Weighting Adjustment Methods for Nonresponse in Surveys
Taylor Lewis, University of Maryland, College Park, MD

ABSTRACT

Unit nonresponse, which occurs when sampled entities fail to respond to a survey request, is a common problem faced by applied survey researchers. One class of techniques to mitigate potential biases in estimates based only on the respondents is to reweight the responding cases to better reflect the target population. This paper discusses several of these techniques, including adjustment cell reweighting, propensity score reweighting, poststratification, and raking. After a brief exposition of the underlying theory, the focus will be on demonstrating the SAS syntax necessary to implement these methods in practice.

BACKGROUND

A scientifically sound method for creating a roster of survey participants is to conduct a probability sample in which \( n \) units are selected at random from a list (sample frame) of \( N \) population units, each with a known and non-zero selection probability, \( \pi_i \). When \( \pi_i = \pi = n/N \) for all \( i \in n \), the method is termed simple random sampling, but equal probabilities of selection is not mandatory—alternative, unequal sample designs can be more precise than simple random sampling depending on the population structure and estimator(s) of interest (Cochran, 1977).

If data are collected on all sample units, one can assign a weight to the \( i^{th} \) sample unit based on the inverse of its selection probability, \( w^{(base)}_i = 1 / \pi_i \), and use it to calculate an unbiased estimate of a population total, \( \hat{Y} = \frac{\sum_{i \in n} w^{(base)}_i Y_i}{\sum_{i \in n} w^{(base)}_i} \) (Horvitz and Thompson, 1952). This weight is often referred to as the base weight and can be interpreted as the number of population units each sample unit represents. For instance, if the \( i^{th} \) sample unit has a weight of 5, it represents itself and 4 other comparable units in the population.

In practice, researchers are rarely able to secure participation and collect complete data on all sample units, especially when the unit being surveyed is a human or a business. Unit nonresponse is the term reserved for situations in which the sample unit fails to respond to a survey request, whereas item nonresponse indicates a particular survey item was left unanswered. A visualization of these two concepts is given in Figure 1.

For sake of an example, assume the goal of some survey is to measure four outcome variables on \( n \) sample units, denoted \( Y_1 \) through \( Y_4 \). A blue box represents a valid response while missing data are represented by a question mark. The data set on the left illustrates unit nonresponse whereas the right illustrates item nonresponse. While item nonresponse is a serious concern and a vast literature has developed proposing ways to compensate for it, as the title suggests, this paper discusses the common methods used in practice to adjust for unit nonresponse.

Arguably, the most common method for treating unit nonresponse is to ignore the \( m \) missing cases and use only the \( r \) complete cases (\( r + m = n \)) when calculating estimates. One immediately evident problem with this approach is that, for any outcome variable with strictly positive values, an estimated total will be biased downward since
Weighting Adjustment Methods for Nonresponse in Surveys, continued

\[ \hat{Y}_r = \sum_{i=r} w_{i(base)} y_i < \sum_{i=n} w_{i(base)} y_i \].

Moreover, the unadjusted sample mean using only the responding cases,

\[ \hat{Y}_r = \frac{\sum_{i=r} w_{i(base)} y_i}{\sum_{i=r} w_{i(base)}} \],

is no longer guaranteed to be unbiased for the population mean, \( \bar{Y} \). Whether or not the estimate is biased hinges on how similar the respondents and nonrespondents are with respect to the outcome variable.

For simple random samples, Groves and Couper (1998) note the bias in \( \hat{y}_r \) is 

\[ bias(\hat{y}_r) = \frac{m}{n}(\hat{y}_r - \hat{y}_m), \]

where \( \hat{y}_m \) is the mean of the \( m \) non-responding cases. In words, the bias equals the nonresponse rate (1 minus the response rate) times the difference in sample means between respondents and nonrespondents. An often overlooked corollary of this formula is that a low response rate (high nonresponse rate) does not imply greater bias; rather, it merely increases the risk of bias. A response rate of 80% could be more detrimental than a response rate of 20% if the former’s respondent/non-respondent mean difference is much larger than the latter’s. Also, the formula is variable-specific, so any evidence (or lack thereof) of bias in a particular estimate should not discredit the quality of all estimates based strictly on responding cases.

The Groves and Couper (1998) bias formula is applicable to the deterministic view of nonresponse, which makes the assumption the population consists of sample units from two mutually exclusive groups or strata: those who always respond and those who never respond. The stochastic view of survey nonresponse is an alternative perspective many believe to be more realistic. This view posits each sample unit is assumed to possess a fixed (but unknown) probability of responding to a survey request. Following the terminology of Rubin and Rosenbaum (1983), this is often called a response propensity and denoted by \( \phi_i \). Bethlehem (1988) showed that the expected nonresponse bias in a mean using only responding sample units equals 

\[ E(bias(\hat{y}_r)) = \sum_{i=1}^{N} (\phi_i - \bar{\phi})(y_i - \bar{y}), \]

where \( \bar{\phi} \) denotes the average response propensity across all population units and the expectation is over (hypothetically) repeated surveys. Thus, the bias is equal to the population covariance between the propensities and the outcome variable divided by the average propensity.

If we are to adopt the stochastic perspective, the three distinct missing data assumptions defined by Little and Rubin (2002) are useful in considering how harmful the nonresponse is and whether or not it can be remedied. The first assumes data are missing completely at random (MCAR), which is equivalent to \( \phi_i = \phi = \bar{\phi} \). Since the \( \phi_i \)'s do not vary, they are clearly uncorrelated with any outcome variable(s). This is the least harmful situation, as responding can be thought of a second stage of sampling. The unbiasedness in a sample mean demonstrated by Horvitz and Thompson (1952) would still hold.

The second assumption is that the data are missing at random (MAR), which implies the \( \phi_i \)'s vary only with regard to a sample unit’s vector of auxiliary variables, denoted \( \chi \). This vector consists of covariates observed for all sample units, most often from the sample frame. Conditional on the covariates, there is no variability in the \( \phi_i \)'s; in other words, there is no additional dependency between the likelihood of unit nonresponse and any outcome variable. This is the situation implicitly assumed by most of the weighting adjustment methods to be demonstrated in this paper.

The third assumption is the most difficult to address, data that are not missing at random (NMAR). This means there is a dependency between the \( \phi_i \)'s and the outcome variable above and beyond what can be accounted for by the auxiliary variables. For example, imagine a mail survey aimed at measuring the proportion of the electorate that voted in the most recent presidential election. If people who did not vote are less inclined to respond the survey request across all auxiliary variables on the sample frame (e.g., race/ethnicity, age, neighborhood), no weighting adjustment technique using those variables will be able to eliminate the bias. Rather sophisticated techniques are required to handle the NMAR situation; they will not be considered in this paper.

There are critics who believe nonresponse adjustment methods are futile and should not be performed. As Brick and Kalton (1996) point out, however, foregoing the adjustments is akin to assuming MCAR, which is far more questionable than the MAR assumption implicit in making them! Despite the fact that any one of the three

---

1 For unequal probability samples in which the base weights differ, the term \( m / n \) would be replaced by the base-weighted nonresponse rate.
classifications outlined by Little and Rubin (2002) is rarely possible to verify (or refute), the plausibility of the MAR assumption increases with a larger number of auxiliary variables.

AN EXAMPLE SURVEY

Assume an employee satisfaction survey was conducted on a sample of individuals who work in a large organization. From a personnel database (sample frame) of 17,219 employees, a sample of 8,583 was drawn, and these employees were sent a personalized link via email to a Web-based survey instrument containing a variety of attitudinal questions and a few demographics. Weekly reminder emails were sent to nonrespondents, but after a few weeks the survey closed with 4,558 completes corresponding to a response rate of 53.1%.

Suppose there is no item nonresponse present in the completed surveys, at least for the outcome variables of interest. Fortunately, there are auxiliary variables maintained in the personnel database known for everyone in the sample and population—candidates for the nonresponse weighting adjustment process. Four of them will be utilized during the exposition of the various weighting methods in this paper: gender, supervisory status, age, and minority status.

The population data are stored in the data set FRAME with the four auxiliary variables defined as follows:

- GENDER – M/F indicator of employee gender
- SUPERVISOR – 0/1 indicator of where the employee is a supervisor
- AGE – integer age of employee at time of survey
- MINORITY – Y/N indicator of minority status

The data set SAMPLE is a subset of FRAME, consisting of records for only the 8,583 sampled employees. In addition to the variables listed above, the data set contains a variable called WEIGHT_BASE equaling the inverse of each employee’s selection probability, or \((8,583 / 17,219)^{-1} = 2.01\). The 0/1 indicator variable RESPOND can be used to extract only the responding cases. There are two key outcome variables in the data set. The first is Q1, a dichotomous 0/1 indicator of whether an employee replied “yes” to the question “I like the kind of work I do.” The second is a continuous variable called LOS (length of service), which houses duration of employment with the organization in years. Since both variables are derived from the survey instrument, they are missing for nonrespondents, or records in which RESPOND=0.

ADJUSTMENT CELL METHOD

The first weighting adjustment method to be demonstrated is the adjustment cell method. The idea is to first partition the sample into \(c = 1, 2, \ldots, C\) mutually exclusive groups, or nonresponse adjustment cells, formed based on one or more auxiliary variables. Within each cell, one shifts the base weights of nonrespondents proportionally to respondents. Operationally, this is accomplished by multiplying the base weights of each respondent by a cell-specific adjustment factor \(f_c\), equaling the sum of the base weights for that cell over the entire sample, \(\sum_{i \in n_c} w_{(base)i}\), divided by the sum of the base weights for just the responding units, \(\sum_{i \in f_c} w_{(base)i}\). At the same time, weights of nonrespondents are set to 0. In other words, one creates a new weight, \(w_{(adj)i}\), as

\[
\left( \frac{\sum_{i \in n_c} w_{(base)i}}{\sum_{i \in f_c} w_{(base)i}} \right) w_{(base)i} = f_c w_{(base)i} \quad \text{for respondents while} \quad w_{(adj)i} = 0 \quad \text{for nonrespondents.}
\]

Note that the adjustment cell method operates under the MAR assumption. Response probabilities are assumed to vary across cell boundaries but not within. That is, data are assumed MCAR within an adjustment cell. Because respondents are viewed as a random subset of the original sample within a cell, one common weight inflation factor is all that is needed.

---

2 There are alternative ways to calculate \(f_c\). Though the form presented here (the inverse of the weighted response rate in cell \(c\)) is perhaps the most commonly used, Little and Vartivarian (2003) argue it can be inefficient and instead suggest using inverse of the unweighted cell response rate, or \(n_c / r_c\).
Returning to our hypothetical survey, suppose that nonresponse adjustment cells are formed by cross-classifying the sample by gender and minority status. The syntax and output below compares the base-weighted sums of sample units and respondents only for each of the four classifications. As should be evident, without any adjustments, the weighted sum within a cell underestimates the weighted sum of the sample units.

```
proc freq data=sample;
  table gender*minority / list nocum;
  weight weight_base;
  title1 'Sum of Base Weights for Sample Units';
run;

proc freq data=sample;
  where respond=1;
  table gender*minority / list nocum;
  weight weight_base;
  title1 'Sum of Base Weights for Respondents';
run;
```

```
Sum of Base Weights for Sample Units

The FREQ Procedure

<table>
<thead>
<tr>
<th>GENDER</th>
<th>Minority</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>N</td>
<td>5291.456</td>
<td>30.44</td>
</tr>
<tr>
<td>F</td>
<td>Y</td>
<td>3683.566</td>
<td>21.19</td>
</tr>
<tr>
<td>M</td>
<td>N</td>
<td>6468.01</td>
<td>37.21</td>
</tr>
<tr>
<td>M</td>
<td>Y</td>
<td>1937.973</td>
<td>11.15</td>
</tr>
</tbody>
</table>

Sum of Base Weights for Respondents

The FREQ Procedure

<table>
<thead>
<tr>
<th>GENDER</th>
<th>Minority</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>N</td>
<td>3177.303</td>
<td>34.42</td>
</tr>
<tr>
<td>F</td>
<td>Y</td>
<td>1861.021</td>
<td>20.16</td>
</tr>
<tr>
<td>M</td>
<td>N</td>
<td>3323.107</td>
<td>36.00</td>
</tr>
<tr>
<td>M</td>
<td>Y</td>
<td>868.7464</td>
<td>9.41</td>
</tr>
</tbody>
</table>
```

Output 1. PROC FREQ Output Summing the Base Weights of Respondents

Below is annotated syntax to create a new, nonresponse-adjusted weight called WEIGHT_ADJ_CELL derived from WEIGHT_BASE. Weights of respondents are inflated by the cell-specific factor, \( f_c \), defined above—called FACTOR_ADJ_CELL in the syntax below—while weights for nonrespondents are set to 0.
* sum the base weights for sample units;
proc freq data=sample noprint;
  table gender*minority / out=counts_s (rename=(count=count_s));
weight weight_base;
run;
* sum the base weights for respondents;
proc freq data=sample noprint;
  where respond=1;
  table gender*minority / out=counts_r (rename=(count=count_r));
weight weight_base;
run;
* create data set of adjustment cell weight inflation factors;
data cell_factors;
  merge counts_s
    counts_r;
  by gender minority;
factor_adj_cell=count_s/count_r;
keep gender minority factor_adj_cell;
run;
* merge these into sample data set and assign new, nonresponse-adjusted weight;
proc sort data=sample; by gender minority; run;
proc sort data=cell_factors; by gender minority; run;
data sample;
  merge sample (in=a)
    cell_factors (in=b);
  by gender minority;
if a;
weight_adj_cell=weight_base*factor_adj_cell*(respond=1);
run;

When summed within a cell for only the respondents, WEIGHT_ADJ_CELL should match the original base-weighted cell sum for the sample units. It is good practice to verify this holds true after completing the weight adjustments, to ensure all has proceeded as planned. From the PROC FREQ output below, we observe the sums match what was previously tabulated.

* verify adjusted weights of respondents match sum of base weights for the full sample;
proc freq data=sample;
  where respond=1;
  table gender*minority / list nocum;
weight weight_adj_cell;
run;

<table>
<thead>
<tr>
<th>GENDER</th>
<th>Minority</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>N</td>
<td>5291.456</td>
<td>30.44</td>
</tr>
<tr>
<td>F</td>
<td>Y</td>
<td>3683.566</td>
<td>21.19</td>
</tr>
<tr>
<td>M</td>
<td>N</td>
<td>6468.01</td>
<td>37.21</td>
</tr>
<tr>
<td>M</td>
<td>Y</td>
<td>1937.973</td>
<td>11.15</td>
</tr>
</tbody>
</table>

Output 2. PROC FREQ Output Summing the Weights of Respondents Inflated within Adjustment Cells

The method is rather straightforward once the cells have been defined. For occasions in which the number of auxiliary variables is small, there is little debate over how to structure the cells, as few options are available. More challenging is the task of restricting a large number of auxiliary variables to only the “best” one(s). Creating cells based on the cross-classification of a large number of variables is prone to empty or extremely small cell sizes, which
can yield unstable weight inflation factors that may increase variances. A common rule of thumb is to require each cell to contain at least 30 sample units (Lohr, 1999). If that threshold is not met, the small cell can be combined with a neighboring cell.

The ideal variables to employ are those highly related to the probability of response and the outcome(s) of interest (Bethlehem, 2002; Kalton and Flores-Cervantes, 2003). In a series of simulations, Little and Vartivarian (2005) found that variables characterized as such have the ability to reduce bias and even variances of sample means. They assert bias can only be reduced when the adjustment cell variable is strongly related to the outcome. Given the multi-purpose nature of most surveys, however, it is difficult to craft a single adjustment cell scheme that is optimal for all outcome variables of interest.

**PROPENSITY CELL METHOD**

Another commonly used technique is adjusting weights within propensity cells. Instead of arbitrarily forming cells based on a purposively selected subset of auxiliary variables, cells are formed by grouping sample units with similar estimated response propensities, which can be denoted \( \phi \). Although there are various ways to get the \( \phi \)'s for \( i \in n \), this paper will demonstrate how to do so via logistic regression.

The idea is to fit a logistic regression model on the entire sample data set using the response indicator variable as the outcome and auxiliary variables as predictors. After fitting the model, one can extract the estimated propensities directly then sort and rank them to form groups, or propensity strata, which serve as the nonresponse adjustment cells. A common number of propensity strata to form is five, a rule of thumb attributable to Cochran (1968), who concluded from empirical evidence that additional strata had little effect on estimates with the potential downside of an increase in variance.

The propensity modeling approach has certain distinct advantages. One is that there are a finite number of adjustment cells, regardless of the number of covariates in the model. Another is that continuous covariates can be more directly incorporated. When using the adjustment cell method, continuous covariates must first be discretized. Lastly, interactions between two or more covariates can easily be specified.

The following syntax estimates response propensities for the 8,583 sample units in our example survey's sample data set based on a logistic regression model which includes the four available covariates as well as all two-way interaction terms. A shorthand way to do this is to string all four covariate variables' names together in the MODEL statement separated by a pipe, and key @2 after the last is specified. The P= option in the OUTPUT statement of PROC LOGISTIC stores the estimated response propensity in a variable named P_HAT. In the PROC RANK step, these are ranked and classified into one of five propensity strata. The number of strata can be altered by the GROUPS= option in the PROC RANK statement. The VAR statement informs PROC RANK of the variable(s) to be ranked and the RANKS statement names the rank group identifier appearing in the output data set, which below is called PROPENSITY_CELL.

```plaintext
proc logistic data=sample;
  class gender minority;
  model respond(event='1') = age|gender|supervisor|minority @2;
  output out=sample p=p_hat;
run;
proc rank data=sample out=sample groups=5;
  var p_hat;
  ranks propensity_cell;
run;
```

Hereafter, the syntax to adjust the weights is functionally equivalent to that shown for the adjustment cell method, except that cells are defined by only one variable, PROPENSITY_CELL. To distinguish from the adjustment cell weight, the new propensity-cell-adjusted weight is called WEIGHT_ADJ_PROP_CELL.

```plaintext
proc logistic data=sample;
  class gender minority;
  model respond(event='1') = age|gender|supervisor|minority @2;
  output out=sample p=p_hat;
run;
proc rank data=sample out=sample groups=5;
  var p_hat;
  ranks propensity_cell;
run;
```
* sum base weights of all sample units;
proc freq data=sample noprint;
   table propensity_cell / out=counts_s (rename=(count=count_s));
weight weight_base;
run;
* sum base weights of responding sample units;
proc freq data=sample noprint;
   where respond=1;
   table propensity_cell / out=counts_r (rename=(count=count_r));
weight weight_base;
run;
* create data set of adjustment cell weight inflation factors;
data cell_factors;
   merge counts_s
      counts_r;
   by propensity_cell;
   factor_prop_cell=count_s/count_r;
   keep propensity_cell factor_prop_cell;
run;
* merge these into sample data set and assign new, nonresponse-adjusted weight;
proc sort data=sample; by propensity_cell; run;
proc sort data=cell_factors; by propensity_cell; run;
data sample;
   merge sample (in=a)
      cell_factors (in=b);
   by propensity_cell;
   if a;
   weight_adj_prop_cell=weight_base*factor_prop_cell*(respond=1);
run;

POSTSTRATIFICATION

Poststratification (Holt and Smith, 1979) was originally proposed as a method to balance a sample’s covariate distribution in the case of complete response (no unit nonresponse), but it is often used in practice to mitigate nonresponse or coverage imparities between the sample frame and target population. The method resembles that of the adjustment cell method, with the principle difference being control totals are not derived from the sample; rather, they are assumed known for the entire population, possibly from an external source.

Unlike many European nations, there is no population registry of United States citizens to serve as a survey sample frame. Many times, surveys of the general public begin with sample frames in which the sample unit is a landline telephone number. Unfortunately, this type of frame suffers from coverage errors, since cell-phone-only households would have a zero probability of selection into the sample, as would households who lack a landline telephone. Blumberg and Luke (2011) estimate the cell-phone-only cohort has reached over 30% of all households in the U.S. as of mid-2011. Assuming no supplemental cell phone frame was used, this type of survey would be wise to employ poststratification to calibrate sample weights to external benchmarks deemed more reliable, perhaps using population estimates maintained by the U.S. Census Bureau.

The benchmark source need not be external. Revisiting our hypothetical employee satisfaction survey, since the data set FRAME contains a record for all employees in the population, covariates can be cross-classified to get the known population total for a given cell, \( N_c \). Once again, this is in contrast to the adjustment cell method, which uses the base-weighted cell total \( \sum_{i \in n_c} w_{(base)i} = N_c \), albeit an unbiased estimate of \( N_c \).

The syntax below calibrates the weights of the 4,558 responding employees to match the known population totals for the cross-classification of gender and minority status. The first PROC FREQ run differs from the one shown in the adjustment cell method’s syntax in that the input data set is FRAME as opposed to SAMPLE and the WEIGHT statement no longer appears. The cell-specific adjustment factor is now defined as \( N_c / \left( \sum_{i \in n_c} w_{(base)i} \right) \), and applied to respondents’ base weight while the nonrespondents’ weights are set to 0. Below, this factor is named
FACTOR_POST and the adjusted weights for respondents are stored in the variable WEIGHT_ADJ_POST.

* sum base weights of all population units;
  proc freq data=frame noprint;
      table gender*minority / list out=counts_p (rename=(count=count_p));
  run;
* sum base weights of responding sample units;
  proc freq data=sample noprint;
      where respond=1;
      table gender*minority / list out=counts_r (rename=(count=count_r));
  weight weight_base;
  run;
* create data set of adjustment cell weight inflation factors;
  data cell_factors;
      merge counts_p
          counts_r;
      by gender minority;
      factor_post=count_p/count_r;
      keep gender minority factor_post;
  run;
* merge these into sample data set and assign new, nonresponse-adjusted weight;
  proc sort data=sample;       by gender minority; run;
  proc sort data=cell_factors; by gender minority; run;
  data sample;
      merge sample (in=a)
          cell_factors (in=b);
      by gender minority;
      if a;
      weight_adj_post=weight_base*factor_post*(respond=1);
  run;
* verify adjusted weights of respondents match base weight sum of full population;
  proc freq data=sample;
      where respond=1;
      table gender*minority / list;
  weight weight_adj_post;
  run;

The output from the last PROC FREQ is shown below. Note that the adjusted weights of respondents now sum to the sample frame’s (known) cross-classified totals of gender and minority status, which differ slightly from the base-weighted estimates using only the sample (shown previously with the adjustment cell method).

<table>
<thead>
<tr>
<th>GENDER</th>
<th>Minority</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>N</td>
<td>5270</td>
<td>30.61</td>
</tr>
<tr>
<td>F</td>
<td>Y</td>
<td>3682</td>
<td>21.38</td>
</tr>
<tr>
<td>M</td>
<td>N</td>
<td>6334</td>
<td>36.78</td>
</tr>
<tr>
<td>M</td>
<td>Y</td>
<td>1933</td>
<td>11.23</td>
</tr>
</tbody>
</table>

Output 3. PROC FREQ Output Summing the Poststratified Weights of Respondents

RAKING

There are occasions when benchmark totals are known, but not in the cross-classified sense necessary to conduct poststratification. For instance, respective marginal distributions of gender and minority status may be available, but not for the joint distribution of the two. Even when the desired joint distributions are available, cross-classifying even
Weighting Adjustment Methods for Nonresponse in Surveys, continued

A moderate number of covariates can result in the small cell size dilemma alluded to earlier. In these situations, raking is an algorithm that can be implemented to ensure the sum of respondents' weights agree with the marginal totals of two or more covariates.

The origins of raking can be traced to Deming and Stephan (1940), who first conceived the technique to ensure certain tabulations from 1940 U.S. Census and samples taken from it agreed with one another. The process begins by making a poststratification adjustment using marginal totals of the first covariate, or raking dimension. Then a similar adjustment is made for the second covariate, acknowledging this may cause the totals with respect to the first covariate to fall out of agreement. Nonetheless, all covariates are calibrated sequentially. After adjusting the last covariate, the algorithm returns the first and cycles through once more. The process continuous until the adjustment factors and/or weights change less than some specified tolerance.

There is little need to write code from scratch to conducting raking, because the %RAKING SAS macro developed by Izrael, Hoaglin, and Battaglia (2000) is free and straightforward to use. Revisiting the hypothetical employee satisfaction survey, suppose one wanted to use all four available covariates as raking dimensions. The variables MINORITY, SUPERVS, and GENDER are all dichotomous, but AGE is continuous. Because raking variables must be categorical, a new variable called AGECAT was created as a 0/1 indicator variable of whether the employee is over the age of 40.

The first step is to compute and store these marginal totals based on the sample frame, accomplished by the PROC FREQ syntax below. The output data sets housing the totals are given the same name as the covariates themselves and the default COUNT variable output by PROC FREQ is renamed to MRGTOTAL, a key variable the %RAKING macro will be seeking.

```sas
* create the marginal totals to pass into the raking macro;
proc freq data=frame noprint;
table gender / out=gender (drop=percent rename=(count=MRGTOTAL));
table agecat / out=agecat (drop=percent rename=(count=MRGTOTAL));
table minority / out=minority (drop=percent rename=(count=MRGTOTAL));
table supervisor / out=supervisor (drop=percent rename=(count=MRGTOTAL));
run;
```

The macro definition can be directly copied and pasted into one's SAS program from the Izrael et al. (2000) paper. Here, it is assumed this has been done and so the macro has been compiled. The following code below invokes the macro:

```sas
%RAKING(inds=respondents, /* input data set */
  outds=respondents_raked, /*output data set */
  inwt=weight_base, /*weight to be raked */
  outwt=weight_raked, /* resulting raked weight */
  freqlist=gender agecat minority supervisor,
  /*list of data sets with marginal freqs or control totals…
   must contain name of raking variable and either variable
   PERCENT or MRGTOTAL */
  varlist=gender agecat minority supervisor,
  /* list of variables to be raked */
  numvar=4, /* number of raking variables */
  cntotal=, /* must be used when raking to PERCENTs */
  trmprec=1, /* weight precision req. to terminate */
  numiter=50) /* max number of iterations */;
```

The first parameter (INDS=) specifies the input data set. Notice how the data set RESPONDENTS is used instead of the full SAMPLE data set as was done in previous examples. There is no parameter to identify a respondent, so the input data set should be filtered accordingly. The next parameter (OUTDS=) names the output data set, which will contain all variables in the input data set plus the raked version of the input weight (INWT= parameter) called weight to be raked. The FREQLIST= parameter is used to list the marginal total data sets, which were defined in the PROC FREQ step above. The VARLIST= parameter identifies the corresponding variables in the respondent data set. The macro also requires the count of variables (or marginal total data sets) be specified in the NUMVAR= parameter. The CNTOTAL= parameter is an optional way to specify the grand total population size, only necessary when margins are given as percentages as opposed to weighted sums. The macro essentially converts percentages to totals by multiplying each by the numeric value given in CNTOTAL=.
The parameter TRMPREC= allows users to control the precision of calibrated (raked) weights relative to the control totals. Setting TRMPREC=1 requires the sum of each covariate categories’ raked weights to be within 1 of the respective control total. The final parameter is NUMITER=, which assigns a maximum number of iterations to be performed if the raked weights’ totals have not converged to the TRMPREC= criteria. Typically, convergence occurs within a few iterations, but not always. For discussions of possible causes for lack of convergence and corresponding remedies, the reader is referred to Brick et al. (2009) and Battaglia et al. (2009).

Below is abridged output after submitting the %RAKING macro with the parameters above. At the first iteration, the macro outputs each covariate categories' base-weighted sum in the “Calculated Margin” column alongside the marginal control total input by the user. The column “Difference” quantifies how far the two lie from one another. As we have already seen, the summed base weights of respondents underestimate the population controls. The macro then proceeds to make a sequence of poststratification adjustments until the convergence criteria is met for all covariates, which for the data at hand occurs after the fourth iteration. Intermediate output is suppressed, but one can see from the final iteration that the “Difference” column is less than 1 for all covariate categories.

<table>
<thead>
<tr>
<th>Obs</th>
<th>GENDER</th>
<th>Calculated</th>
<th>Control</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>5038.32</td>
<td>8952</td>
<td>3913.68</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>4191.85</td>
<td>8267</td>
<td>4075.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>AGE CAT</th>
<th>Control</th>
<th>Calculated</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>13531</td>
<td>13399.88</td>
<td>131.117</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3688</td>
<td>3819.12</td>
<td>-131.117</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>MINORITY</th>
<th>Control</th>
<th>Calculated</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>11604</td>
<td>12204.82</td>
<td>-600.815</td>
</tr>
<tr>
<td>2</td>
<td>Y</td>
<td>5615</td>
<td>5014.18</td>
<td>600.815</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>SUPERVISOR</th>
<th>Control</th>
<th>Calculated</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>15321</td>
<td>15246.19</td>
<td>74.8070</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1898</td>
<td>1972.81</td>
<td>-74.8070</td>
</tr>
</tbody>
</table>
.Raw text:

<table>
<thead>
<tr>
<th>Margin</th>
<th>Control</th>
<th>Calculated</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td>GENDER</td>
<td>Total</td>
<td>margin</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>8952</td>
<td>8951.72</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>8267</td>
<td>8267.28</td>
</tr>
<tr>
<td>======</td>
<td>======</td>
<td>===========</td>
<td>===========</td>
</tr>
<tr>
<td>17219</td>
<td>17219.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Raking by AGECAT, iteration - 4

<table>
<thead>
<tr>
<th>Margin</th>
<th>Control</th>
<th>Calculated</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td>AGECAT</td>
<td>Total</td>
<td>margin</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>13531</td>
<td>13530.93</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3688</td>
<td>3688.07</td>
</tr>
<tr>
<td>======</td>
<td>======</td>
<td>===========</td>
<td>===========</td>
</tr>
<tr>
<td>17219</td>
<td>17219.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Raking by MINORITY, iteration - 4

<table>
<thead>
<tr>
<th>Margin</th>
<th>Control</th>
<th>Calculated</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td>Minority</td>
<td>Total</td>
<td>margin</td>
</tr>
<tr>
<td>1</td>
<td>N</td>
<td>11604</td>
<td>11603.91</td>
</tr>
<tr>
<td>2</td>
<td>Y</td>
<td>5615</td>
<td>5615.09</td>
</tr>
<tr>
<td>======</td>
<td>======</td>
<td>===========</td>
<td>===========</td>
</tr>
<tr>
<td>17219</td>
<td>17219.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Raking by SUPERVISOR, iteration - 4

<table>
<thead>
<tr>
<th>Margin</th>
<th>Control</th>
<th>Calculated</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td>SUPERVISOR</td>
<td>Total</td>
<td>margin</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>15321</td>
<td>15321.00</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1898</td>
<td>1898.00</td>
</tr>
<tr>
<td>======</td>
<td>======</td>
<td>===========</td>
<td>===========</td>
</tr>
<tr>
<td>17219</td>
<td>17219.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**** Program terminated at iteration 4 because all calculated margins differ from Marginal Control Totals by less than 1

Output 4. Abridged Output from the %RAKING Macro

Enhancements to the original %RAKING macro were proposed by Izrael et al. (2004) and Izrael et al. (2009). In the former, additional parameters were added to help visualize the convergence process to facilitate remediing convergence failures. The latter introduced capabilities to trim the raked weights, which some researchers prefer to do in an effort to control variance increases attributable to highly variable weights. The core raking procedure embedded in the macros is essentially the same, however, so for simplicity only the original macro described in Izrael
et al. (2000) was illustrated in this paper.

**USING THE NEW WEIGHT**

To use the nonresponse-adjusted weight, simply list the variable in the WEIGHT statement of any applicable procedure. Several SAS PROCs prefixed with SURVEY are available for the analysis of survey data sets such as the one illustrated in this paper. For instance, PROC SURVEYMEANS is the analogous procedure to PROC MEANS with built-in capabilities to account for complex survey features (such as the unequal respondent weights) when computing measures of variability. The syntax below requests the estimate and standard error of our two key estimates, a 0/1 indicator of whether employees replied yes to the question “I like the kind of work I do” (Q1) and a continuous variable housing length of service, in years, with the organization (LOS). There are five separate calls to PROC SURVEYMEANS: one to get the unadjusted figures using the base weights and four subsequent calls using the four adjusted weights.

```sas
proc surveymeans data=sample nobs mean stderr;
  var Q1 LOS;
  weight weight_base;
  title1 '1) Unadjusted Estimates (Using WEIGHT_BASE)';
run;
proc surveymeans data=sample nobs mean stderr;
  var Q1 LOS;
  weight weight_adj_cell;
  title1 '2) Estimates Using Adjustment Cell Weight';
run;
proc surveymeans data=sample nobs mean stderr;
  var Q1 LOS;
  weight weight_adj_prop_cell;
  title1 '3) Estimates Using Propensity Cell Weight';
run;
proc surveymeans data=sample nobs mean stderr;
  var Q1 LOS;
  weight weight_adj_post;
  title1 '4) Estimates Using Poststratified Weight';
run;
proc surveymeans data=respondents_raked nobs mean stderr;
  var Q1 LOS;
  weight weight_raked;
  title1 '5) Estimates Using Raked Weight';
run;
```

### 1) Unadjusted Estimates (Using WEIGHT_BASE)

The SURVEYMEANS Procedure

**Data Summary**

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>8583</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of Weights</td>
<td>17381.004</td>
</tr>
</tbody>
</table>

**Statistics**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Error of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>4558</td>
<td>0.739798</td>
<td>0.006499</td>
</tr>
<tr>
<td>LOS</td>
<td>4558</td>
<td>16.596361</td>
<td>0.171057</td>
</tr>
</tbody>
</table>

### 2) Estimates Using Adjustment Cell Weight

The SURVEYMEANS Procedure
3) Estimates Using Propensity Cell Weight

The SURVEYMEANS Procedure

Data Summary

Number of Observations  8583
Sum of Weights          17381.004

Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Error of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>4558</td>
<td>0.736251</td>
<td>0.006591</td>
</tr>
<tr>
<td>LOS</td>
<td>4558</td>
<td>16.624118</td>
<td>0.173255</td>
</tr>
</tbody>
</table>

4) Estimates Using Poststratified Weight

The SURVEYMEANS Procedure

Data Summary

Number of Observations  8583
Sum of Weights          17219

Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Error of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>4558</td>
<td>0.737722</td>
<td>0.006554</td>
</tr>
<tr>
<td>LOS</td>
<td>4558</td>
<td>16.554057</td>
<td>0.172186</td>
</tr>
</tbody>
</table>

5) Estimates Using Raked Weight

The SURVEYMEANS Procedure

Data Summary

Number of Observations  4558
Sum of Weights          17219

It is immediately evident that estimates hardly change after introducing any one of the nonresponse-adjusted weights. Going back to the discussion in the background section, this inefficaciously could be a function of a weak relationship between the auxiliary variables and the outcome and/or a weak relationship between the auxiliary variables and the probability of an employee responding. It is possible, however, more marked changes would be observed for other outcome variables or other estimators.

One can also see how the adjusted weights yield slightly higher standard error estimates relative to the base weights. Again, this is likely attributable to the additional variability the adjusted weights introduce (Kish, 1992). Indeed, there is no variability in the base weights, which are assigned as $w_i = 2.01$ for all $i \in n$.

PROC SURVEYMEANS also outputs the sum of the weights for the variable listed in the WEIGHT statement. Note how the poststratified and raked weights sum to the known sample frame total, $N = 17,219$, while the adjustment cell and propensity cell methods sum to the unbiased estimate of the sample frame total computed from the base weights, $\sum_{i \in n} w_{\text{base}i} = \hat{N} = 17,381.004$. This is expected, because the former two methods calibrate weights to counts of the FRAME data set, whereas the latter two calibrate to the sum of the base weights in the SAMPLE data set.

DISCUSSION

This paper has explored some of the common weight adjustment procedures used in practice to compensate for unit nonresponse in surveys, which occurs when sampled units fail to respond to a survey request. While the theory behind the methods may seem complex to the novice reader, it is hoped the SAS syntax demonstrated to carry out these methods was straightforward.

The methods discussed in this paper are not all-inclusive. For instance, Deville and Särndal (1992) note how poststratification and raking are but two special cases of a larger class of calibration estimators. Additionally, alternative methods to define nonresponse adjustment cells are available, including regression tree modeling via algorithms such as CART, SEARCH, or CHAID (Breiman et al., 1993). These alternatives, however, operate under the same central objective of the adjustment cell and propensity cell methods: to use covariates known for the entire sample to explain the survey response process. Regardless of how cells are created, MCAR missingness is assumed within each.

Lastly, although a separate section was devoted to each method, it should be stressed that practitioners often use two or more methods in combination with one another. For instance, nonresponse adjustments might be implemented via propensity cells, with the resulting weights poststratified or raked to known totals, possibly with a different set of covariates used in each step.

REFERENCES


**CONTACT INFORMATION**

Your comments and questions are valued and encouraged. Contact the author at:

Name: Taylor Lewis  
Enterprise: University of Maryland  
Address: 1218 LeFrak Hall  
City, State ZIP: College Park, MD 20742  
Work Phone: 202-686-0796  
E-mail: tlewis@survey.umd.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.