Weighting Methods in Survey Sampling

Chiao-chih Chang* Ferry Butar Butar†

Abstract
It is said that a well-designed survey can best prevent nonresponse. However, no matter how well a survey is designed, in practice, nonresponse almost always occurs. The easiest way to deal with nonresponse is to ignore it, but frequently, ignoring nonresponse results in poor survey quality. Item nonresponse and unit nonresponse are two types of nonresponse. Imputation procedures are popular remedies for the former while weighting methods are commonly used to compensate for the latter. This paper focuses on weighting methods that help reduce nonresponse bias.

Key Words: nonresponse, adjustment, calibration, missing data

1. Introduction

In survey sampling, a good sample is desirable for making good inferences about a population. The bias of a properly calculated estimate results from sampling error and non-sampling errors. Practically, sampling error always exists because of sample-to-sample variation. The only sure way to avoid sampling error is to study the entire population, $i.e.$, census, and that is impractical for a huge population. Selection bias is an example of non-sampling error, and it includes nonresponse which might greatly bias estimates calculated without adjustments.

If there is nonresponse together with under-coverage in survey sampling, missing data will result. Under-coverage occurs when not all of the elements in the target population are included in the sampling frame. Nonresponse can be divided into item nonresponse and unit nonresponse. In item nonresponse, at least one but not all of the measurements of interest are obtained from the sampled unit. In unit nonresponse, the sampled unit provides nothing. However, some information usually is available by other means.

Example 1.1. *Suppose a telephone survey is conducted to estimate household electricity consumption in a certain region, and a sample of n individuals is drawn from the residential phone directory. Therefore, the target population includes all households in the region and the sampling frame is the list of people in the telephone directory. Each of the n individuals is asked the type of housing unit, the number of bedrooms, and the average monthly electrical use (see Table 1).*

Lohr (2009) stated that missing data and haphazard mistakes in data collection are often the biggest causes of error in a survey. Designing a good survey and questionnaire, and being careful in collecting data can prevent poor response rates and mistakes in data collection. Follow-up and callback are common procedures used to increase response rate. In the telephone survey example, it is possible that the selected person unavailable at the first call is contacted and there is a response at the kth callback. k is a positive integer. Since making phone calls may be expensive and time-consuming, it is impractical to try to contact someone a number of times. Furthermore, it is usually the case that some people refuse to be interviewed no matter how many times an interviewer tries.

*Department of Mathematics and Statistics Sam Houston State University, Huntsville, TX 77341
†Department of Mathematics and Statistics Sam Houston State University, Huntsville, TX 77341

Table 1: Hypothetical data for example

Individual	Housing Unit Type	Number of Bedrooms	Average Monthly Electric Usage (kWh)
a	House	3	2,000
b	House	-	-
c	-	-	3,000
d	Apartment	-	500
f	-	-	-
g	-	2	1,000
h	Other	2	-
i	-	3	-
⋮	⋮	⋮	⋮

Unit nonresponse: f.
Item nonresponse: b, c, d, g, h, i.

Since nonresponse almost always occurs in a survey, one should be cautioned if the nonrespondents differ critically from the respondents, especially when the response rate is low. Due to the presence of nonresponse and the plausible difference between respondents and nonrespondents, the sample is considered to be representative for the population of people who will answer survey questions but not for the target population. Therefore, inferences based only on the respondents do not seem to be valid. When nonresponse is inevitable and non-ignorable, there are ways to compensate.

2. Simple Random Sampling (SRS)

Simple random sampling is the most common probability sampling procedure. Although it is referred to as simple random sampling, this procedure is very important because fully understanding simple random sampling is a prerequisite for further studies in other sampling techniques. In simple random sampling, every element in the population has an equal probability of being selected in the sample. Intuitively, each subset of a fixed number of elements in the population has an equal probability of being selected as a sample.

2.1 Estimation in Simple Random Sampling

In many sampling studies, the most common objective is to estimate the population total, mean, or proportion. The estimation of population total is emphasized here because estimated population total divided by the population size yields the estimated population mean and estimating proportion is a special form of estimating mean. To derive the estimator of the population total, let $\mathcal{U} = \{1, 2, ..., N\}$ and let $\mathcal{S}$ be a set of n elements chosen from $\mathcal{U}$. Let $\mathcal{U}_y = \{y_i \mid i \in \mathcal{U}\}$ denote the population of size N, then y_i is the survey characteristic of the ith unit and the simple random sample is denoted by $\mathcal{S}_y = \{y_i \mid i \in \mathcal{S}\}$. Let t and t_s be population total and sample total, respectively, then $t = \sum_{i=1}^{N} y_i$ and $t_s = \sum_{i \in \mathcal{S}} y_i$. The

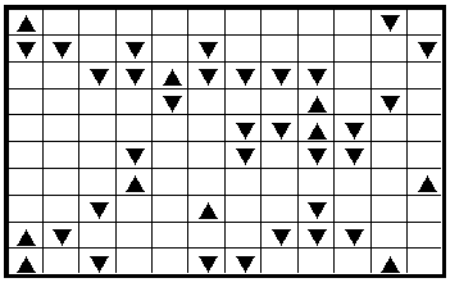

FIGURE 2.1 An SRS of size 40 drawn from a population of size 120

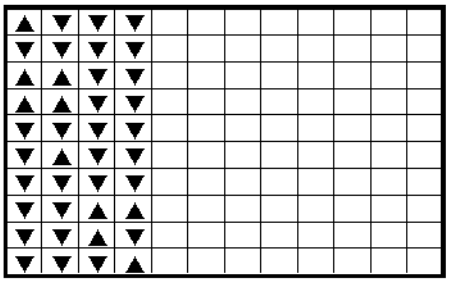

FIGURE 2.2 A rearrangement of FIGURE 2.1

unbiased estimator of the population mean,

$$\overline{Y} = \frac{t}{N}, \tag{1}$$

is simply the sample mean, $\overline{y} = \dfrac{t_s}{n}$. The population total is estimated by substituting $\overline{y}$ for $\overline{Y}$ in (1), so $\hat{t} = \frac{N}{n} t_s$. Note that t and $\overline{Y}$ are unknown since, in a survey, a sample is a small portion of the population, *i.e.*, $n < N$. In fact, n is much smaller than N, and this is the idea of sampling using n units to represent N units, i.e, one unit in the sample represents N/n units in the population. Suppose that the perfect representative sample is given, then a logical relation between the sample total and the population total should be $t_s : t = n : N$. From this relation, one can see that the estimation of the population total is

$$\hat{t} = \frac{N}{n} t_s. \tag{2}$$

Let $w_i = N/n$ for all $i \in \mathcal{S}$, then (2) may be rewritten as

$$\hat{t} = \sum_{i \in \mathcal{S}} w_i y_i, \tag{3}$$

where w_i is called the sampling weight. Sampling weights are calculated to help simplify the calculation in many sample surveys. More importantly, making adjustments on sampling weights may help reduce nonresponse bias.

Example 2.1. *A simple random sample of size 40 drawn from a population of size 120 is illustrated in Figure 2.1. In this sample design,* $w_i = \frac{120}{40}$ *for all* $i \in \mathcal{S}$*, i.e., one selected unit represents 3 units including itself in the population.*

Cornfield (1944) introduced a useful method of deriving the expected value and the variance of an estimator in sampling without replacement from a finite population. In order to use the method the researcher must show that $\hat{t}$ is an unbiased estimator of t and to obtain the variance of $\hat{t}$. Let Z_i be a random variable such that

$$Z_i = \begin{cases} 1, & i \in \mathcal{S} \\ 0, & i \in \mathcal{S}' \end{cases}.$$

Note that $E(Z_i) = n/N$, and $E(Z_i Z_j) = \frac{n}{N} \frac{n-1}{N-1}$. By definition, given n and N, the variance of $\hat{t}$ is $Var\left(\hat{t} \mid n, N\right) = E\left[\left(\hat{t} - t\right)^2 \mid n, N\right]$, and it can be shown $Var\left(\hat{t} \mid n, N\right) = N^2 \left(1 - \frac{n}{N}\right) \frac{S^2}{n}$, where $S^2 = (N-1)^{-1} \sum_{i=1}^{N} (y_i - \overline{Y})^2$, where the unknown S^2 can be

estimated by the sample variance, $s^2 = (n-1)^{-1} \sum_{i \in \mathcal{S}} (y_i - \overline{y})^2$. Thus,

$$\widehat{Var\left(\hat{t} \mid n, N\right)} = N^2 \left(1 - \frac{n}{N}\right) \frac{s^2}{n} \tag{4}$$

is an unbiased estimator of $Var\left(\hat{t} \mid n, N\right)$.

2.2 Nonresponse in Simple Random Sampling

Two widely discussed remedies for nonresponse are imputation and weighting adjustment. In imputation, the researcher may fill in missing values by plausible values which are generated from observed variables, hence imputation is the common remedy for item nonresponse. There are many imputation procedures used such as hot-deck imputation, cold-deck imputation, regression imputation, multiple imputation, and fractional imputation. However, our research focuses on weighting adjustment, in which sampled units are classified into classes by auxiliary information, and then each responding sampled unit is assigned an adjustment weight calculated by taking the inverse of the response rate of the corresponding class. Responding sampled units in a class with low response rate are assigned higher weights than those in a class with a high response rate. Weighting adjustment is the common remedy for unit nonresponse, and it does not require filling in gaps in the data. A weighted estimator is unbiased under the assumption the probability of nonresponse is the same for all units within a class. Although the assumption is usually not satisfied in practice, it is more reasonable than to assume that the probability of nonresponse is the same for all units in the sample, hence nonresponse is ignorable. As a result, weighting adjustment might not eliminate nonresponse bias but it is useful for reducing nonresponse bias.

Example 2.2. *The small data set in Table 2. is created to mimic the situation in Example 1.1 in which one can see both item nonresponse and unit nonresponse. This sample of size, $n = 40$, is drawn from the population of size, $N = 120$. The goal is to estimate t, the total average monthly electricity consumption in the region of interest. Y is the study variable; X_1, X_2, and X_3 are the auxiliary variables. X_3 is obtained from other available sources but not from the sampled units, hence it is always observed. Y, X_1, and X_2 are subject to nonresponse. Suppose that nonresponse can be totally ignored and only observed Ys are considered in the estimation, then $\hat{t}$ and $\widehat{Var\left(\hat{t} \mid n, N\right)}$ can be calculated by using (3) and (4), respectively. Although $\hat{t}$ may not be affected by the nonresponse, $\widehat{Var\left(\hat{t} \mid n, N\right)}$ becomes large because n is 26 instead of 40. Consequently, the confidence interval may be too wide and does not provide much information.*

Ignoring nonresponse is usually not a good idea. The following are some possible approaches to handling this data set.

- *Fill in every gap in the data set by some imputation techniques.*
- *Form classes by X_3 and perform weighting adjustment.*
- *Fill in the gaps in auxiliary information and form classes by the best possible set of auxiliary variables. Then, perform weighting adjustment.*
- *Fill in the gaps for the data points with sufficient auxiliary information. Weighting adjustment can then compensate for the remaining nonresponse. Note that item nonresponse may be treated as unit nonresponse when a respondent answers too few questions.*

Table 2: Hypothetical sample of 40 from a population of 120

ID	X_1	X_2	X_3	Y
1	3	-	1	2500
2	2	1	1	500
9	3	3	1	-
10	1	-	1	-
12	1	4	2	4000
19	3	1	1	1500
23	3	2	2	-
28	4	2	1	2500
30	3	-	1	1000
32	2	1	1	2000
33	2	2	1	3000
36	3	2	1	3000
37	-	-	1	-
43	-	-	2	-
44	1	3	1	2000
52	3	1	1	1000
53	1	-	1	4500
58	4	-	1	500
60	3	2	1	2000
63	3	2	1	1500

ID	X_1	X_2	X_3	Y
65	1	2	1	1500
66	3	2	1	2500
70	3	1	1	500
73	3	2	1	-
75	1	-	2	2000
79	4	-	1	-
83	1	3	1	3500
84	1	3	2	-
85	3	1	1	1000
86	2	3	1	3000
88	-	-	2	-
89	2	2	2	-
95	1	-	2	-
96	1	2	1	2500
99	-	-	2	-
101	3	1	1	2000
104	3	1	1	1500
110	2	-	1	2500
112	1	-	1	-
117	2	3	2	-

Note: X_1=1 house, =2 mobile home, =3 apartment,=4 other; X_2=1 one bedroom, =2 two bedrooms,= 3 three bedrooms,=4 at least four bedrooms; X_3=1 the community was establishd after 1950, =2 otherwise; Y denotes average monthly electrical usage in kWh.

Lohr (2009) noted that response rates can be manipulated by defining them differently and presented several formulas that are used in surveys for calculating response rates. Here, for demonstration purposes, assume that only one variable is to be observed in a survey. Therefore, the calculation of response rates is straightforward. Let Y be the study variable subject to nonresponse, and the response rates are calculated by dividing the number of observed Ys by the number of units in sample. Let m be the number of observations obtained from the simple random sample of size n drawn from the population of size N and let $r \cdot 100\%$ denote the response rate, where $r = \frac{m}{n}$. Figure 2.1 displays the sample with 100%, a response rate that is nearly impossible to achieve in surveys. The following example is an altered Example 2.1 in which nonresponse occurs. It shows the effect of ignoring nonresponse.

Example 2.3. *Consider the situation in Example 1.1 again. The goal is to estimate the total average monthly electricity consumption in the region of interest. Let Y denote the average monthly electrical usage. Suppose that the sample is shown in Figure 2.3 in which $\triangledown$ and $\triangle$ represent selected persons who do not give information about their average monthly electrical usage while $\blacktriangledown$ and $\blacktriangle$ represent selected persons who answer the question about their average monthly electrical usage. The response rate is 65% because $m = 26$ and $n = 40$.*

Since the sample is drawn from a telephone directory, selected persons' addresses are usually available notwithstanding nonresponse. This additional information enables the researcher to find out in what community a person lives. Suppose it is observed that housing units in old communities are generally larger than those in new communities. Also, it is reasonable to state that people who live in large properties normally have higher electrical

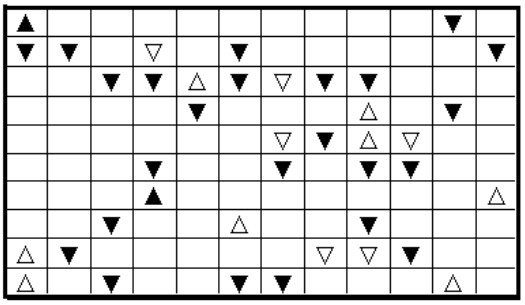

FIGURE 2.3 An SRS of size 40 drawn from a population of size 120

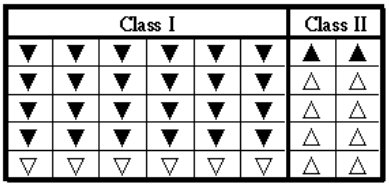

FIGURE 2.4 The SRS in FIGURE 2.3 with classification and rearrangement

usage than those who live in small properties. Ideally, one would like to form classes by some characteristics and then view the group of respondents in each class as a simple random sample of the corresponding class. Consider only Y *and* X_3*, then the data set in Table 2. can be classified into two groups by the age of the community,* X_3*. One can visually tell the difference in response rate between the two classes in Figure 2.4. Class* I *is comprised of* $n_1 = 30$ *people and* $m_1 = 24$ *out of* 30 *people are respondents. Thus, the response rate in class* I *is* 80%*, i.e.,* $r_1 = 0.8$*. Class* II *is comprised of* $n_2 = 10$ *people and only* $m_2 = 2$ *out of* 10 *people are respondents. Thus, the response rate in class* II *is* 20%*, i.e.,* $r_2 = 0.2$*.*

Suppose that people in class II *consume more electricity than people in class* I*, then ignoring nonresponse can seriously bias the estimate. To show that, use the data of the* 26 *respondents to estimate the total average monthly electricity consumption,* t*. This can be thought of as taking a simple random sample of size,* $m = 26$*, drawn from the population of size,* $N = 120$*, so here,* $\mathcal{S}$ *is a set containing* 26 *respondents' IDs. The sampling weight is* N/m*. Use (3) to find* $\hat{t} \approx 249,231$*.*

As an example, assume that the sample is representative and the classification is perfect, then the group of respondents in each class can be viewed as a simple random sample obtained from a corresponding subpopulation. Let N_1 *and* N_2 *be the number of members in subpopulation* I *and the number of members in subpopulation* II*, respectively. Since the sample is representative, then* $\frac{n_c}{N_c} = \frac{n}{N}$ *holds for* $c = 1, 2$*. Therefore,* $N_1 = 90$ *and* $N_2 = 30$*. Let* $\overline{y}_1$ *and* $\overline{y}_2$ *be the mean average monthly electrical usage in class* I *and* II*, respectively. So,* $\overline{y}_1 = 2,000$ *and* $\overline{y}_2 = 3,000$*. Let* $\overline{Y}_1$ *and* $\overline{Y}_2$ *be the mean average monthly electrical usage in subpopulation* I *and* II*, respectively, then* $t = N_1\overline{Y}_1 + N_2\overline{Y}_2$*. The reasonable estimate of* t *should be:* $\hat{t}_{adj} = N_1\overline{y}_1 + N_2\overline{y}_2 = 270,000$*.*

If $\overline{y}_c$ *is close to* $\overline{Y}_c$ *for* $c = 1, 2$*, then the difference between* $\hat{t}_{adj}$ *and* t *will be small. Assume that* $\overline{y}_1 = \overline{Y}_1$ *and* $\overline{y}_2 = \overline{Y}_2$*, then* $\hat{t}_{adj} = t$*. Compare* $\hat{t}$ *with* $\hat{t}_{adj}$*, one may conclude that* $\hat{t}$ *underestimates* t*, and* $\left|\hat{t} - t\right|$ *is large if:*

- *for a fixed* $r_2 - r_1 \neq 0$*,* $\left|\overline{Y}_2 - \overline{Y}_1\right|$ *is large.*
- *for a fixed* $\overline{Y}_2 - \overline{Y}_1 \neq 0$*,* $|r_2 - r_1|$ *is large.*

If $r_2 = r_1$ *or* $\overline{Y}_2 = \overline{Y}_1$*, then* $\hat{t} = t$*.*

2.3 Weighting Adjustment in Simple Random Sampling

In general, suppose that the classification forms C classes. The set, $\mathcal{S}$, containing the identification numbers of sampled units can be partitioned so that $\mathcal{S} = \bigcup_{c=1}^{C} \mathcal{S}_c$. Let $\mathcal{U} =$

Table 3: The composition of sample response and nonresponse

c	1	2	...	C	Note
m_c	m_1	m_2	$\cdots$	m_C	$m = \sum_{c=1}^{C} m_c$
$n_c - m_c$	$n_1 - m_1$	$n_2 - m_2$	$\cdots$	$n_C - m_C$	$n - m$
n_c	n_1	n_2	$\cdots$	n_C	$n = \sum_{c=1}^{C} n_c$
N_c	N_1	N_2	$\cdots$	N_C	$N = \sum_{c=1}^{C} N_c$
r_c	$\frac{m_1}{n_1}$	$\frac{m_2}{n_2}$	$\cdots$	$\frac{m_C}{n_C}$	$\frac{m}{n}$

$\mathcal{S} \cup \mathcal{S}'$. There exist partitions $\mathcal{U}_1, \mathcal{U}_2, \cdots, \mathcal{U}_C$, such that $\mathcal{S}_c \subset \mathcal{U}_c$ for $c = 1, 2, \cdots, C$. Let $\mathcal{A}_c$ be the set that contains the identification numbers of responding units in class c, so $\mathcal{A}_c \subset \mathcal{S}_c$ for $c = 1, 2, \cdots, C$. Let N_c, n_c, and m_c denote the number of elements in $\mathcal{U}_c$, $\mathcal{S}_c$, and $\mathcal{A}_c$, respectively, and assume that $m_c > 1$ for $c = 1, 2, \cdots, C$. So, $m = \sum_{c=1}^{C} m_c$, $n = \sum_{c=1}^{C} n_c$, and $N = \sum_{c=1}^{C} N_c$. The overall response rate and response rate for each class can be calculated and compared by constructing a table similar to Table 3. Suppose that each of the n_c units has the same probability of responding, then m_c responding units may be interpreted as a simple random sample drawn from the n_c sampled units. To interpret in terms of weighting, the m_c units is to represent the n_c units, so each of the m_c units has a weight of $\frac{n_c}{m_c}$. For $c = 1, 2, \cdots, C$, let y_{ci} be the observation of the study having the identification number, i, and let $t_{sc} = \sum_{i \in \mathcal{S}_c} y_{ci}$. Let $t = \sum_{c=1}^{C} t_c$ and $t_c = \sum_{i \in \mathcal{U}_c} y_{ci}$, then t_c can be estimated by $\frac{N_c}{n_c} \cdot t_{sc}$. Since not every y_{ci} where $i \in \mathcal{S}_c$ is available, t_{sc} should be estimated. Let $t_{ac} = \sum_{i \in \mathcal{A}_c} y_{ci}$, then $\hat{t}_{sc} = \frac{n_c}{m_c} \cdot t_{ac}$. If N_c is known for $c = 1, 2, \cdots, C$, then $\hat{t}_c = \frac{N_c}{n_c} \hat{t}_{sc} = \frac{N_c}{n_c} \frac{n_c}{m_c} \cdot t_{ac} = \frac{N_c}{m_c} \sum_{i \in \mathcal{A}_c} y_{ci}$.

It can be regarded as using a simple random sample of size m_c to represent the population of size N_c. Therefore, the adjusted estimator of t is

$$\hat{t}_{adj} = \sum_{c=1}^{C} \sum_{i \in \mathcal{A}_c} w_{ci} y_{ci} \tag{5}$$

where $w_{ci} = N_c/m_c$, and it can be expressed as $w_{ci} = w_{b_{ci}} w_{adj_{ci}}$ where $w_{b_{ci}} = N_c/n_c$ is called the base weight in class c and $w_{adj_{ci}} = n_c/m_c$ is called the nonresponse adjustment weight in class c. In (5), N_c should be known for all $c = 1, 2, \cdots, C$, and $\hat{t}_{adj}$ is also called the post-stratified estimator of t that may also be employed in adjusting under-coverage. Generally, $\hat{t}_{adj}$ should provide a better estimate than $\hat{t} = \sum_{i \in \mathcal{A}} \frac{N}{m} y_i$ where nonresponse is totally ignored.

To show that $\hat{t}$ is an unbiased estimator of t, follow the method proposed by Cornfield

(1944) and define:

$$Z_{ci} = \begin{cases} 1, & i \in \mathcal{S}_c \\ 0, & i \in \mathcal{S}_c' \end{cases} \quad \text{and} \quad A_{ci} = \begin{cases} 1, & i \in \mathcal{A}_c \\ 0, & i \in \mathcal{A}_c' \end{cases}.$$

Thus, $E(Z_{ci}) = n_c/N_c$, $E(A_{ci}) = m_c/n_c$, and $E(Z_{ci}A_{ci}) = m_c/N_c$. For $c = 1, 2, \cdots, C$, let $\mathbf{u}_c = (m_c, n_c, N_c)'$, then it can be shown that $E\left(\hat{t}_c \mid \mathbf{u}_c\right) = t_c$. Since the m_c responding units can be regarded as a simple random sample of size m_c representing the population of size N_c for $c = 1, 2, \cdots, C$, then

$$Var\left(\hat{t}_c \mid \mathbf{u}_c\right) = N_c^2 \left(1 - \frac{m_c}{N_c}\right) \frac{S_c^2}{m_c} \tag{6}$$

where $S_c^2 = (N_c - 1)^{-1} \sum\limits_{i \in \mathcal{U}_c} (y_{ci} - t_c/N_c)^2$. Note that (6) is just the analogous to the variance of t seen in Section 2.1. Therefore, $E\left(\hat{t}_{adj} \mid \mathbf{u}\right) = E\left(\sum\limits_{c=1}^{C} \sum_{i \in \mathcal{A}_c} \frac{N_c}{m_c} y_{ci} \mid \mathbf{u}\right) = \sum\limits_{c=1}^{C} E\left(\hat{t}_c \mid \mathbf{u}_c\right) = \sum\limits_{c=1}^{C} t_c = t$, and $Var\left(\hat{t}_{adj} \mid \mathbf{u}\right) = Var\left(\sum\limits_{c=1}^{C} \hat{t}_c \mid \mathbf{u}\right) = \sum\limits_{c=1}^{C} Var\left(\hat{t}_c \mid \mathbf{u}_c\right) = \sum\limits_{c=1}^{C} N_c^2 \left(1 - \frac{m_c}{N_c}\right) \frac{S_c^2}{m_c}$, where $\mathbf{u} = (m_1, \cdots, m_C, n_1, \cdots, n_C, N_1, \cdots, N_C)'$. In the post-stratified estimator, the estimated variance (5) is

$$\widehat{Var\left(\hat{t}_{adj} \mid \mathbf{u}\right)} = \sum_{c=1}^{C} N_c^2 \left(1 - \frac{m_c}{N_c}\right) \frac{s_c^2}{m_c} \tag{7}$$

where $s_c^2 = (m_c - 1)^{-1} \sum\limits_{i \in \mathcal{A}_c} (y_{ci} - t_{ac}/m_c)^2$; $E\left(s_c^2\right) = S_c^2$.

So far, the estimation can be done when N_c is available for $c = 1, 2, \cdots, C$. However, if this is not the case, consider the example of the perfect simple random sample in Example 2.3 in which the ratio of each sub sample size to its corresponding subpopulation size is equal to the ratio of the sample size to the population size. Although this condition is almost impossible to satisfy in a single sampling activity, the sampling distribution of n_c/N_c should be centered at n/N. That is $E(n_c/N_c) = n/N$ and $E(N_c/n_c) = N/n$. Therefore, t can be estimated by

$$\tilde{t}_{adj} = \sum_{c=1}^{C} \sum_{i \in \mathcal{A}_c} w_{ci}^* y_{ci} \tag{8}$$

where $w_{ci}^* = \frac{N}{n} \frac{n_c}{m_c}$ and $\tilde{t}_{adj}$ is called the weighting class estimator of t.

Let $\mathbf{u}^* = (m_1, \cdots, m_C, n_1, \cdots, n_C)'$; therefore, it can be shown that the expected value of $\tilde{t}_{adj}$ given $\mathbf{u}^*$, that is $E\left(\tilde{t}_{adj} \mid \mathbf{u}^*\right) = t + \sum\limits_{c=1}^{C} \frac{n_c}{n} \left(\frac{N}{N_c} t_c - t\right) \neq t$. The variance of $\tilde{t}_{adj}$ given $\mathbf{u}^*$ can be estimated by substituting $\frac{n_c}{n} \cdot N$ for N_c in (7). Thus,

$$\widetilde{Var\left(\tilde{t}_{adj} \mid \mathbf{u}^*\right)} = \sum_{c=1}^{C} \left(N \cdot \frac{n_c}{n}\right)^2 \left(1 - \frac{n}{N} \frac{m_c}{n_c}\right) \frac{s_c^2}{m_c}. \tag{9}$$

Consider the weighting class estimator. Oh and Scheuren (1983) proposed that an approximate $100(1 - \alpha)\%$ confidence interval for t may be constructed by

$$\tilde{t}_{adj} \pm z_{\alpha/2} \sqrt{\widetilde{MSE\left(\tilde{t}_{adj} \mid \mathbf{u}^*\right)}} \tag{10}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$th percentile of the standard normal distribution and

$$\widetilde{MSE\left(\tilde{t}_{adj} \mid \mathbf{u}^*\right)} = \widetilde{Var\left(\tilde{t}_{adj} \mid \mathbf{u}^*\right)} + \left[\widetilde{Bias\left(\tilde{t}_{adj} \mid \mathbf{u}^*\right)}\right]^2$$
$$= \sum_{c=1}^{C}\left(N\cdot\frac{n_c}{n}\right)^2\left(1-\frac{n}{N}\frac{m_c}{n_c}\right)\frac{s_c^2}{m_c} + \frac{N-n}{N-1}\sum_{c=1}^{C}\frac{n_c}{n^2}\left(N\cdot\frac{t_{ac}}{m_c}-\tilde{t}_{adj}\right)^2.$$

3. Stratified Random Sampling (STR)

In simple random sampling, it is possible that some certain groups of sampling units in the population are underrepresented and hence, will result a bad sample. Stratified random sampling enables one to avoid having a bad sample and to improve the precision of estimates for a fixed sample size or to achieve the same level of precision with a sample size not as large as in simple random sampling. In stratified random sampling, the population of size N is divided into $H \geq 2$ subpopulations called strata. Let stratum h have size N_h for $h = 1, 2, \cdots, H$. The sample is obtained by randomly selecting n_h sampling units from stratum h for $h = 1, 2, \cdots, H$. Therefore, H simple random samples are obtained in the stratified random sampling. Stratified random sampling works well when the variation in the measurement of interest is low within the strata but is high between the strata.

Suppose that the total sample size is n so $n = \sum_{h=1}^{H} n_h$. To allocate the sample among the strata may depend on the sample design or some constraints such as costs. Discussions of the allocation of sampling units among the strata can be found in Cochran (1977) and Lohr (2009).

3.1 Estimation in Stratified Random Sampling

Since a procedure of simple random sampling is implemented in each stratum, the results of the preceding sections can be applied directly to each stratum. Let index set, $\mathcal{U} = \{1, 2, \cdots, N\}$, denote the population of size N. Let $\mathcal{U}_h$ denote stratum h of size N_h for $h = 1, 2, \cdots, H$ such that $\mathcal{U} = \bigcup_{h=1}^{H} \mathcal{U}_h$, $N = \sum_{h=1}^{H} N_h$, and $\mathcal{U}_h \cap \mathcal{U}_k = \emptyset$ for $h \neq k$. Let $\mathcal{S}_h$ denote the simple random sample of size n_h drawn form $\mathcal{U}_h$ for $h = 1, 2, \cdots, H$ and let $\mathcal{S} = \bigcup_{h=1}^{H} \mathcal{S}_h$. Therefore the total sample size is $n = \sum_{h=1}^{H} n_h$. Let y_{hi} be the measurement of interest in stratum h, then $t_h = \sum_{i\in\mathcal{U}_h} y_{hi}$ and $t_{str} = \sum_{h=1}^{H} t_h$ are the total in stratum h and the population total, respectively. The unbiased estimator of t_h is

$$\hat{t}_h = \sum_{i\in\mathcal{S}_h} w_{str_{hi}} y_{hi} \tag{11}$$

where $w_{str_{hi}} = N_h/n_h$, and the variance of $\hat{t}_h$ given n_h and N_h is $Var\left(\hat{t}_h \mid n_h, N_h\right) = N_h^2(1-n_h/N_h)S_h^2/n_h$ where $S_h^2 = (N_h-1)^{-1}\sum_{i\in\mathcal{U}_h}\left(y_{hi}-\overline{Y}_h\right)^2$ and $\overline{Y}_h = t_h/N_h$. The sample variance obtained from stratum h is $s_h^2 = (n_h-1)^{-1}\sum_{i\in\mathcal{S}_h}\left(y_{hi}-\overline{y}_h\right)^2$ where $\overline{y}_h = n_h^{-1}\sum_{i\in\mathcal{S}_h} y_{hi}$ and $E\left(s_h^2\right) = S_h^2$. Therefore,

$$\widehat{Var\left(\hat{t}_h \mid n_h, N_h\right)} = N_h^2\left(1-\frac{n_h}{N_h}\right)\frac{s_h^2}{n_h}. \tag{12}$$

Since the population total is $t_{str} = \sum_{h=1}^{H} t_h$, then it can be estimated by $\hat{t}_{str} = \sum_{h=1}^{H} \hat{t}_h$. Thus, $\hat{t}_{str} = \sum_{h=1}^{H} \sum_{i \in \mathcal{S}_h} w_{str_{hi}} y_{hi}$, and $E\left(\hat{t}_{str} \mid \mathbf{q}\right) = \sum_{h=1}^{H} E\left(\hat{t}_h \mid n_h, N_h\right) = \sum_{h=1}^{H} t_h = t_{str}$, where $\mathbf{q} = (n_1, \cdots, n_H, N_1, \cdots, N_H)$. The variance of $\hat{t}_{str}$ given $\mathbf{q}$ can be estimated by

$$\widehat{Var\left(\hat{t}_{str} \mid \mathbf{q}\right)} = \sum_{h=1}^{H} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h} \tag{13}$$

because $Var\left(\hat{t}_{str} \mid \mathbf{q}\right) = Var\left(\sum_{h=1}^{H} \hat{t}_h \mid \mathbf{q}\right) = \sum_{h=1}^{H} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$.

3.2 Nonresponse in Stratified Random Sampling

In stratified random sampling, the researcher deals with multiple simple random samples. Suppose that there are H strata and a simple random sample of size n_h is obtained from each of the H strata having size N_h. Thus, there are H simple random samples. Consider only unit nonresponse and for $h = 1, 2, \cdots, H$, let m_h denote the number of the observations obtained from the n_h sampled units. If $m_h < n_h$ (which is almost always the case in surveys) then nonresponse occurs in the sample obtained from stratum h and estimation without adjustments might not produce the best results for making inferences about the subpopulation as well as the population.

3.3 Weighting Adjustment in Stratified Random Sampling

Suppose that a stratified random sampling procedure is carried out and there exists nonresponse. Assume that it is possible to divide the sample selected from stratum h into C_h classes by some available auxiliary information for $h = 1, 2, \cdots, H$ such that every sampled units in the same class has an equal probability of responding. Let set, $\mathcal{S}_{hc}$, denote class c having size n_{hc} in the sample selected from stratum h for $c = 1, 2, \cdots, C_h$ and $h = 1, 2, \cdots, H$, then $\mathcal{S}_h = \bigcup_{c=1}^{C_h} \mathcal{S}_{hc}$ where $\mathcal{S}_{hc} \cap \mathcal{S}_{hk} = \emptyset$ for $h \neq k$ and $n_h = \sum_{c=1}^{C_h} n_{hc}$. For $h = 1, 2, \cdots, H$, let $\mathcal{U}_{h1}, \mathcal{U}_{h2}, \cdots$, and $\mathcal{U}_{hC_h}$ be the substrata of $\mathcal{U}_h$ and each of them has size $N_{h1}, N_{h2}, \cdots$, and N_{hC_h}, respectively, so that $N_h = \sum_{c=1}^{C_h} N_{hc}$ and $\mathcal{S}_{hc} \subset \mathcal{U}_{hc}$.

Let $\mathcal{A}_{hc}$ denote the group of responding units in class c of the sample drawn from stratum h and let m_{hc} be the size of $\mathcal{A}_{hc}$. For $h = 1, 2, \cdots, H$, r_{hc} denotes the response rate in $\mathcal{S}_{hc}$. In post-stratified approach, the base weight, $w_{b_{hci}} = N_{hc}/n_{hc}$, indicates that one unit in $\mathcal{S}_{hc}$ is to represent N_{hc}/n_{hc} units in substratum hc. The nonresponse adjustment weight, $w_{adj_{hci}} = n_{hc}/m_{hc}$, indicates that one unit in $\mathcal{A}_{hc}$ is to represent n_{hc}/m_{hc} units in $\mathcal{S}_{hc}$. The final weight, $w_{hci} = N_{hc}/m_{hc}$ is

$$w_{hci} = w_{b_{hci}} \times w_{adj_{hci}}. \tag{14}$$

This indicates that one unit in $\mathcal{A}_{hc}$ represents N_{hc}/m_{hc} units in substratum hc.

Using the techniques in the preceding sections, one can find an unbiased estimator, given $\mathbf{u}_{hc} = (m_{hc}, n_{hc}, N_{hc})'$, for t_{hc}, the total of substratum hc:

$$\hat{t}_{hc} = \sum_{i \in \mathcal{A}_{hc}} w_{hci} y_{hci}. \tag{15}$$

The variance of $\hat{t}_{hc}$ given $\mathbf{u}_{hc}$ may be estimated by

$$\widehat{Var\left(\hat{t}_{hc} \mid \mathbf{u}_{hc}\right)} = N_{hc}^2 \left(1 - \frac{m_{hc}}{N_{hc}}\right) \frac{s_{hc}^2}{m_{hc}}$$

where $s_{hc}^2 = \dfrac{1}{m_{hc}-1} \sum\limits_{i \in \mathcal{A}_{hc}} \left(y_{hci} - \dfrac{t_{a_{hc}}}{m_{hc}}\right)^2$ and $t_{a_{hc}} = \sum\limits_{i \in \mathcal{A}_{hc}} y_{hci}$. Moreover, for $h = 1, 2, \cdots, H$, the post-stratified estimator of the total of stratum h is

$$\left(\hat{t}_h\right)_{adj} = \sum_{c=1}^{C_h} \sum_{i \in \mathcal{A}_{hc}} w_{hci} y_{hci} \tag{16}$$

and $E\left[(\hat{t}_h)_{adj} \mid \mathbf{u}_h\right] = t_h$ where $\mathbf{u}_h = (m_{h1}, \cdots, m_{hC_h}, n_{h1}, \cdots, n_{hC_h}, N_{h1}, \cdots, N_{hC_h})'$ since $E\left(\hat{t}_{hc} \mid \mathbf{u}_{hc}\right) = t_{hc}$. Thus, for $h = 1, 2, \cdots, H$, the variance of $\left(\hat{t}_h\right)_{adj}$ given $\mathbf{u}_h$ can be estimated by $\widehat{Var}\left[\left(\hat{t}_h\right)_{adj} \mid \mathbf{u}_h\right]$, by summing $\widehat{Var\left(\hat{t}_{hc} \mid \mathbf{u}_{hc}\right)}$ from $c = 1$ to C. Therefore, using a post-stratified approach, the population total can be estimated by

$$\left(\hat{t}_{str}\right)_{adj} = \sum_{h=1}^{H} \sum_{c=1}^{C_h} \sum_{i \in \mathcal{A}_{hc}} w_{hci} y_{hci} \tag{17}$$

and it is unbiased given $\mathbf{u}_\bullet = (\mathbf{u}_1; \mathbf{u}_2; \cdots; \mathbf{u}_H)$ since $E\left[\left(\hat{t}_h\right)_{adj} \mid \mathbf{u}_h\right] = t_h$. The variance of $\left(\hat{t}_{str}\right)_{adj}$ given $\mathbf{u}_\bullet$ may be estimated by

$$\widehat{Var}\left[\left(\hat{t}_{str}\right)_{adj} \mid \mathbf{u}_\bullet\right] = \sum_{h=1}^{H} \sum_{c=1}^{C_h} N_{hc}^2 \left(1 - \frac{m_{hc}}{N_{hc}}\right) \frac{s_{hc}^2}{m_{hc}}. \tag{18}$$

Suppose that the N_{hc} is not available for $c = 1, 2, \cdots, C_h$ and $h = 1, 2, \cdots, H$, then consider the weighting class estimator: $\tilde{t}_{hc} = \sum\limits_{i \in \mathcal{A}_{hc}} w_{hci}^* y_{hci}$ where $w_{hci}^* = w_{str_{hi}} w_{adj_{hci}}$. That is replacing $w_{b_{hci}}$, $c = 1, 2, \cdots, C_h$ with $w_{str_{hi}} = \dfrac{N_h}{n_h}$ in (14). Furthermore, $\left(\tilde{t}_h\right)_{adj} = \sum\limits_{c=1}^{C_h} \tilde{t}_{hc}$ and $\left(\tilde{t}_{str}\right)_{adj} = \sum\limits_{h=1}^{H} \left(\tilde{t}_h\right)_{adj}$. Therefore, using weighting class approach, the population total may be estimated by

$$\left(\tilde{t}_{str}\right)_{adj} = \sum_{h=1}^{H} \sum_{c=1}^{C_h} \sum_{i \in \mathcal{A}_{hc}} w_{hci}^* y_{hci} \tag{19}$$

Let $\mathbf{u}_h^* = (m_{h1}, \cdots, m_{hC_h}, n_{h1}, \cdots, n_{hC_h})'$ and let $\mathbf{u}_\bullet^* = (\mathbf{u}_1^*; \mathbf{u}_2^*; \cdots; \mathbf{u}_H^*)$. To show that $E\left[\left(\tilde{t}_{str}\right)_{adj} \mid \mathbf{u}_\bullet^*\right] \neq t_{str}$, define:

$$Z_{hci} = \begin{cases} 1, & i \in \mathcal{S}_{hc} \\ 0, & i \in \mathcal{S}_{hc}' \end{cases} \quad \text{and} \quad A_{hci} = \begin{cases} 1, & i \in \mathcal{A}_{hc} \\ 0, & i \in \mathcal{A}_{hc}' \end{cases}.$$

It can be shown that $E\left(Z_{hci}\right) = \dfrac{n_{hc}}{N_{hc}}$, $E\left(A_{hci}\right) = \dfrac{m_{hc}}{n_{hc}}$, $E\left(Z_{hci} A_{hci}\right) = \dfrac{m_{hc}}{N_{hc}}$ for $c = 1, 2, \cdots, C_h$ and $h = 1, 2, \cdots, H$, and $E\left[\left(\tilde{t}_{str}\right)_{adj} \mid \mathbf{u}_\bullet^*\right) = t_{str} + \sum\limits_{h=1}^{H} \sum\limits_{c=1}^{C_h} \dfrac{n_{hc}}{n_h} \left(\dfrac{N_h}{N_{hc}} t_{hc} - t_h\right)$.

The variance of $\left(\tilde{t}_{str}\right)_{adj}$ given $\mathbf{u}_{\bullet}^{*}$ may be estimated by replacing N_{hc} with $\frac{n_{hc}}{n_h} \cdot N_h$ in (18). That is

$$Var\left[\widetilde{\left(\tilde{t}_{str}\right)_{adj}} \mid \mathbf{u}_{\bullet}^{*}\right] = \sum_{h=1}^{H}\sum_{c=1}^{C_h}\left(N_h \cdot \frac{n_{hc}}{n_h}\right)^2\left(1 - \frac{n_h}{N_h}\frac{m_{hc}}{n_{hc}}\right)\frac{s_{hc}^2}{m_{hc}}. \tag{20}$$

If proportional allocation is used in the sample design, then $\frac{n_h}{N_h} = \frac{n}{N}$ for all $h = 1, 2, \cdots, H$. In (19), w_{hci}^{*} becomes $\frac{N}{n} \cdot w_{adj_{hci}}$ and (20) can be written as

$$Var\left[\widetilde{\left(\tilde{t}_{str}\right)_{adj}} \mid \mathbf{u}_{\bullet}^{*}\right] = \left(\frac{N}{n}\right)^2\sum_{h=1}^{H}\sum_{c=1}^{C_h} n_{hc}^2\left(1 - \frac{n}{N}\frac{m_{hc}}{n_{hc}}\right)\frac{s_{hc}^2}{m_{hc}}.$$

Oh and Scheuren (1983) proposed $MSE\left[\widetilde{\left(\tilde{t}_{str}\right)_{adj}} \mid \mathbf{u}_{\bullet}^{*}\right]$ as

$$MSE\left[\widetilde{\left(\tilde{t}_{str}\right)_{adj}} \mid \mathbf{u}_{\bullet}^{*}\right] = Var\left[\widetilde{\left(\tilde{t}_{str}\right)_{adj}} \mid \mathbf{u}_{\bullet}^{*}\right] + \left\{Bias\left[\widetilde{\left(\tilde{t}_{str}\right)_{adj}} \mid \mathbf{u}_{\bullet}^{*}\right]\right\}^2,$$

and $\left\{Bias\left[\widetilde{\left(\tilde{t}_{str}\right)_{adj}} \mid \mathbf{u}_{\bullet}^{*}\right]\right\}^2 = \sum_{h=1}^{H}\frac{N_h - n_h}{N_h - 1}\sum_{c=1}^{C_h}\frac{n_{hc}}{n_h^2}\left[N_h \cdot \frac{t_{a_{hc}}}{m_{hc}} - \left(\tilde{t}_h\right)_{adj}\right]^2$.

4. Two-Stage Cluster Sampling

A multi-stage random sample is constructed by selecting a sample in at least two stages. Sampling in stages enables one to reduce the population and to combine sampling procedures. Suppose the population can be divided into a number of subpopulations. Contrary to stratification, it is expected that the variance with respect to the measurements of interest is high within each subpopulation, whereas the variance with respect to the measurements of interest is low between subpopulations. The subpopulations are called clusters here, and they are usually naturally formed from, for example, communities, city blocks, and schools. In the first stage, a random sample of clusters is obtained, and a number of sub-clusters are formed within each selected cluster for selection in the next stage. The subgroups formed for selection in each stage prior to the final stage can be called the population units in general, and the process of selecting population units within each population unit obtained in the previous stage can be repeated until the sizes of the population units meet the requirements. In the last stage, a random sample of sampling units is obtained from each selected population unit.

4.1 Estimation in Two-Stage Cluster Sampling

Suppose that the population consists of L clusters, then in the first stage, let a simple random sample of l clusters be obtained. Let τ be the population total and let t_g be the total of cluster g for $g = 1, \cdots, L$. Let index set $\mathcal{U} = \{1, 2, \cdots, L\}$ denote the population, then $\tau = \sum_{g \in \mathcal{U}} t_g$. Let $\mathcal{S}$ denote the simple random sample of size l drawn from $\mathcal{U}$, then an unbiased estimator of τ is

$$\hat{\tau} = \sum_{g \in \mathcal{S}} w_g t_g \tag{21}$$

where $w_g = L/l$. The variance of $\hat{\tau}$ given l and L can be estimated by $\widehat{Var\left(\hat{\tau} \mid l, L\right)} = L^2\left(1 - l/L\right)\dfrac{s_t^2}{l}$ where $s_t^2 = (l-1)^{-1}\sum\limits_{g\in\mathcal{S}}\left(t_g - \bar{t}\right)^2$ and $\bar{t} = l^{-1}\sum\limits_{g\in\mathcal{S}} t_g$. These estimators are usable in one-stage cluster sampling, but additional work is required in two-stage cluster sampling.

Since, in the second stage, simple random sample g is to be drawn from cluster g for all $g \in \mathcal{S}$, then t_g in (21) is not observed but should be estimated. Let $\mathcal{U}_g$ denote cluster g having size N_g, and let $\mathcal{S}_g$ denote simple random sample g having size n_g drawn from $\mathcal{U}_g$ for $g \in \mathcal{S}$. Let y_{gi} be the measurement of interest obtained from sampling unit i in cluster g, then for $g \in \mathcal{S}$, an unbiased estimator of t_g is

$$\hat{t}_g = \sum_{i\in\mathcal{S}_g} w_{gi} y_{gi} \tag{22}$$

where $w_{gi} = \dfrac{N_g}{n_g}$. The variance of $\hat{t}_g$ given n_g and N_g can be estimated by $\widehat{Var\left(\hat{t}_g \mid n_g, N_g\right)}$

Combining (21) and (22), the unbiased estimator of τ for two-stage cluster sampling is given by

$$\hat{\hat{\tau}} = \sum_{g\in\mathcal{S}}\sum_{i\in\mathcal{S}_g} w_g w_{gi} y_{gi}. \tag{23}$$

Lohr (2009) proved the variance of $\hat{\hat{\tau}}$ has two components: the variability between clusters and the variability of sampling units within each cluster. The variance of $\hat{\hat{\tau}}$ can be estimated by

$$\widehat{Var\left(\hat{\hat{\tau}} \mid l, L, \mathbf{q}\right)} = L^2\left(1 - \frac{l}{L}\right)\frac{s_{\hat{t}}^2}{l} + \frac{L}{l}\sum_{g\in\mathcal{S}} N_g^2\left(1 - \frac{n_g}{N_g}\right)\frac{s_g^2}{n_g} \tag{24}$$

where $\mathbf{q} = \left\{(n_g, N_g)' \mid g \in \mathcal{S}\right\}$, $s_{\hat{t}}^2 = (l-1)^{-1}\sum\limits_{g\in\mathcal{S}}\left(\hat{t}_g - \hat{\bar{t}}\right)^2$, and $\hat{\bar{t}} = l^{-1}\sum\limits_{g\in\mathcal{S}}\hat{t}_g$.

4.2 Nonresponse in Two-Stage Cluster Sampling

Nonresponse appears in the second stage of two-stage cluster sampling. If a single cluster is selected in the first stage, then the nonresponse problem in the sample drawn from this cluster is identical to that in a regular simple random sampling. However, here one deals with l simple random samples with nonresponse. For $g \in \mathcal{S}$, suppose that there are m_g responding units in simple random sample g, then the response rate for sample g is $r_g = \dfrac{m_g}{n_g}$. Nonresponse appears if $r_g < 1$.

4.3 Weighting Adjustment in Two-Stage Cluster Sampling

Assume that it is possible to divide sample g into C_g classes by some available auxiliary information for all $g \in \mathcal{S}$ so that each unit in a class has an equal probability of responding. Let $\mathcal{S}_{gc}$ denote class c of sample g consisting of n_{gc} units and let $\mathcal{A}_{gc} \subset \mathcal{S}_{gc}$ denote a set of m_{gc} responding units in class c of sample g for $c = 1, 2, \cdots, C_g$ within each $g \in \mathcal{S}$, then $n_g = \sum\limits_{c=1}^{C_g} n_{gc}$ and $m_g = \sum\limits_{c=1}^{C_g} m_{gc}$. For $g \in \mathcal{S}$, let $\mathcal{U}_{g1}, \mathcal{U}_{g2}, \cdots$, and $\mathcal{U}_{gC_g}$ be the sub-cluster of $\mathcal{U}_g$ having size $N_{g1}, N_{g2}, \cdots$, and N_{gC_g}, respectively, so that $N_g = \sum\limits_{c=1}^{C_g} N_{gc}$

and $\mathcal{S}_{gc} \subset \mathcal{U}_{gc}$. Suppose that N_{gc} are known for all $g \in \mathcal{S}$ and $c = 1, 2, \cdots, C_g$. The post-stratified estimator for the total of cluster g where $g \in \mathcal{S}$ is

$$\left(\hat{t}_g\right)_{adj} = \sum_{c=1}^{C_g} \sum_{i \in \mathcal{A}_{gc}} w_{gci} y_{gci} \tag{25}$$

where $w_{gci} = w_{b_{gci}} w_{adj_{gci}}$, $w_{b_{gci}} = \dfrac{N_{gc}}{n_{gc}}$, and $w_{adj_{gci}} = \dfrac{n_{gc}}{m_{gc}}$. Thus, $w_{gci} = \dfrac{N_{gc}}{m_{gc}}$. The variance of $\left(\hat{t}_g\right)_{adj}$ given $\mathbf{u}_g = \left(m_{g1}, \cdots, m_{gC_g}, n_{g1}, \cdots, n_{gC_g}, N_{g1}, \cdots, N_{gC_g}\right)'$ is estimated by $\widehat{Var}\left[\left(\hat{t}_g\right)_{adj} \mid \mathbf{u}_g\right] = \sum_{c=1}^{C_g} N_{gc}^2 \left(1 - \dfrac{m_{gc}}{N_{gc}}\right) \dfrac{s_{gc}^2}{m_{gc}}$ where $s_{gc}^2 = \dfrac{1}{m_{gc}} \sum_{i \in \mathcal{A}_{gc}} \left(y_{gci} - \overline{y}_{gc}\right)^2$ and $\overline{y}_{gc} = \dfrac{1}{m_{gc}} \sum_{i \in \mathcal{A}_{gc}} y_{gci}$.

Since $\hat{\bar{\tau}} = \sum_{g \in \mathcal{S}} w_g \hat{t}_g$, then $\hat{\bar{\tau}}_{adj} = \sum_{g \in \mathcal{S}} w_g \left(\hat{t}_g\right)_{adj}$. When using post-stratified estimators in the second stage to compensate for nonresponse, the unbiased estimator of the population total is

$$\hat{\bar{\tau}}_{adj} = \sum_{g \in \mathcal{S}} w_g \sum_{c=1}^{C_g} \sum_{i \in \mathcal{A}_{gc}} w_{gci} y_{gci}. \tag{26}$$

The variance of $\hat{\bar{\tau}}_{adj}$ given $\mathbf{u}_\bullet = \{\mathbf{u}_g \mid g \in \mathcal{S}\}$ may be estimated by

$$\widehat{Var}\left(\hat{\bar{\tau}}_{adj} \mid l, L, \mathbf{u}_\bullet\right) = L^2 \left(1 - \frac{l}{L}\right) \frac{s_{\hat{t}'}^2}{l} + \frac{L}{l} \sum_{g \in \mathcal{S}} \sum_{c=1}^{C_g} N_{gc}^2 \left(1 - \frac{m_{gc}}{N_{gc}}\right) \frac{s_{gc}^2}{m_{gc}}$$

where $s_{\hat{t}'}^2 = (l-1)^{-1} \sum_{g \in \mathcal{S}} [\left(\hat{t}_g\right)_{adj} - \bar{\hat{t}}_{adj}]^2$, and $\bar{\hat{t}}_{adj} = l^{-1} \sum_{g \in \mathcal{S}} \left(\hat{t}_g\right)_{adj}$.

If using weighting class adjustment, then one can replace w_{gci} with $w_{gci}^* = \dfrac{N_g}{n_g} \cdot w_{adj_{gci}}$ in (25). Thus, the weighting class estimator of t_g for $g \in \mathcal{S}$ is

$$\left(\tilde{t}_g\right)_{adj} = \sum_{c=1}^{C_g} \sum_{i \in \mathcal{A}_{gc}} w_{gci}^* y_{gci}. \tag{27}$$

Let $\mathbf{u}_g^* = \left(m_{g1}, \cdots, m_{gC_g}, n_{g1}, \cdots, n_{gC_g}\right)'$. Given $\mathbf{u}_g^*$, for $g \in \mathcal{S}$, the variance of $\left(\tilde{t}_g\right)_{adj}$ and squared bias of $\left(\tilde{t}_g\right)_{adj}$ may be estimated by

$$\widetilde{Var}\left[\left(\tilde{t}_g\right)_{adj} \mid \mathbf{u}_g^*\right] = \sum_{c=1}^{C_g} \left(N_g \cdot \frac{n_{gc}}{n_g}\right)^2 \left(1 - \frac{n_g}{N_g} \frac{m_{gc}}{n_{gc}}\right) \frac{s_{gc}^2}{m_{gc}}$$

and

$$\left\{\widetilde{Bias}\left[\left(\tilde{t}_g\right)_{adj} \mid \mathbf{u}_g^*\right]\right\}^2 = \frac{N_g - n_g}{N_g - 1} \sum_{c=1}^{C_g} \frac{n_{gc}}{n_g^2} \left[N_g \overline{y}_{gc} - \left(\tilde{t}_g\right)_{adj}\right]^2,$$

respectively.

When using weighting class estimators in the second stage to compensate for nonresponse, the estimator of τ is

$$\tilde{\bar{\tau}}_{adj} = \sum_{g \in \mathcal{S}} w_g \sum_{c=1}^{C_g} \sum_{i \in \mathcal{A}_{gc}} w_{gci}^* y_{gci}. \tag{28}$$

Given $\mathbf{u}_{\bullet}^{*} = \left\{ \mathbf{u}_{g}^{*} \mid g \in \mathcal{S} \right\}$, the variance of $\tilde{\bar{\tau}}_{adj}$ and squared bias of $\tilde{\bar{\tau}}_{adj}$ may be estimated respectively by

$$\widetilde{Var}\left(\tilde{\bar{\tau}}_{adj} \mid l, L, \mathbf{u}_{\bullet}^{*}\right) = L^{2}\left(1-\frac{l}{L}\right)\frac{s_{\tilde{t}'}^{2}}{l} + \frac{L}{l}\sum_{g\in\mathcal{S}}\sum_{c=1}^{C_g}\left(N_g \cdot \frac{n_{gc}}{n_g}\right)^{2}\left(1-\frac{n_g}{N_g}\frac{m_{gc}}{n_{gc}}\right)\frac{s_{gc}^{2}}{m_{gc}}$$

and

$$\left[\widetilde{Bias}\left(\tilde{\bar{\tau}}_{adj} \mid l, L, \mathbf{u}_{\bullet}^{*}\right)\right]^{2} = \frac{L}{l}\sum_{g\in\mathcal{S}}\frac{N_g - n_g}{N_g - 1}\sum_{c=1}^{C_g}\frac{n_{gc}}{n_g^{2}}\left[N_g \overline{y}_{gc} - \left(\tilde{t}_g\right)_{adj}\right]^{2}$$

where $s_{\tilde{t}'}^{2} = (l-1)^{-1}\sum_{g\in\mathcal{S}}[\left(\tilde{t}_g\right)_{adj} - \bar{\tilde{t}}_{adj}]^{2}$, and $\bar{\tilde{t}}_{adj} = l^{-1}\sum_{g\in\mathcal{S}}\left(\tilde{t}_g\right)_{adj}$.

An approximate $100\,(1-\alpha)\,\%$ confidence interval for τ may be constructed by

$$\tilde{\bar{\tau}}_{adj} \pm z_{\alpha/2}\sqrt{\widetilde{MSE}\left(\tilde{\bar{\tau}}_{adj} \mid l, L, \mathbf{u}_{\bullet}^{*}\right)} \tag{29}$$

where $\widetilde{MSE}\left(\tilde{\bar{\tau}}_{adj} \mid l, L, \mathbf{u}_{\bullet}^{*}\right) = \widetilde{Var}\left(\tilde{\bar{\tau}}_{adj} \mid l, L, \mathbf{u}_{\bullet}^{*}\right) + \left[\widetilde{Bias}\left(\tilde{\bar{\tau}}_{adj} \mid l, L, \mathbf{u}_{\bullet}^{*}\right)\right]^{2}$ and $z_{\alpha/2}$ is the $(1-\alpha/2)$th percentile of the standard normal distribution.

5. Conclusion

Two very common weighting adjustments for unit nonresponse, post-stratification and weighting class methods, are described in simple random sampling, stratified random sampling, and two-stage cluster sampling. It is expected that these methods can be used to reduce nonresponse bias. It is inappropriate to fully rely on these weighting adjustments because the assumption of forming a group in which the nonresponse is completely negligible is rarely satisfied in the real world. A researcher should always try to avoid low response rates. Furthermore, item nonresponse and unit nonresponse normally appear together; thus, some imputation techniques may be necessary as complementary remedies since filling in missing values by using some known information could be efficient and useful for obtaining desirable classification in weighting adjustments.

REFERENCES

Cochran, W. G. (1977), *Sampling Techniques*, *third edition*. Wiley, New York.

Cornfield, J. (1944). "On Samples from Finite Populations," *Journal of the American Statistical Association*, 39: (226), 236-239.

Holt, D., and Elliot, D. (1991), " Methods of Weighting for Unit Non-Response," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 40: (3), 333-342.

Holt, D., and Smith, T. M. F. (1979), " Post Stratification," *Journal of the Royal Statistical Society. Series A (General)*, 142: (1), 33-46.

Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, *second edition*. Wiley, New York.

Lohr, S. L. (2009), *Sampling: Design and Analysis*, *second edition*. Brooks/Cole, Cengage Learning, Boston, MA.

Oh, H. L., and Scheuren, F. J. (1983), " Weighting Adjustment for Unit Nonresponse," In W. G. Madow, I. Olkin, and D. B. Rubin (Eds.), *Incomplete Data in Sample Surveys*, *Volume 2: Theory and Bibliographies* (pp. 143-184). Academic Press, New York.

Smith, T. M. F. (1991)," Post-Stratification" *Journal of the Royal Statistical Society. Series D (The Statistician)*, 40: (3), 315-323.