Reproducing Nonresponse Adjustments in Replicate Weights

Pedro J. Saavedra[1], R. Lee Harding[2]
[1]ICF Macro, 11875 Beltsville Dr, Calverton, MD 21230
[2] ICF Macro, 11875 Beltsville Dr, Calverton, MD 21230

Abstract
Creation of replicate weights for the jackknife procedure should reproduce the adjustment patterns of the main weighting. Often this is not possible because all the information is not available or the process is too cumbersome. When nonresponse adjustments using propensity scores are done, and there is no poststratification, there are four possible ways of adjusting replicate weights. Complete reproduction of the weighting procedure may not be necessary. One can build replicates with the initial sample and reproduce the propensity categories, accept the categories but recalibrate the adjustments, estimate the change in adjustment coefficients from the respondents only, or apply the same adjustments as were applied in the original sample. Using simulated data, estimates from the four approaches were compared under different conditions. While complete reproduction gave the based prediction of the standard deviation of the estimates over many samples, it was counterproductive when a nonresponse bias remained after the adjustments.

Key Words: Jackknife, Propensity Categories, Variance Estimation, Weight Adjustments

1. Introduction

There are many instances where nonresponse adjustments are applied to a sample, but there is no post-stratification. In such instances the adjustments are likely to add to the variance of the estimates and should therefore be included in the variance estimation procedure. Replicate methods such as the jackknife are well suited for estimations of this type, as the procedure can be applied to each replicate as one calculates the replicate weights.

However, to include fully the nonresponse adjustment procedures in the variance estimates, one should include the initial sample, not just the respondent sample. And there is an issue as to whether a recalculation of propensity scores or categories should be conducted separately for each replicate. This paper conducts a simulation to compare the variance estimates using four different approaches of including nonresponse adjustments in jackknife variance estimates. All four will be applied to the same adjustment methods to examine the degree to which the different ways of including the nonresponse adjustments when preparing the variance estimates affect the results.

2. Simulation procedure

2.1 Description of Data

For the purposes of simulation we desired a relatively large dataset with extensive demographic and analytic variables. The National Health and Nutrition Examination Survey (NHANES) is rich data. The 2005 – 2006 NHANES interview data contains over 10,000 respondents with complete demographic variables as well as numerous analytic variables. The NHANES data has the added bonus of being gathered in two parts. Respondents are interviewed and then asked to be examined at a later date. Not all respondents interviewed responded to the exam thus creating a natural nonrespondent category within the data.

The dependent variable was Body Mass Index (BMI). Since BMI was not obtained for minors, these were excluded from the sample. Using stepwise regression, BMI was then predicted for all members of the sample, with a number of demographic predictors. These included Hispanic origin, African American, sex, citizenship, age and the square of age. The equation used the 5,200 adult respondents, and a predicted value and a standard error of the prediction were obtained (R2=.057). Then the weights were divided by 1000 and rounded using a random number. Each respondent was reproduced by this number, creating a frame of over 200,000. It was then reduced to exactly 200,000.

Two variables were then created. The first used the equation BMISIM=x'+e1s where x' was the predicted value, s was the standard error of the predicted value and e1 was normally distributed random variable with a mean of 0 and a standard deviation of 1. The send variable was BMIALT= x'+10e2s where e2 was a different normally distributed variable with the same distribution. The idea was that BMIALT would be correlated with BMISIM moderately. Indeed, a correlation of r =.52 was found.
Now the lowest quintiles of values for BMIALT were designated nonrespondents. Thus the chances are than any random sample of the population would have about 20% nonrespondents. The simulated frame and the designation of nonrespondents were created only once. Then 1,000 random samples of 2,000 each were drawn from the simulated frame.

2.2 Method of Nonresponse Adjustment

Each of the 1,000 samples of 2,000 had approximately 400 nonrespondents. The demographic variables known to be correlated with BMISIM and with nonresponse were available. A regression equation was obtained for each sample, and the predicted values were divided into quartiles. An initial weight of 100 was simply the inverse of the probability of selection.

Now, let w_1 be the weight for the initial sample (100 in this case). Then for each propensity category, we can obtain $w_2 = (r+n)/r$ where r is the number of respondents among the initial sample in the propensity category and n is the number of nonrespondents in the initial sample in the propensity category. Now $w = w_1 w_2$ becomes the adjusted weight.

These adjusted weights turned out to be biased, as the mean BMISIM of the frame was 28.57 and the mean of the means of the 1,000 samples was 28.61 (t=38.42 p<.0001). While relatively small, this difference is important to consider in the evaluation of the results.

3. The Four Approaches

Four different methods of trying to capture the nonresponse adjustment were used. In each approach a delete-a-group jackknife was used. The sampled records were ordered by sequence number and the records were numbered 1 through 20, with the 21st record being assigned to group 1 again, and so forth. What differed was the treatment of the non-response adjustment.

3.1 The Simple Approach

Since the weights will be different for the different categories, one can simply do a jackknife from the response data sent only. If one uses the delete-a-group jackknife and there are m groups, the weights can simply be multiplied by m/(m-1) after deleting one group. This approach does not take into account where the weights came from, but will account for the weight discrepancy.

3.2 Adjusting within the Respondent Sample

The second approach also uses only the respondent sample, but recalculates the weights, assuming proportional deletion from the nonrespondent set. For each replicate the w1 is calculated as before. The r in the propensity category is calculated from the sample. Thus if in category 1 there were originally 150 respondents, and now there are 120, we let r=120m/(m-1). However, since we do not have access to the initial sample, we simply use the same n as before.

3.3 Using the Initial Sample

Here the jackknife is applied to the initial sample. The groups are formed at random, or following the systematic sampling order. The same propensity categories are used as in the total sample, but n and r are recalculated with the replicate sample. Since the numerator and denominator of the ratio would be multiplied by m/(m-1) this term can be dispensed with. This approach might seem to come closer to capturing the variance of the adjustments, except that the actual propensity equation might have been different with a different sample.

3.4 Replicating the Equation

In order to capture the equation's variability, this approach creates a new logistic regression equation for each replicate and uses quartiles for the replicate. The only part of the adjustment that the equation may not capture is the selection of the variables that are used to predict propensity to respond.

4. Examination of the Results

For each of multiple samples, four standard errors were calculates using the jackknife, one for each of the four methods. Two other standard errors were directly calculated from the simulations. One was the root mean square deviation of the thousand estimates from the population mean. The other was simply the standard deviation of the 1,000 estimates of the mean. The values were .053 for the former and .034 for the latter. If in fact the propensity categories do not capture nonresponse bias, it would seem that most of the variance estimates would not capture it either. In fact, consider the four methods with respect to the root mean square error of the estimates around the population mean.

Table 1: Estimates with Respect to RMS Error around the Population Mean

Method	*Mean Std. Error*	*Mean Deviation*	*Mean Absolute Deviation*	*Root Mean Square Deviation*
1	0.0367	-0.0165	0.0165	0.0175
2	0.0292	-0.0240	0.0240	0.0243
3	0.0280	-0.0252	0.0252	0.0256
4	0.0305	-0.0226	0.0226	0.0232

As can be seen all four estimates are an underestimate. Which means that the naïve method (#1), which yields the largest estimate is the one that comes closest. All the pairs of estimates between methods, by the way, are significant at the .001 level.

However, if we consider the RMS Deviations around the mean of the estimates, we get a different picture.

Table 2: Estimates with Respect to RMS Error around the Mean of the Estimates

Method	*Mean Std. Error*	*Mean Deviation*	*Mean Absolute Deviation*	*Root Mean Square Deviation*
1	0.0367	0.0029	0.0052	0.0066
2	0.0292	-0.0046	0.0052	0.0062
3	0.0280	-0.0058	0.0062	0.0073
4	0.0305	-0.0032	0.0050	0.0061

In this case the estimates are much closer, and the gold standard is between estimate 4 and estimate 1. Whether one uses the absolute deviations of the mean square deviations, it seems that estimate 3 is significantly worst that the others with 1, 2 and 4 being too close to make a determination (the differences are not statistically significant, save that 1 is worse than 4 at the .05 level).

The conclusion is that if we set as the gold standard the standard deviation of repeated samples (whether or not the estimates are biased) deriving new propensity categories for each replicate may yield the best results, but the marginal accuracy of the estimates is minimal. The naïve solution (using the weights as if they were a given) may be a good conservative estimate.

References

Kott, Phil S., *Using the Delete-a-Group Jackknife Variance Estimator in NASS Surveys*. U.S. Dept. Agr., National Agricultural Statistics Service (1997).