, secondary and combined). The school sample is selected using a systematic probability proportionate to size sampling procedure. The measure of size is the square root of the number of teachers in the school. Before sample selection, the school frame is sorted by a specific nonrandom order. The school districts that include a sampled school comprise the school district sample. In order to simplify the computation of the district selection probabilities, it is important, within each stratum, to keep schools belonging to the same district together.

Weighting

The school weight for school i (W_i) is:

$$W_i = 1/p_i$$

p_i: is the selection probability for school i.

The district weight for district d (W_d) is:

$$W_d = 1/(1-(1-p_{de})(1-p_{ds})(1-p_{dc}))$$

p_{de}: $\sum_{i \in S_{de}} p_i$

S_{de}: the set of all elementary schools in district d

p_{ds}: $\sum_{i \in S_{ds}} p_i$

S_{ds}: the set of all secondary schools in district d

p_{dc}: is $\sum_{i \in S_{dc}} p_i$

S_{dc}: is the set of all combined schools in district d

If p_{de}, p_{ds} or p_{dc} is greater than or equal to one then the district is selected with certainty and $W_d = 1$.

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. The r^{th} district half-sample replicate is defined to be the set of districts that have schools in the r^{th} school half-sample replicate. Since the SASS half-sample variances are based on 48 replicates, the simulations will be based on 48 half-sample replicates.

The school replicate weight is:

$$RW_i = 2/p_i .$$

The district replicate weight is:

$$RW_d = 1/(1-(1-p_{de}/2)(1-p_{ds}/2)(1-p_{dc}/2))$$

The probabilities are divided by 2 because with half the sample, each school has half the chance of being selected.

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 $1-P_h$, where P_h is the average of the selection probabilities for the selected units within stratum h. For the district P_h's h represents a State.

The third BHR variance estimate (**BHR SRS FPC**) adjusts the first variance estimator by $1-n_h/N_h$, 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. For the district adjustments h represents a State.

School-bootstrap Frame

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

For each selected school i and associated district d, W_d bootstrap-districts (bd) are generated, as well as, W_i/W_d bootstrap-schools (bi) within each bootstrap-district. If W_d or W_i/W_d have a noninteger component then a full school is generated with a reduced selection probability and weight. As shown below, the bootstrap expectation of the bootstrap weights (W_{bi} or W_{bd}) equals the full-sample weight (W_i or W_d). The bi^{th} bootstrap-school has the following measure of size (m_{bi}):

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

$$I_{bd} = \begin{cases} 1 & \text{if bd is an integer component of } W_d \\ C_{bd} & \text{if bd is a noninteger component of } W_d, \\ & \quad C_{bd} \text{ being the noninteger component} \end{cases}$$

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

The sum of the m_{bi}s, generated from a selected school, equals one; so one bootstrap-school would be selected to represent school i, provided the bootstrap stratum sample size and sort order are the same as in the original design.

Each bootstrap-school, bi, generated within a bootstrap-district, bd, has the bd^{th} bootstrap-district's id. If the d^{th} district has selected schools in the elementary and secondary strata then the bd^{th} bootstrap-district id generated in the elementary stratum should match to the bd^{th} bootstrap-district id in the secondary stratum. This relationship should exist for all school levels that are selected for the district. This is important to compute the appropriate bootstrap-district weights.

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 Efon (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. The Sitter's bootstrap sample size (n^*) is the sample size which make 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-10 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 either I_{bd} or I_{bi} are not equal to 1. Instead, of using 1, as Sitter does when counting units; $I_{bd} * I_{bi}$ is used to calculate N_g^*. To reduce the incidence of $I_{bd} * I_{bi}$ being not equal to 1, the districts are ignored when determining n^*. This is accomplished by generating a bootstrap frame as described above, assuming $W_d = 1$ (i.e., W_d never has a noninteger component). 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 either I_{bd} or I_{bi} are not equal to 1, and the bootstrap-school is in two sampling intervals then N_g^* is increased by the product of the two modifications described above. If n is large, n^* should not be affected much by these modifications.

Sorting the School-Bootstrap Frame

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. Four orderings will be tested in this simulation study.

Sort Method 1

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-districts and bootstrap-schools generated within each pair of schools are assigned bootstrap-district and bootstrap-school random numbers, respectively. Bootstrap-schools are sorted by the school pair random number; within each pair, bootstrap-schools are sorted by the bootstrap-district random number; and within the bootstrap-district, the bootstrap-schools are sorted by the bootstrap-school random number.

Sort Method 2

If the weights are relatively uniform within the set of paired schools, method 1 may underestimate the true variance. Sort method 2, tries to adjust for this. Sitter (1990) shows when the sample weights are uniform that his n^* will equal n-1. Hence, for this simulation, when n^* is between n and n-2, it will be assumed the stratum weights are relatively uniform and sort method 1 may underestimate the true variance. Instead, the bootstrap-schools are sorted by the bootstrap-district random number; and within the bootstrap-district, the bootstrap-schools are sorted by the bootstrap-school random number. If $n^* < n-2$, for a stratum, then the bootstrap-schools are randomized as described in sort method 1.

Sort Method 3

Sort method 3 is the same as sort method 2, except that the weights are assumed to be uniform when n^* is between n and n-3, instead of sort method 2's n and n-2. In this case, the bootstrap-schools are sorted by the bootstrap-district random number; and within the bootstrap-district, the bootstrap-schools are sorted by the bootstrap-school random number. If $n^* < n-3$, for a stratum, then the bootstrap-schools are randomized as described in sort method 1.

Sort Method 4

Sort method 4 does not use the school pairings; instead, bootstrap-schools are placed in a district and school random order. With this sort, the bootstrap-schools are sorted by the bootstrap-district random number; and within the bootstrap-district, the bootstrap-schools are sorted by the bootstrap-school random number.

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 must be randomize 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.

Number of Replicates and Bootstraps

Since the 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:

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

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

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

E_*: is expectation over the bootstrap samples

Since the available data is defined by the districts 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 $bi\in i$ selected in the B^{th} bootstrap sample.

The bootstrap-district weights, W_{bd} is:

$$W_{bd} = I_{bd}/(1-(1-p_{bde})(1-p_{bds})(1-p_{bdc}))$$

p_{bde}: is $\sum_{bi\in S_{bde}} p_{bi}$

S_{bde}: is the set of all elementary bootstrap-schools in bootstrap-district pd

p_{bds}: is $\sum_{bi\in S_{bds}} p_{bi}$

S_{bds}: is the set of all secondary bootstrap-schools in bootstrap-district bd

p_{bdc}: is $\sum_{bi\in S_{bdc}} p_{bi}$

S_{bde}: is the set of all combined bootstrap-schools in bootstrap-district bd

If p_{bde}, p_{bds} or p_{bdc} is greater than or equal to one then the bootstrap-district is selected with certainty and $W_{bd} = 1$.

$E_*(\sum_{bd} W_{bd})=\sum_{bd} I_{bd} =\sum_{d} W_d$, as desired.

Since the available data is defined by the districts selected in the original sample, a bootstrap district weight indexed by d (BW_d) is required:

$$BW_d = \sum_{bd\in S_{dB}} W_{bd}$$

S_{dB}: the set of all $bd\in d$ selected in the B^{th} bootstrap sample.

Sample Estimate

For each of the simulation samples, totals, averages and ratios are computed within a number of the States and the District of Columbia, using variables available on the sample frame. For district samples, two averages are computed using teachers and schools; two ratios are computed using students, teachers and schools; and five totals are computed using students, teachers, graduates, schools and districts. For the school samples, two averages are computed using teachers and students; one ratio is computed using students and teachers; three totals are computed using students, teachers and schools. For each of the 60 simulation samples, the sample estimates and respective sample variances are computed for both district and school samples. 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.

A number of other analysis statistics are used. They are 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 variance estimates are appropriate then the coverage rates should be close .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. For a given estimator, it is based on the simple variance of the simulation estimates for that estimator

Tables 1-6 presents the coverage rate Bias's.

CV of Variance Estimate (**CV**)

To measure the variability of the variance estimate, the coefficient of variation (CV) of the variance estimate is calculated.

$$CV = [(1/59) \sum_{t=1}^{60} (V_t - \bar{V})^2]^{1/2} / \bar{V}$$

V_t: is the variance estimate for the t^{th} simulation estimate,

$\bar{V}$: is the average variance estimate across the 60 simulation samples.

Table 7 presents the CV of the variance estimates averaged across the States included in the study.

Results

Due to the time to complete the simulations, simulations for 4 large States (more than 2,000 schools) did not include bootstrap sort 1 or sort 2. First, tables 1-6 are discussed which are based on the 25 States in the simulations. The worst variance estimator is **BHR no FPC**. A large percent of the time the one σ coverage rates are better 2σ coverage rates than one σ coverage rates (i.e., Bias GE 0.14). The worst case is in table 5 with 68% of the estimates being better 2σ coverage rates than one σ coverage rates. One reason for this is because the sampling rates are very high in some States. The other two BHR variance estimate are better; but in 4 out of the 6 tables, there are still a reasonable number of estimates that are better 2σ coverage rates. In table 2, 24% of the estimates are better 2σ coverage rates. In general, the BHR variances tend to be overestimates.

An additional problem with the two FPC adjusted BHR variance estimates is that a number of the coverage rates are better $.5\sigma$ coverage rates than one σ coverage rates (i.e., Bias LT -0.14). The worst cases are found in table 4, where the Prob and SRS adjusted estimates have 60% and 36% of the coverage rates being better $.5\sigma$ coverage rates, respectively.

The best bootstrap variance estimator is the **bootstrap sort 4** estimator, with the **bootstrap sort 3** estimator a close second. There are still some coverage rates that are better 2σ coverage rates, but now the worst case is table 2 with 16%. The bootstrap variances for school estimates do tend to be underestimates, while district estimates tend to be overestimates. However, except for school ratios, the **bootstrap sort 4** estimator appears to be better than any of the BHR estimators. For school ratios, **BHR prob FPC** or **BHR SRS FPC** appear to be best. Some of the sort 4 estimates are better $.5\sigma$ coverage rates, but except for school ratios, the BHR FPC adjusted estimates are still worst overall with respect to this point. The worst **bootstrap sort 4** coverage rates are in table 3 (school ratios) with 20% being better $.5\sigma$ coverage rates. However, the absolute bias of the standard errors for these 20%, averages less than -0.04. Since the -0.04 bias is so small, even for school ratios **bootstrap sort 4** performs well.

If there is a desire to make an FPC adjustment for large sampling rates, the **bootstrap sort 4** appears to be the best variance estimator from those tested. However, if the desire is to always provide a conservative variance estimate then the **BHR no FPC** is the most conservative.

The major drawback with Bootstrap variances is that the calculation of the bootstrap replicate weights is far more complicated and computer intensive than the calculation of BHR replicate weights. However, this work only needs to be done once. Given the bootstrap weights, any BHR variance program can compute the bootstrap variance estimates, without any special adjustments. The bootstrap weights use most of the sample cases in each replicate, so when computing variances for ratios, there is not as much need to worry about zero denominators, as is the case with BHR variances.

When the sampling rates are lower one expects the **BHR No FPC** to provide good results. Although not presented here, this is true for the States in this study with low sampling rates. For these States, the **bootstrap sort 4** also provides good results, especially for school estimates.

Table 7 presents the CV of the variance estimates. For the most part, the BHR CV's are smaller than the Bootstrap CV's. However, the differences are small. For practical purposes, BHR and Bootstrap CV's are the same. One reason for this result is that the BHR replicates are only partially balanced.

References

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

Kaufman, Steven(1992). Balanced Half-Sample Replication with Aggregation Units. ASA 1992 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 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for School Averages by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	38.10	9.52	0.00	0.00	8.00	8.00	8.00
[-0.14., 0.0)	52.38	61.90	56.00	56.00	16.00	16.00	4.00
[0.0, 0.14)	9.52	23.81	32.00	32.00	64.00	56.00	44.00
GE 0.14[2]	0.00	4.76	12.00	12.00	12.00	20.00	44.00

Table 2 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for School Totals by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	42.86	9.52	8.00	4.00	8.00	8.00	0.00
[-0.14., 0.0)	33.33	57.14	44.00	52.17	12.00	8.00	12.00
[0.0, 0.14)	23.81	23.81	36.00	28.00	56.00	60.00	48.00
GE 0.14[2]	0.00	9.52	12.00	16.00	24.00	24.00	40.00

Table 3 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for School Ratios by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	66.67	52.38	20.00[3]	20.00[3]	0.00	0.00	0.00
[-0.14., 0.0)	33.33	38.10	64.00	64.00	48.00	48.00	16.00
[0.0, 0.14)	0.00	9.52	16.00	16.00	44.00	44.00	60.00
GE 0.14[2]	0.00	0.00	0.00	0.00	8.00	8.00	24.00

[1] Coverage rates in this category are better .5σ coverage rates than σ coverage rates

[2] Coverage rates in this category are better 2σ coverage rates than σ coverage rates

[3] The absolute bias of the standard errors in these States, averages less than -0.04

Table 4 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for District Averages by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	19.05	19.05	16.00	16.00	60.00	36.00	0.00
[-0.14., 0.0)	38.10	38.10	32.00	32.00	24.00	32.00	20.00
[0.0, 0.14)	33.33	33.33	44.00	44.00	16.00	28.00	44.00
GE 0.14[2]	9.52	9.52	8.00	8.00	0.00	4.00	36.00

Table 5 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for District Totals by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	14.29	19.05	16.00	20.00	32.00	16.00	0.00
[-0.14., 0.0)	52.38	52.38	44.00	40.00	60.00	48.00	4.00
[0.0, 0.14)	28.57	23.81	36.00	36.00	8.00	36.00	28.00
GE 0.14[2]	4.76	4.76	4.00	4.00	0.00	0.00	68.00

Table 6 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for District Ratios by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	0.00	0.00	0.00	0.00	28.00	4.00	0.00
[-0.14., 0.0)	42.86	42.86	32.00	36.00	56.00	56.00	0.00
[0.0, 0.14)	57.14	52.38	64.00	60.00	16.00	40.00	40.00
GE 0.14[2]	0.00	4.76	4.00	4.00	0.00	0.00	60.00

Table 7 -- CV of the Variance (**CV**) by Type of Estimate and Type of Variance Estimator

	CV Type of Variance	Type of Estimate: School AVE	School RATIO	School TOTAL	District AVE	District RATIO	District TOTAL
B	No FPC	0.31	0.32	0.31	0.40	0.28	0.37
H	SRS FPC	0.31	0.32	0.32	0.43	0.35	0.38
R	PROB FPC	0.31	0.32	0.32	0.43	0.35	0.42
B	Sort 4	0.30	0.35	0.32	0.50	0.36	0.35
O	Sort 3	0.31	0.35	0.33	0.52	0.36	0.36
O	Sort 2	0.34	0.36	0.36	0.52	0.34	0.36
T	Sort 1	0.33	0.39	0.36	0.53	0.34	0.35