Statistical Science
2007, Vol. 22, No. 2, 171–174
DOI: 10.1214/088342307000000186
Main article DOI: 10.1214/088342306000000691

Comment: Struggles with Survey Weighting and Regression Modeling

Roderick J. Little

I appreciate the opportunity to comment on Andrew Gelman's interesting paper. As an admirer of Gelman's work, it is a pleasure to read his take on the topic of survey weighting, which I have always found fascinating. Since I support Gelman's general approach, I focus on reinforcing some points in the article and commenting on some of the modeling issues he raises.

As a student of statistics, I first encountered weights as the inverse of the residual variance for handling nonconstant variance in regression. I then had a course on sample surveys, where the weights were the inverse of the probability of selection. When these two sets of weights are different, which should be used? This question remained a mystery for many years, and only later did I come to appreciate that it reflects fundamental philosophical differences of design-based versus model-based survey inference.

The design-based approach treats the survey outcomes as fixed, with randomness arising from the distribution of sample selection. Sampling weights, defined as the inverse of the probability of selection, play a pivotal role in design-based inference in yielding estimates that are design unbiased or consistent. Similarly with poststratification, the weight is proportional to the ratio of population and sample counts in the poststrata, and as such involves the distribution of the sample counts rather than outcomes. If the "probability of selection" is replaced by the "probability of inclusion," then nonresponse weighting also enters the picture as the inverse of the estimated probability of response given selection.

The regression approach is model-based, and puts the emphasis on predicting values for nonsampled units in the population. Gelman uses the Bayesian paradigm to generate predictions, but to me the key issue is whether the objective is viewed as prediction. The Bayesian paradigm seems to me (and I think to Gelman) the most natural and compelling framework for prediction (Little, 2004, 2006), but in many situations one can get quite far with likelihood-based methods that do not explicitly add a prior distribution. In summary:

$$\text{design-based} = \text{weighting};$$

$$\text{model-based} = \text{prediction}.$$

This statement is an oversimplification. Design-based weights arise in the context of particular prediction models, so the approaches intersect. A simple example is the stratified mean for stratified samples, which arises as the prediction estimate for a regression on dummy variables for strata. More generally, Little (1991) provides an approximate Bayesian interpretation of design-weighted estimates of regression parameters. Prediction and weighting can be combined, and hybrid approaches are increasingly popular. In particular, Särndal, Swensson and Wretman (1992) take the prediction estimate from a model and then calibrate it by adding weighted sums of residuals, to yield protection against model misspecification. Robins and colleagues (Scharfstein, Rotnitzky and Robins, 1999; Bang and Robins, 2005) use the term "doubly-robust" to describe such estimators, and have popularized them in the general statistics literature; I would be interested in Gelman's views on this alternative approach. My own view is that robustness can be achieved within a pure prediction paradigm by judicious choice of model; see Firth and Bennett (1998), Little (2004) and Little and Zheng (2007).

Design weighting, as represented by the Horvitz–Thompson (HT) estimator and variants, has the virtue of simplicity, and by avoiding an explicit model it has an aura of robustness to model misspecification. It is the "granddaddy of doubly-robust estimators," since it is a prediction estimator for a model where the ratios of outcomes to selection probabilities are exchangeable, and it is consistent when either this model or the weights are correctly specified (Firth and Bennett, 1998). However, unthinking application of the HT estimator is dangerous, since inferences based on it can

Roderick J. Little is Professor, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, USA (e-mail: rlittle@umich.edu).