VARIANCE ESTIMATION FOR THE SPECIAL STUDIES OF THE NATIONAL ASSESSMENT FOR EDUCATIONAL PROGRESS (NAEP)

Hyunshik Lee, Westat, Jae-Kwang Kim, Hankuk University of Foreign Studies, Sylvia Dohrmann, Westat
Hyunshik Lee, Westat, 1650 Research Boulevard, Rockville, MD 20850

KEY WORDS: Multi-phase sampling; Clustering, Double expansion weight, Ratio adjustment, Weight Winsorization, Poststratification.

1. Introduction

The multi-phase sample design is often employed in sample surveys for various reasons. It has a long history, first introduced by Neyman (1938). Traditionally, the technique is used to collect some auxiliary data that are not available for the sampling frame from a large first-phase sample to use the data at the second-phase sampling. The technique, sometimes called "double sampling" has many applications in different forms (e.g., Rao, 1973; Cochran, 1977; Breidt and Fuller, 1993; Rao and Sitter, 1995; Hidiroglou and Särndal, 1998; Fuller, 1998).

In this paper, we focus on variance estimation for multi-phase sampling with stratified PPS sampling at the first phase and multi-stage cluster sampling at the second phase. Particularly, we are interested in the jackknife technique for variance estimation. The jackknife method has considerable advantage of being flexible in incorporation of series of weight adjustments often applied to sample survey data.

Rao and Shao (1992) proposed a consistent jackknife variance estimator for the reweighted expansion estimator (REE) in the context of hot deck imputation treating the respondents as the second-phase sample. Kott and Stukel (1997) considered the same problem and concluded that the jackknife variance estimator works well for the reweighted expansion estimator (REE) if the first-phase sampling is with replacement. Rao and Sitter (1997) considered a case of two-phase sampling with nontrivial sampling rate, where the first-phase strata are the same as the second-phase strata. Binder et al. (2000) studied the variance estimation problem for a similar two-phase sample design but without the restriction of the with-replacement sampling assumption. However, they used the Taylor linearization method. Also, Kim, Navarro, and Fuller (2000) tackled the same problem using the jackknife variance estimator but with with-replacement sampling assumption at the first-phase. Lee and Kim (2002) considered a similar case but with without-replacement sampling assumption at the first-phase.

The NAEP special studies used stratified two-phase design and at both phases, a complex design was employed. In this paper, our main concern is how to reasonably estimate the variance incorporating this design feature and complicated weighting procedures that include base sampling weights, nonresponse adjustment, weight truncation, and post-stratification.

In the next section, the sample design of the special studies are detailed. In Section 3, we describe the weighting procedures and define the estimator. Section 4 presents the proposed variance estimator, which is followed by discussion in Section 5.

2. Sample Design of the Special Studies

The two special studies used two different grade student samples: grade 4 for the oral reading study and grade 8 for the writing online study.

The sample design for each of the studies was basically two-phase sample design but a complex design was used at both phases. A similar design was used for the two study samples so, in the following we describe the basic features of the sample design without referring to the grade or the study.

A school sample was first drawn as a subsample of the NAEP main study sample. For the main study, the grade schools were first stratified by public and private schools, for which very different designs were used.

The public schools were further stratified by states and independent state samples were selected with the goal of producing state by state estimates. From this a national subsample that has a representation from each state was taken. The private school sample was a national sample from the beginning but it was also subsampled to be used for special studies. The two national samples (public and private) were then combined to form a single national sample, which can safely be regarded as the 1st-phase sample.

There was a concern of cost to administer such a wide spread sample because it involves testing of students at each sample school. To address this concern, it was decided to use geographically compact clusters of schools. Such clusters were formed from the combined national sample disregarding public and private distinction and then a sample of clusters was selected by PPS systematic sampling with the measure of size (MOS) being defined by the number of 1st-phase sample schools in the cluster. Because of this sample based cluster formation and sampling, we should treat the clustering and subsequent sampling as 2nd-phase sampling.

The sampled clusters at the 2nd-phase were further subsampled independently and student samples were then selected within each school in the final school sample. This part of the second-phase sampling design is itself three-stage sampling, where the first-stage is cluster sampling of schools, the second-stage is subsampling of schools from the sampled clusters, and the third-stage is student sampling within sampled schools, all nested as in the usual multi-stage sampling.

3. Special Studies Weighting and Estimator

The final weights were developed in two steps: first the school weights and then the student weights.

3.1 School Weighting

The base school sampling weights are simple double expansion weights, namely the first phase school weights are multiplied by the second-phase school weights. To the resulting weights a ratio adjustment was then applied to improve the estimate as well as to make a nonresponse adjustment using the school grade enrollment as the auxiliary variable. The adjustment cells were formed by cross-classes defined by public/private, four NAEP Regions (similar to the Census regions), and three classes of school location. Small cells were collapsed. These adjustment cells are indexed by *l*, which is numbered from 1 to *L*. The set of sample schools in class *l* is then denoted by C_l. Other notation we need to define the ratio adjustment is listed below. This notation is defined with variance estimation by the jackknife method in mind.

- h: index for the variance strata, $h = 1, \ldots, H$;
- i: index for the variance units within a variance stratum, i = 1 or 2;
- j: index for the sampled schools within a variance unit;
- C_l: set of sample schools in ratio adjustment class *l*, $l = 1, \ldots, L$;
- A_S: set of sample schools in the second phase sample;
- x_{hij}: student enrollment for school *j* in variance unit *i* in variance stratum *h*;
- X_l: class population total of *x*-values for ratio adjustment class *l*;
- w_{hij}: the 1st-phase sampling weight for school *j* in variance unit *i* in variance stratum *h*;
- w^*_{hij}: the 2nd-phase school weight for school h_{ij} that includes both cluster sampling and within-cluster subsampling; and
- a^*_{hij}: the school nonresponse adjustment factor for school h_{ij} if respondent, = 0 otherwise.

The variance strata and variance units are formed at the 1st-phase sample with the restriction that H must be less than or equal to 62, and there are two variance units in each variance stratum. This results in putting more than one school in a variance unit.

The 1st-phase school weights are defined straightforwardly as the inverse of the first-phase school selection probability.

Now, the second-phase school weight is defined. Let there be M clusters of schools and without loss of generality, the first m clusters be assumed selected by PPS sampling. The measure of size (MOS) is determined by the number of schools in the cluster. Further, let p_k be the subsampling probability within *k*-th sampled cluster, $k = 1, ..., m$. However, most of time, the subsampling rate is one. Then the 2nd-phase sampling weight for school h_{ij} that belongs to cluster *k* is defined by

$$w^*_{hij} = \frac{\sum_{k'=1}^{M} \text{MOS}_{k'}}{m\text{MOS}_k} \times \frac{1}{p_k},$$

where

$$\text{MOS}_k = \sum_{h=1}^{H} \sum_{i=1}^{2} \sum_{j=1}^{n_{hi}} I_{hij}(k).$$

The indicator variable $I_{hij}(k)$ takes a value of 1 if school h_{ij} belongs to cluster *k*, a value of 0 otherwise. Note that variance unit h_i has n_{hi} schools.

It would have been possible to use different weighting cells for the ratio adjustment and school nonresponse adjustment but it was decided to use the same classes for both and, thus, the two steps can be combined. This combination in effect eliminates the nonresponse adjustment because the nonresponse adjustment factor, which should be the same within each class, would be cancelled out. As a result, the ratio and nonresponse adjusted full sample school weights for on $h'i'j' \in C_l$ are given by

$$\omega_{h'i'j'} = X_l \frac{w_{h'i'j'} w^*_{h'i'j'} a^*_{h'i'j'}}{\sum_{hij \in A_S \cap C_l} w_{hij} w^*_{hij} a^*_{hij} x_{hij}} .$$

3.2 Student Weighting

The student weights were then developed starting with the school weights derived above by applying student sampling weights and host of other adjustments including student nonresponse and poststratification adjustments.

3.2.1 Student Base Weights

In order for a student to be selected into the special study, the student must have been selected for the main NAEP study, and selected for the special study. The student base weight is thus the product of a number of weights and factors shown below:

1. Special study school weight $\left(\omega_{hij}\right)$;
2. Main NAEP within-school student selection weight;
3. School substitution adjustment factor;
4. Year-round adjustment factor;
5. Student session factor;
6. Subject spiral adjustment factor (2 for the special studies); and
7. Inverse of the student within-school probability of selection into the special study.

The student weights such obtained were then adjusted for nonresponse.

3.2.2 Student Nonresponse Adjustment

Only students having enough data to be used for analysis (as determined by Education Testing Services, which is an agency that analyzes the NAEP data) were considered respondents to the special study. Students who were excluded from the main study were considered ineligible for the special study. All other students (including absent students) were considered nonrespondents to the special study.

Student nonresponse adjustment cells were formed using the relative age of the student (whether or not the student is old relative to the other students) and student race (in 6 categories). If the cells formed by crossing of these two variables contained fewer than 10 students (or 5 students for the replicates), or the adjustment factor exceeded 3, the cell was combined with an adjacent cell.

3.2.3 Weight Winsorization

Extreme student weights were Winsorized (truncated or frequently called trimmed) using the multiple median rule, which is used for State NAEP. In the outlier literature, trimming means to throw away an extreme observation, while Winsorization means to modify the outlying value to an acceptable value. Thus, Winsorization is more appropriate than trimming in this context.

In this method, Winsorization groups are defined, and the median of the full-sample student nonresponse adjusted weight is calculated within each Winsorization group. Any full-sample student nonresponse adjusted weight in the Winsorization group greater than a multiple of the median is Winsorized. Within each Winsorization group, the Winsorization factor calculated for the full-sample weight is also applied to the replicate weights.

All students in private schools were in one Winsorization group while students in public schools were in four Winsorization groups defined by NAEP region. The Winsorization factors ranged from 2 to 3.

3.2.4 Poststratification

The resulting weights were then further adjusted through poststratification to reduce the mean squared error of the estimate. In this procedure, student nonresponse adjusted and Winsorized weights are further adjusted so that their sums for various subgroups are equal to the composite estimates of the same subgroups derived from several years of the Current Population Survey (CPS) data. These same totals were used for other NAEP 2002 studies.

The poststratification cells were formed using student race (in 4 categories), age (in 2 categories), and

Census region (in 4 categories for the white/other race category only) for a total of 14 cells. If the adjustment factors for a cell exceeded 2, the cell was combined with an adjacent cell.

3.2.5 Total Estimator

An estimator $\hat{Y}$ of the total of a *y*-variable of interest is given by

$$\hat{Y} = \sum_{h=1}^{H} \sum_{i=1}^{2} \sum_{j=1}^{n_{hi}} \sum_{k \in S_{hij}} \omega_{hij} \xi_{hijk} y_{hijk} ,$$

where S_{hij} is the special study student sample from school h_{ij}, ξ_{hijk} is the student weight factor in the final student weight, and y_{hijk} is *y*-value for student *k* in sample S_{hij}.

4. Jackknife Variance Estimator

To estimate the variance of this estimator, the JK2 jackknife method is used. This method assumes that there are two variance units per variance stratum and one replicate is randomly created from each variance stratum. Therefore, the number of replicates is equal to the number of the variance strata. It is further assumed that the variance units are independently sampled. The variance stratum is different from the design stratum but is created in such a way that the sample design is properly reflected in variance estimation. There are *H* (actually 62) replicates, each corresponding to a variance stratum, and the variance estimator is given by

$$v_J\left(\hat{Y}\right) = \sum_{r=1}^{H} \left(\hat{Y}^{(r)} - \hat{Y}\right)^2 ,$$

where $\hat{Y}^{(r)}$ is the *r*-th replicate estimate, which is defined by

$$\hat{Y}^{(r)} = \sum_{h=1}^{H} \sum_{i=1}^{2} \sum_{j=1}^{n_{hi}} \sum_{k \in S_{hij}} \omega_{hij}^{(r)} \xi_{hijk}^{(r)} y_{hijk} .$$

Once replicated ω-weights are obtained, the replication of the ξ-weight factor could be done in a usual manner and thus, not discussed here.

The calculation of the replicated ω-weights starts with the 1st-phase school weights. Define 1st-phase replicate school weights by

$$w_{hij}^{(r)} = \begin{cases} 0 & \text{if } h = r \text{ and } hi \text{ is randomly chosen} \\ & \quad \text{to be deleted} \\ 2w_{hij} & \text{if } h = r \text{ and } hi \text{ is randomly chosen} \\ & \quad \text{to be kept} \\ w_{hij} & \text{if } h \neq r \end{cases}$$

and then the 2nd-phase replicate weight $w_{hij}^{*(r)}$ is calculated by

$$w_{hij}^{*(r)} = \frac{\sum_{k'=1}^{M} \mathrm{MOS}_{k'}^{(r)}}{m \mathrm{MOS}_{k}^{(r)}} \frac{1}{p_k}$$

with

$$\mathrm{MOS}_k^{(r)} = \sum_{h=1}^{H} \sum_{i=1}^{2} \sum_{j=1}^{n_{hi}} w_{hij}^{(r)} w_{hij}^{-1} I_{hij}(k) .$$

Note that the indicator variable $I_{hij}(k)$ is random, so is MOS_k because it depends on the random 1st-phase sample. Thus, this randomness should be properly reflected in the 2nd-phase replicate weights as shown above. The school nonresponse adjustment factor can be replicated in the usual manner; that is the replicate school nonresponse adjustment factor $a_{hij}^{*(r)}$ is calculated by using $w_{hij}^{*(r)}$. Then the replicated ω-weights are given by

$$\omega_{h'i'j'}^{(r)} = X_l \frac{w_{h'i'j'}^{(r)} w_{h'i'j'}^{*(r)} a_{h'i'j'}^{*(r)}}{\sum_{hij \in A_S \cap C_l} w_{hij}^{(r)} w_{hij}^{*(r)} a_{hij}^{*(r)} x_{hij}} .$$

The rest of replication (i.e., replication of ξ-weight factor) can be proceeded in a usual manner.

5. Discussion

The jackknife method has considerable advantage over the Taylor method when the weighting procedure is very complex such as the NAEP special study discussed in this paper. However, incorporation of multi-phase sampling in the jackknife method has been somewhat elusive until recently. Considerable progress has been made as briefly surveyed in the introduction and still unfolding as discussed in this paper and others presented in this session. However useful they may be, unfortunately, we cannot show the results of the actual implementation in the NAEP special studies because the NAEP data have not been officially released.

6. References

Binder, D.A., Babyak, C., Brodeur, M., Hidiroglou, M.A., and Jocelyn, W. (2000). Variance estimation for two-phase stratified sampling. *The Canadian Journal of Statistics*, 28, pp. 751-764.

Breidt, K., and Fuller, W.A. (1993). Regression weighting for multi-phase samples. *Sankhya*, 55, pp. 297-309.

Cochran, W.G. (1977). *Sampling techniques*, 3rd edition. New York: Wiley.

Fuller, W.A. (1998). Replication variance estimation for two-phase samples. *Statistica Sinica*, 8, pp. 117-132.

Hidiroglou, M.A., and Särndal, C.E. (1998). Use of auxiliary information for two-phase sampling. *Survey Methodology*, 24, pp. 11-20.

Kim, J.K. (2000). *Variance estimation after imputation*. Ph.D. dissertation, Iowa State University.

Kim, J.K., Navarro, A., and Fuller, W. (2000). Variance Estimation for 2000 Census Coverage Estimates. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pp. 515-520.

Kott, P.S., and Stukel, D.M. (1997). Can the Jackknife Be Used with a To-Phase Sample? *Survey Methodology*, 23, pp. 81-89.

Lee, H., and Kim, J.K. (2002). Jackknife Variance Estimation for Two-Phase Samples with High Sampling Fractions. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pp.2024-2028.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association*, 33, pp. 101-116.

Rao, J.N.K. (1973). On double sampling for stratification and analytic surveys. *Biometrika*, 60, pp. 125-133.

Rao, J.N.K., and Shao, J. (1992). Jackknife variance estimation with survey data under hot desk imputation. *Biometrika*, 79, pp. 811-822.

Rao, J.N.K., and Sitter, R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, 82, pp. 453-460.

Rao, J.N.K., and Sitter, R. (1997). Variance estimation under stratified two-phase sampling with applications to measurement bias. *Survey Measurement and Process Quality*, pp. 753-768.

Westat (2000). *WesVar 4.0: User's Guide*. Rockville: Westat, Inc.