Discussion of "Weighting and prediction in sample surveys," by R. J. Little[1]

Andrew Gelman[2]

7 Feb 2009

Survey weights, like sausage and legislation, are designed and best appreciated by those who are placed a respectable distance from their manufacture. For those of us working inside the factory, vigorous discussion of methods is appreciated. I enjoyed Rod Little's review of the connections between modeling and survey weighting and have just a few comments.

I like Little's discussion of model-based shrinkage of post-stratum averages, which, as he notes, can be seen to correspond to shrinkage of weights. I would only add one thing to his formula at the end of his Example 3, which is that his regression model can include poststratum-level predictors; for example, if poststrata are indexed by sex, age, ethnicity, and education, the model could include indicators for each of these factors, and even two-way effects as necessary. This seems to be where he is leading in his Example 4.

I also found Little's discussion of probability proportional to size (pps) sampling very helpful; this is a problem that I have found difficult to attack using model-based methods. The spline model for the response given stratum size seems like a good way to go. My only comment here is that I have always associated pps sampling with two-stage cluster sampling, in which clusters are sampled pps and then a fixed-size sample is drawn from each cluster. In this case, the classical pps unit weights are all equal, and it is hard for me to believe that a model-based approach can improve much upon this, at least in settings in which the measures of size used in the sampling are not far from the actual sizes of the clusters.

As Little emphasizes, weights and other survey adjustment procedures are intended to correct for known differences between sample and population. I would rephrase his claim that "model-based statisticians cannot avoid weights," and instead say that statisticians cannot avoid *adjustment*, but this adjustment could take other forms, such as my personal favorite of model-based poststratification (Gelman and T. C. Little, 1997, Gelman, 2007).

Don Rubin once told me he would prefer to do all survey adjustment using multiple imputation; for example, in a survey of 1000 American adults, he would impute the missing responses for the other 250 million. I asked him if that was impractical, and he replied that the imputation could only realistically be performed conditional on information available on all 250 million; i.e. Census demographics, and thus the imputation would in fact be equivalent to fitting a regression

---

[2] Department of Statistics and Department of Political Science, Columbia University, gelman@stat.columbia.edu, http://www.stat.columbia.edu/~gelman

model of the response conditional on key demographic variables recorded in the survey and then summing over Census numbers to get national estimates. Depending on the method used to estimate the regression, it might be possible to approximate such an estimate as a weighted average over the sample (Little, 1993, Gelman, 2006) but it would be stretching it to call this a use of weights. In addition, under this approach, the approximate weights depend on the fitted model and thus on the outcome being modeled. Having a different weight for each question on the survey would seem to go beyond the usual conception of survey weighting.

Even in the design-based world, survey weights are not always based on selection probabilities. Consider the following poststratification example: A national survey of American adults is conducted and yields 600 female respondents and 400 males. The standard poststratified estimate is to take 0.52 times the average response for the women plus 0.48 times the average for the men, which corresponds to unit weights of 0.52/0.60 for each woman and 0.48/0.40 for each man. These are not inverse selection probabilities but rather are based on the known proportions of men and women in the sample and population. The weights are not even *estimated* inverse selection probabilities, a fact which we can see by noting that, even the actual selection probabilities were given to us, we would not use them: the poststratification weights are better. Which is perfectly consistent with the points Little makes in his article.

Additional references

Gelman, A. (2007). Struggles with survey weighting and regression modeling (with discussion). *Statistical Science.*

Gelman, A., and Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology* **23**, 127-135.