Multidimensional Control Totals for Poststratified Weights

Darryl V. Creel and Mansour Fahimi
RTI International, 6110 Executive Blvd., Suite 905, Rockville, MD 20852

Abstract

The weighting process for surveys with complex designs can be a major undertaking, often requiring several adjustment steps. While calculating the sampling weights, care must be applied to ensure that the resulting survey estimates are as efficient as possible – a nontrivial process that must arrive at an equitable balance between the two competing features of weights: bias reduction and variance inflation. By nature, complex surveys often encounter unanticipated weighting issues that further complicate this process, proper resolutions of which should take into account the specific analytical objectives of the given survey. This paper discusses a number of such issues and the solutions that were implemented while calculating the analysis and replicate weights for the 2004 National Study of Postsecondary Faculty (NSOPF:04) survey.

Keywords: Complex survey weighting, poststratification, raking

1. Introduction

Virtually, all surveys encounter some form of nonresponse at the item level, unit level, or both levels. While item nonresponse is typically handled via imputation, unit nonresponse is often addressed through weight adjustments. Falling to account for nonresponse can result in survey estimates that carry varying degrees of bias. A general approach that attempts to minimize this potential bias involves adjusting the sampling (or design) weights so that the responding units can also represent those units that have failed to respond. Variations of this approach include weighting class adjustments, poststratification, and raking, as well as, weight adjustment techniques that involve statistical modeling through logistic regression and generalized exponential models.

To account for the nonresponse these weighting approaches rely on the assumption that the data is missing at random. That is, the response propensity, ϕ, depends only on the auxiliary variables, $x_i, i \in \{1,...,P\}$, where P is the number of auxiliary variables used in the model. Ratio adjustments techniques, such as weighing class adjustments and poststratification models assume all respondents in a class have the same response propensity, i.e., $\phi_{Ci} = \phi_{C,} \forall i \in C.$ These classes can be formed by any number of the P auxiliary variables. With raking adjustments the model assumes that all respondents that fall in the cross classification of the values of the two auxiliary variables H and K have the same response propensity, i.e., $\phi_{hki} = \phi_{hk,} \forall i \in \{h \in \{1,...,H\} \, and \, k \in \{1...K\}$, where the number of auxiliary variables can be extended to include all P auxiliary variables. The logistic regression model assumes each unit has a response propensity, i.e., $\phi_i = \frac{1}{1+e^{\alpha + \sum \beta_i x_i}}, \forall i,$ where the x_i is the *P-variate* auxiliary variable. In this case, not implementing any method to account for nonresponse amounts to assuming that everyone has the same response propensity, i.e., $\phi_i = \phi_{,} \forall i.$

The section on weighting alternatives will discuss how these response propensities are employed in traditional methods of calculating weighting adjustment factors. The section addressing the issues in NSPOF:04 describes a number of specific issues that were encountered when calculating the analysis weights for the NSOPF:04 and how they were addressed. The section covering the diadnostics for weighting describes a number of diagnostic tests that were used to assess the effects of weighting. The final section provides a conclusion.

2. Weighting Alternatives

This paper discusses four common alternatives that are often used for calculating survey weights. The first is the weighting class adjustment, for which the estimated response probability for weighting class C will be the sum of the weights for the responding units in class C

divided by the sum of the weights of the eligible units in the given class. That is, the estimated response propensity for class *C* is:

$$\hat{\phi}_C = \frac{\sum_{k\in R} w_k}{\sum_{j\in E} w_j},$$

where w_k or w_j is the unit's weight in class *C*, *R* is the subset of responding units in class *C*, and *E* is the set of eligible units in class *C*. Subsequently, the adjusted weight for the responding unit *i* in class *C* is given by:

$$w_i^* = \frac{w_i}{\hat{\phi}_C} = w_i \frac{\sum_{j\in E} w_j}{\sum_{k\in R} w_k}.$$

While the poststratification alternative is very similar to the weighting class adjustment, with this method it is the frame counts or some other external totals that are used within each class for calculation of the adjustment factors. Here, the poststratified adjusted weight for a responding unit *i* in class *C* is given by:

$$w_i^* = w_i \frac{N_C}{\sum_{k\in R} w_k},$$

where N_C is the control total for class *C*. Unlike the weighting class adjustment method where the auxiliary information is only needed for the sampled units, here, it will be necessary to have the auxiliary information for all sampling units on the target universe belonging to each class. As a result this method of adjustment can also account for frame inadequacies such as undercoverage.

Weighting class adjustments and poststratification force the sum of the weights for the classes to reflect the joint distribution of the auxiliary variables used to form each class. Alternatively, in the absence of the joint distribution, it is possible to have the sum of the weights for the classes match the marginal distributions of the auxiliary variables used to form these classes. Under this alternative, the distribution of weights does not match the joint distribution of the auxiliary variables, which are usually unknown, rather they match the marginal distributions of the variables that form the classes. With this method, which is typically referred to as raking, the following criteria will be simultaneously satisfied:

$$T_{Rh} = \sum_{k=1}^{K}\sum_{i=1}^{n_{h,k}} w_{hki} \text{ and } T_{Ck} = \sum_{h=1}^{H}\sum_{i=1}^{n_{h,k}} w_{hki},$$

where T_{Hh} and T_{Kk} represent the distributions of the first and second auxiliary variables, with n_k and n_h identifying the number of responding units at each level of the given auxiliary variable. This approach can also be extended to the *P* auxiliary variables.

When many variables are used to form adjustment classes, it is possible to end up with sparse adjustment classes. Logistic regression modeling, which calculates the estimated response propensity as $\hat{\phi}_i = \frac{1}{1+e^{\hat{\alpha}+\sum\hat{\beta}_i x_i}}, \forall i$, can offer one way around this problem. By producing a response probability for each unit, the adjusted weight for the responding units is given by:

$$w_i^* = w_i \frac{1}{\hat{\phi}_i} = w_i(1+e^{\hat{\alpha}+\sum\hat{\beta}_i x_i}).$$

Alternatively, the estimated response propensity can be sorted and grouped together by similar values to form classes, within which a method of adjustment is implemented. This alternative has the advantage that all of the available information from auxiliary variables is distilled into one number and used to create weighting classes.

The final methodology is a general weight calibration[1] model called the generalized exponential model (GEM), which is a generalization of the commonly used raking-ratio method (Folsom and Singh, 2000). Here, specific distance functions are minimized by incrementally perturbing the initial weights within certain bounds while simultaneously matching the control totals. Explicitly, this method is a generalization of Deville and Sarndal's (1992) logit method:

[1] Weight calibration describes weighting adjustments that simultaneously adhere to control totals.

$$a_k(\lambda) = \frac{l(u-1) + u(1-l)\exp(Ax_k'\lambda)}{(u-1) + (1-l)\exp(Ax_k'\lambda)},$$

$$where\, l < 1 < u \, and \, A = \frac{u-l}{(u-1)(1-l)}.$$

$$As\, l \to 0 \, and \, u \to \infty, a_k(\lambda) \to \exp(x_k'\lambda).$$

GEM has a centering constant and does not required that the bounds have to be uniform. The centering constant is c_k and the specific bounds (l_k, u_k), where *k* is a unit in the sample for the adjustment factor

$$a_k(\lambda) = \frac{l_k(u_k - c_k) + u_k(c_k - l_k)\exp(A_k x_k'\lambda)}{(u_k - c_k) + (c_k - l_k)\exp(A_k x_k'\lambda)},$$

$$where\, l_k < c_k < u_k \, and \, A_k = \frac{u_k - l_k}{(u_k - c_k)(c_k - l_k)}.$$

$$As\, l_k \to 1, c_k \to 2, and\, u \to \infty, a_k(\lambda) \to 1 + \exp(x_k'\lambda).$$

For our situation, GEM allows for different bounds to be placed on the three possible adjustment factors - low extreme, nonextreme, and high extreme - where adjustment factors are considered extreme when they fall outside of the interval defined by the median$\pm$3×interquartile range. The control of the bounds for the extreme weights, along with the incorporation of control totals, gives GEM the flexibility to create nonresponse and poststratification adjustment factors without creating extreme weights. This program has been used to create the NSOPF:04 analysis weights.

3. Issues Addressed in NSOPF:04

NSOPF:04, a study of postsecondary faculty and instructional staff, has relied on a two-stage sample. The first stage consisted of sampling institutions while in the second stage faculty members were selected amongst the sampled institutions. Consequently, the fianl analysis weight for a responding faculty was calculated as the product of his/her institution and individual weights. Moreover, a separate set of weights were calculated to be used for analysis of the institution survey data, since not all institutions from which faculty participated in NSOPF:04 had responded to the corresponding institution survey.

For calculation of the faculty analysis weights, the initial faculty weights – the product of the faculty design sampling weight and his/her corresponding institution weight – were adjusted for multiplicity, unknown eligibility, and nonresponse before they were poststratified to control totals. Moreover, replicate weights were created for calculation of standard errors that included a similar form of poststratification. As discussed in Issue: Cell Collapsing section, certain adjustment classes had to be collapsed to ensure proper convergence in GEM.

A similar set of adjustments were implemented for calculation of the institution analysis weights, including an adjustment for nonresponse. The final weights for each institution were then calibrated to ensure that estimates of total part- and full-time faculty members obtained from the two surveys would coincide, as discussed in the Issue: Control Totals section.

4. Issue: Identifying Respondents

The NSOPF:04 survey employed two questionnaires, one for gathering basic information from each sample institution and a second and more detailed one for interviewing faculty members selected from lists secured from sampled institutions. Sampled institutions that responded to the institution questionnaire were considered respondents to the institution survey. Institutions providing faculty lists without responding to the institution questionnaire were included only to the extent their selection probabilities were needed to calculate the faculty design weights.

5. Issue: Control Totals

For NSOPF:04 there were three possible sets of estimates for faculty counts: weighted estimates from the institution survey, weighted estimates from the faculty survey, and an auxiliary set of estimates obtained from the Winter 2003-04 Employees by Assigned Position Survey (EAP:03). The latter source was used to create postratification totals, or control totals, for the faculty analysis weights, aggregates of which were used to calibrate the institution analysis weights.

One of the necessary conditions for poststratification is that the needed population counts should be known. Since the EAP:03 source only contained information for instructional faculty, it could not directly be used to create the needed control totals, as NSOPF:04

also included non-faculty who provide instruction but do not have teaching as their principal activity. Consequently, estimates of total for the latter group had to be estimated from the survey data prior to the poststratification step. For this purpose, the survey data were divided into two subsets: one for which the EAP:03-based population counts were known and one for which we the needed counts were unknown. The first subset was poststratified to the EAP:03 totals, while nonresponse-adjusted weights for the latter subset were used to estimate the outstanding sub-population. In the final step, the resulting two datasets along with the corresponding weights were merged to create the combined set of analysis weights. As a result of this process, the weighted totals from the combined dataset aggregate to estimates that are slightly larger than what can be obtained from the EAP:03.

6. Issue: Cell Collapsing

In order to construct the poststratification adjustment factors for faculty, the corresponding input weights were adjusted along two dimensions, as well as their interactions. One dimension consisted of the joint distribution of employment status (full- and part-time) within each of the 10 institution types that were used for sample selection. The other dimension was indexed by institution type, race/ethinicity, and gender of faculty members. In two of the resulting poststrata respondents had to be collapse along the race categories to ensure stable adjustment factors. Analogously, similar measures where taken when poststratifying the replicate weights. For this purpose, a somewhat coarser set of the poststratification control totals were constructed by eliminating the three-way interactions. That is, final resplicate weights were poststratified along the main dimensions and the resulting three two-way interactions.

7. Diagnostics for Weighting

Survey weights must be calculated while taking into account the potential inflation that can results in estimates of standard errors. While GEM provides a affective tool keep such ill effects under control as various weight adjustment factors are calculated, it was necessary to monitor the measure of variance inflation, often refer to as unequal weighting effect (UWE) defined by Kish (1992):

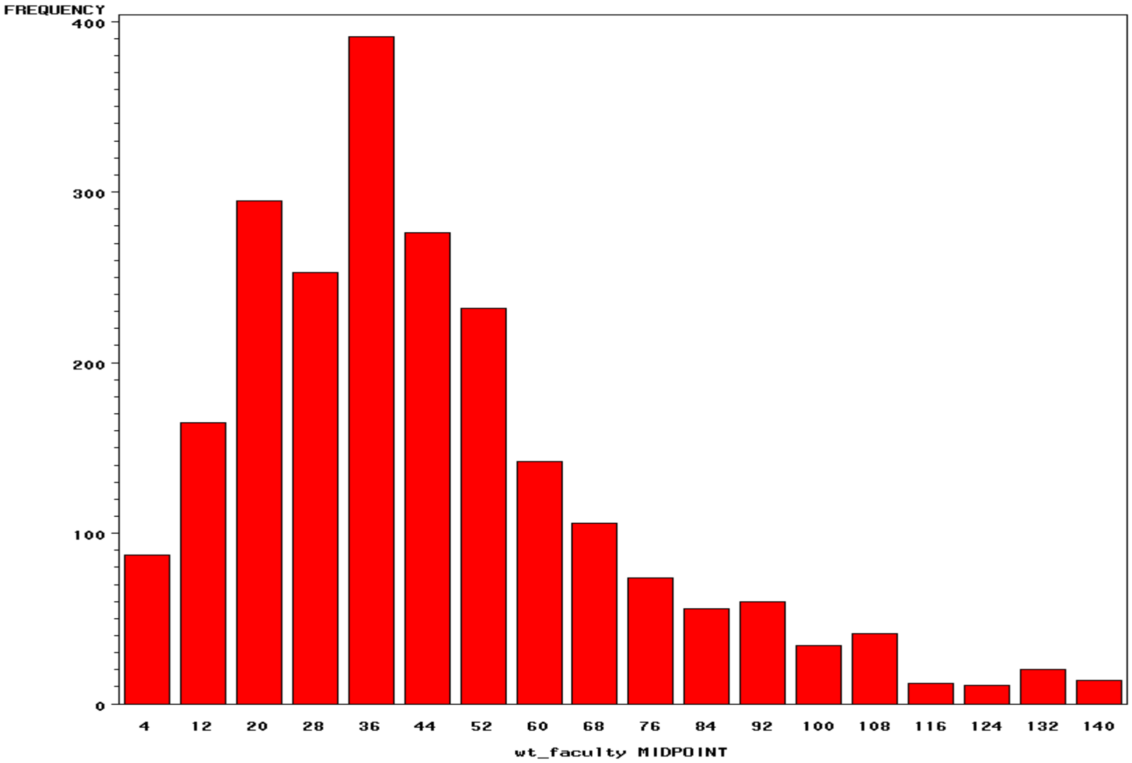
$$1+((cv(w))^2 = \frac{n\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2}$$

where *cv(w)* represents the coefficient of variation of the weights. This measure was calculated at each stage of the weighting process to assess the impact of the adjustments on the weights. Table 1 provides a summary of the number of responding units and the resulting UWE factors by institution type. We also monitored the distribution of the adjustement factors and the distribution of the weights at each stage of the weighting process. An example of the weight distribution is in Chart 1.

Table 1. Analysis Weight Information

Stratum	Count	Minimum Weight	Median Weight	Mean Weight	Maximum Weight	Coefficient Of Variation	UWE
Overall	26108	0.04	37	46	621	0.76	**1.58**
1	7464	0.49	35	42	167	0.56	**1.32**
2	2617	1.63	48	56	177	0.57	**1.33**
3	513	1.53	30	41	195	0.90	**1.82**
4	6416	0.28	44	57	216	0.79	**1.63**
5	106	1.13	124	134	621	0.81	**1.65**
6	3161	0.04	36	44	187	0.69	**1.47**
7	2269	0.41	40	44	143	0.61	**1.37**
8	2523	0.21	22	27	151	0.86	**1.73**
9	190	1.37	21	45	365	1.29	**2.67**
10	849	0.10	30	42	207	1.02	**2.03**

Chart 1: Example of Weight Distribution



The distribution of the analysis weights is skewed to the right, but there are no extreme outliers. As a final check, point estimates and their associated standard errors were calculated for a number of the key outcome measures. These estimates were checked against previous estimates of the same measures for external validity, when available.

8. Conclusion

Using GEM allowed us the ability to maintain control over the adjustment factors at each stage of the weighting process. Controlling the adjustment factors at each stage minimized the creation of extreme weights and helped keep the design effects for unequal weighting relatively low.

References

Deville, J. and Sarndal, C. (1992). *Calibration Estimation in Survey Sampling*. Journal of the American Statistical Association, 87, 376-382.

Folsom RE Jr, Singh AC (2000). *A Generalized Exponential Model for a Unified Approach to Sampling Weight Calibration for Outlier Weight Treatment, Nonresponse Adjustment, and Poststratification*. Proceedings of the American Statistical Association, Section on Survey Research Methods, pp 598-603.

Kish, L. (1992). *Weighting for Unequal* P_i. Journal of Official Statistics, 8, 183-200.