Weighting Methods in School-based Surveys

Ronaldo Iachan[1], Pedro Saavedra[2], James. G. Ross[3]
[1] Macro International, Calverton, MD
[2] Macro International, Calverton, MD
[3] Macro International, Calverton, MD

Abstract
Data from school survey samples are typically weighted to account for unequal probabilities of selection, and non-response at the student and school levels. Weight adjustments are usually performed in two stages, the first one based on weighting class adjustments and the other based on post-stratification. Weighting class adjustments are based on weighting cells defined by school or classroom using student enrollment data. Post-stratification adjustments capitalize on external data available for cells (pos-strata) defined by grade, race/ethnicity and gender. This paper investigates alternative approaches to weight adjustments using data from state surveys and national surveys and their implications for design effect and bias adjustment. Alternatives vary in the definition of weighting cells and post-stratum cells with special attention to the potential interaction between class/school and demographic characteristics (e.g., race/ethnicity and gender) used in post-stratification. We compare these approaches using both empirical and simulated results.

Key Words: Weighting, Post-stratification, School surveys

1. Introduction

Data from school survey samples are typically weighted to account for unequal probabilities of selection, and non-response at the student and school levels. This paper looks at student surveys, rather than other school surveys that may involve data collected for teachers, staff, principals, etc. Student surveys require another level of sample selection, and data collection, that lead to interesting angles and complexities in the weighting process.

Weight adjustments are usually performed in two stages, the first one based on weighting class adjustments and the other based on post-stratification. Weighting class adjustments are based on weighting cells defined by school or classroom using student enrollment data. Post-stratification adjustments capitalize on external data available for cells (post-strata) defined by grade, race/ethnicity and gender.

This paper investigates alternative approaches to weight adjustments using data from state surveys and national surveys and their implications for design effect and bias adjustment. Alternatives vary in the definition of weighting cells and post-stratum cells with special attention to the potential interaction between class/school and demographic characteristics (e.g., race/ethnicity and gender) used in post-stratification. We compare these approaches using both empirical and simulated results.

We also discuss a couple of issues that transcend school survey data but are motivated by the context. One issue is the assignment of a single race to multiple race respondents, coupled with the race/ethnicity question. The other issue is the sequencing of trimming and post-stratification steps in the weighting process.

2. Overview of sampling weights

This section describes the basic weighting procedures used in state and national surveys based on stratified multistage samples. These sample designs select areas and schools with probabilities proportional to size (PPS) at the first and second stages, respectively; students are selected with equal probabilities at the third stage of sampling.

At the first stage, primary sampling units (PSUs) are selected with PPS where the measure of size (MOS) is the aggregate of the school MOS values over all eligible schools in the PSU. By selecting an approximately constant number of students in each school, the design is nearly self-weighting in that sampling weights are approximately equal for students before adjustments for non-response at all levels.

Most surveys designed by Macro adopt a weighted measure of size designed to oversample minority groups—typically blacks and Hispanics. The school MOS incorporates coefficients for black and Hispanic student enrollments within each school. The following steps encapsulate the typical approach used for computing sampling weights with the notation that S(i,j,k) is the school size (MOS) and S(i,j) the sum of school sizes within the PSU. The sampling weight is a product of three sampling stage weights.

a) Selection of n(i) PSUs in stratum-i with probabilities proportional to S(i)

$$W_{ij}^{P} = \frac{S_i}{n_i S_{ij}}$$

b) Selection of 2 schools in PSU(j) in stratum-i with probabilities proportional to S(i,j)

$$W_{ijk} = \frac{S_{ij}}{2S_{ijk}}$$

c) Selection and response factors for r(i,j,k) respondents out of a school enrollment E(i,j,k)

$$W_{ijkl}^{R} = \frac{E_{ijk}}{r_{ijk}}$$

Sampling weights can be large for small schools that are sampled with lower probabilities of selection. We reduce these effects by combining small schools or by capping weights.

3. Non-response adjustments

The adjustment factor for school non-response is typically based on the sums of weighted size measure (MOS) over all schools (k) in PSU-j in stratum-i. In the numerator, the sum is over all selections; in the denominator, the sum is over participating schools only.

$$F_i = \left(\frac{\sum_k W_{ijk}^{S} W_{ij}^{P} S_{ijk}}{\sum_{k \in Stratum\, i\, respondents} W_{ijk}^{S} W_{ij}^{P} S_{ijk}} \right)$$

The student weight is the product of the inverse of the probability of selection, a non-response adjustment and a ratio adjustment to control to known school enrollment totals. These adjustments simplify to the ratio of school enrollment and respondents, computed within each school or within a subgroup (e.g., by grade and/or by gender).

We usually form adjustment cells by grade and gender within each school.[1] Alternative adjustment cells are based on classes rather than schools. These adjustments are made possible by enrollment form data made available at the school level or class level. These data list the number of students by grade and gender for each school (or for each class within the school).

Adjustments by grade and gender can lead to very large weights in schools (or classes) where:

- Student reports a grade different from almost every other student in the school
- Student has a gender different from the dominant gender in the school

The first situation often follows from incorrectly grades reported by students, an error that can be corrected with the use of age data also available in the survey. The second situation arises in single-gender schools that may include an errant student of the other gender, either by mistake in reporting or as an actual exception.

4. Post-stratification

For national surveys, we obtain national estimates of racial/ethnic percentages from at least two sources. Private school enrollments by grade and five racial/ethnic groups are obtained from the Private School Universe Survey (PSS). Public school enrollments by grade, gender, and five racial/ethnic categories are obtained from the Common Core of Data (CCD). Both data sources are available from the National Center for Education Statistics (NCES). These databases are combined to produce the enrollments for all schools, and to develop population totals (or percentages) to use as controls in the post-stratification step. In state surveys, the adjustments can also make use of finer updated data available from the State (Dept. of Education).

One difficulty in post-stratification by race/ethnicity is the way ethnicity interacts with race classifications, and the way students report multiple races. In other words, different race/ethnicity cross-classifications and multiple race classifications are used in the survey and in the external population source data.

In addition, multiple race respondents need to be assigned to one single race to make the data consistent with the single-race classification available in the population control distribution. Macro statisticians have developed an algorithm for assignment of a single race that capitalizes on the empirical probabilities of respondents choosing a given single race given multiple race configurations.

[1] To avoid any excessively large weights, we apply a collapsing algorithm for non-response cells. We collapse if the non-response adjustment for the cell exceeds a cap.

5. Examples from simulations and other studies

Using both empirical data (past studies) and/or simulation studies, we compared some of the alternative methods that can be used at each stage. For example, non-response adjustments can be done at the classroom level, at the school level or by adjustment categories than transcend schools. The fact that ethnic groups can be concentrated in the same class can interact with the later post-stratification adjustments. In addition, it can be argued that post-stratification adjustments make non-response adjustments unnecessary.

In one of our simulations studies, we used data from a state survey to examine the relative effectiveness of two alternative non-response adjustment methods using classes or schools as weight adjustment cells. The cells are defined by grade within each major geographic area. We also examined cells defined by gender and by race, but do not present here the data (leading to similar conclusions).

To quantify how close or how far each cell total is from the population total, Tables 1-2 show the type of adjustments that would be used: the ratio of the population total to weight sum. Table 1 shows the racial distribution (Percent White), and Table 2 is focused on the gender distribution (Percent Male),

Table 1: Comparing adjustment methods in terms of the racial distribution

Method Area	By Class	By School	By Area	School Ratio	Population Total
1	71.14%	71.36%	71.60%	70.44%	72.81%
2	51.19%	50.68%	50.00%	50.66%	57.09%
3	57.74%	57.63%	57.39%	56.44%	61.47%
4	79.84%	80.05%	80.14%	79.71%	84.57%
5	76.83%	76.88%	76.88%	76.88%	85.69%
6	42.90%	42.93%	41.61%	42.90%	41.38%
7	54.77%	54.73%	54.36%	57.81%	56.01%
8	67.62%	68.50%	69.37%	70.22%	71.98%

9	62.28%	62.65%	63.02%	63.36%	62.11%

Table 2: Comparing adjustment methods in terms of the gender distribution

Method Area	By Class	By School	By Area	School Ratio	Population Total
1	50.15%	50.38%	50.60%	51.87%	46.32%
2	51.08%	51.27%	51.11%	51.19%	51.68%
3	52.13%	52.33%	52.27%	52.41%	50.78%
4	47.04%	47.75%	47.81%	47.82%	48.20%
5	51.46%	51.15%	51.16%	51.14%	49.59%
6	47.18%	46.94%	45.35%	46.90%	50.49%
7	54.64%	54.75%	54.92%	53.05%	50.39%
8	49.44%	49.44%	48.95%	47.83%	50.22%
9	51.54%	51.35%	51.70%	50.87%	51.40%

It could be expected that the finer adjustments by class would lead to distributions that better reflect the population distribution. These tables suggest, however, that none of the methods is uniformly superior to the others.

Table 3 shows the variability induced by the adjustments at the class level (Method#1) or at the school level (Method#2). The table shows that the first method leads to much greater design effect components due to unequal weighting.

Table 3: Design effects for two adjustment methods

Area	Method 1: Classes	Method 2: Schools
1	1.0162	1.0047
2	1.0674	1.0195
3	1.0064	1.0003
4	1.0279	1.0015
5	1.0102	1.0000
6	1.1099	1.0000
7	1.0196	1.0059
8	1.0969	1.0132
9	1.0436	1.0033

6. Discussion

Post-stratification weight adjustments based on race/ethnicity sometimes lead to very large adjustment factors. One notable (or notorious) example is for Native Americans, since many more students self-report as Native Americans in the survey as would be classified by external data used in post-stratification. In other words, weighted sums exceed, and sometimes far exceed, population control totals for these groups.

On the other hand, because minorities are concentrated in certain schools in some districts, it is possible that one particular ethnic group may end up with very large post-stratified weights.

Another difficulty is the way ethnicity interacts with race classifications, and the way students report multiple races. In other words, different race/ethnicity cross-classifications and multiple race classifications are used in the survey and in the external population source data. In addition, multiple race respondents need to be assigned to one single race to make the data consistent with the single-race classification available in the population control distribution.

Macro statisticians have developed and used different algorithms for assigning a single race that are based either on a) an established hierarchy of single-race priorities, or b) an empirical distribution of single-race choices made by respondents. The latter distribution is supported by data from a survey where a forced single choice question followed the multiple choice question. Using those data, we estimated the empirical probabilities for each single choice component of a multiple-race configuration.

Missing data on race and/or ethnicity is another growing complication that stems from the new OMB race/ethnicity guidelines. When a separate Hispanic ethnicity question is provided prior to the race question, a large proportion of respondents simply skip the race question. For example, in a recent national survey 48% of those reporting Hispanic ethnicity failed to complete the race question. This same survey shows that more than 1/3 of the multiple race cases reported were for the White and American Indian and Alaskan Native (AIAN) combination.

The divergence in race/ethnicity classifications is not only between the self-reported survey data and the population control data but also introduced by the use of another data source as sampling frame and in the design. These data (for example, QED data) are used to assign higher probabilities of selection to schools, and areas, with greater concentrations of minority students. To the extent that the race/ethnicity classification is different in this source, the eventual sampling weights will be more unequal.