Journal of Official Statistics, Vol. 28, No. 3, 2012, pp. 363–365

Discussion

Michael P. Cohen[1]

Hearty thanks go to Rod Little for his interesting and thought-provoking article. I have heard Chief Scientist Little (to use his government title) speak many times and have always admired his principled approach to statistics and his engaging style of presentation. As the associate editor for this discussion article, I also thank Jean-François Beaumont, Philippe Brion, Alan Dorfman, Risto Lehtonen, and Paul Smith for their review work and discussions. I particularly note the great variety among them of insightful comments. I have the advantage of having read their discussions but also the responsibility to keep my discussion as distinct as possible.

I share Rod's belief in the value of unified approaches, but there is one distinction that may be worth maintaining. To quote Morris Hansen (1987, p. 180): "It is important to distinguish two distinct types of inferences based on sample survey and census results. The first is descriptive, that is the aim is to describe the characteristics of a specified finite population. For this case a complete census is sometimes available or can be taken. If the census covers the desired subjects and is complete and accurate, it would be sufficient. Ordinarily, a complete census is not feasible, especially for providing current information on many studies and topics, and a sample survey is used to provide estimates of what would be obtained by a complete census. The second type of inference is concerned with the causes that produced certain characteristics of the population. Such problems may be particularly important but inferences may be more difficult. The distinction between the two types of inference is sometimes not clearly recognized, especially since both may use the same data and the same or similar methodology."

The word *causes* in Hansen's quotation may be too strong in that, especially for exploratory work, associations are also important. I nevertheless think descriptive statistics have aspects that differ from analytical statistics. The most obvious is the need for the finite population correction (fpc). But the reason for the fpc is that descriptive statistics describe actual populations, not hypothetical ones.

A nice illustration of this distinction is the U.S. National Assessment of Educational Progress (NAEP) conducted by the National Center for Education Statistics. Because interest is in the actual school, an fpc is applied at the school level. For students, the interest is analytical (after all, the students who were assessed may not even be in the same grade by the time the results are announced), so no fpc at the student level is employed. See Kali et al. (2011) and Rizzo and Rust (2011) for full discussion.

[1]The Center for Survey Methods, American Institutes for Research, 1000 Thomas Jefferson Street NW, Washington, DC 20007-3835 U.S.A. Email: mcohen@air.org

I entered the U.S. government believing that bias is only important to the extent that its square is one of the components of mean squared error, along with variance. But my first supervisor, the late and I would say great Curtis Jacobs, soon convinced me that for *descriptive statistics* bias is paramount. The reason is that a policy of an agency of using estimates with little or no bias for their descriptive statistics leads to an overall pattern of estimates with no tendency to favor one side over another in any controversy.

Also in terms of agency policy, the use of design-based confidence intervals for descriptive statistics seems appropriate. A policy of using 90% confidence intervals (the standard at the U.S. Census Bureau, I believe) should mean that the intervals computed by the agency cover the true value on average 90% of the time. For the Bayesian confidence (or creditable) intervals, the estimate is in the interval for 90% of the draws of the parameters from the posterior distribution, not the criterion of interest. Rod proposes calibrating the creditable intervals to have good frequentist properties, but his purpose is to ensure robustness to model failure. An admitted weakness of the design-based approach is that one usually computes estimated variances and invokes a normality assumption to calculate confidence intervals. For large samples approximate normality can be justified by an appropriate version of the central limit theorem, but what about smaller samples? This is an area where more research is needed.

Does small area estimation contradict what I have been saying about descriptive statistics? No, small area estimation has to be analytical. In most cases, small area estimation is resorted to when one cannot get descriptive statistics that satisfy the usual statistical standards for descriptive statistics. (I was surprised by Rod's American Community Survey example at the beginning of Section 4.4. In my dealings with the U.S. Census Bureau, they always insisted on a coefficient of variation (CV) no greater than, if I recall correctly, 30%.) The discontinuity that Rod mentions in going from the small area estimate to the direct estimate occurs just at the point where the direct estimate begins to meet statistical standards for descriptive statistics.

One final point on descriptive statistics: yes, one has to adopt some form of a missing at random assumption (or other non-design based assumption) in dealing with nonresponse. We make every effort to keep the response rate high to minimize the impact of this assumption, but the need for it is just a fact of life.

Turning to analytical statistics, I see great potential in the calibrated Bayes approach that Rod advocates. More specifically, my preference is for *nonparametric* Bayesian methods (NBM). The appealing property of NBM is that the posterior can approximate *any* distribution as the sample size grows larger, so the Bayesian modeling is unrestrictive, at least in large samples. The influence of the prior can be made small, even in small samples. Although much progress has been made in NBM recently, it is far from ready to become "an inferential paradigm for official statistics." But it may someday. The work of Aitkin (2008) is very promising, especially for categorical data. The Polya posterior, also mentioned by Beaumont, also has great potential, especially when its extension to unequal probability sampling is fully developed. For the Polya posterior, in addition to the works cited by Beaumont, let me cite the book by Ghosh and Meeden (1997); there is much recent work by Glen Meeden and his colleagues. Of course, there is the work by Zheng and Little (2003 and the references in Rod's article) on splines. At the 2011 Joint Statistical Meetings, there was a fascinating talk by Sahar Z. Zangeneh on "Bayesian Nonparametric

Estimation of Finite Population Quantities in Absence of Design Information on Nonsampled Units" (unpublished). Her work is joint with Bob Keener and, yes, Rod Little.

In closing, the road may be rocky in the journey toward a single unified approach to official statistics, even for analytical estimates. But there are many clear paths toward improvements. For example, for many surveys government agencies do not do single imputation, much less multiple imputation or one of its competitors. Data analysts have to resort to deletion methods to analyze the data. For many statistics such as price indexes, variances may not be estimated at all on a routine basis. As a survey statistician retired from the government, I want to express my gratitude to Rod for coming into the government from his prestigious academic position at the University of Michigan and for devoting his time and energy to his important work at the U.S. Census Bureau.

Finally, I want to dedicate this discussion to the memory of David Binder (1949–2012), a pioneer in the application of nonparametric Bayesian methods to surveys, a great statistician, and a wonderful guy.

1. References

Aitkin, M. (2008). Applications of the Bayesian Bootstrap in Finite Population Inference. Journal of Official Statistics, 24, 21–51.

Ghosh, M. and Meeden, G. (1997). Bayesian Methods for Finite Population Sampling. London: Chapman and Hall.

Hansen, M.H. (1987). Some History and Reminiscences on Survey Sampling. Statistical Science, 2, 180–190.

Kali, J., Burke, J., Hicks, L., Rizzo, L., and Rust, K. (2011). Incorporating a First-Stage Finite Population Correction (FPC) in Variance Estimation for a Two-Stage Design in the National Assessment of Educational Progress (NAEP). Proceedings of the American Statistical Association, Section on Survey Research Methods, 2576–2583.

Rizzo, L. and Rust, K. (2011). Finite Population Correction (FPC) for NAEP Variance Estimation. Proceedings of the American Statistical Association, Section on Survey Research Methods, 2501–2515.

Zheng, H. and Little, R.J. (2003). Penalized Spline Model-Based Estimation of the Finite Population Total from Probability-Proportional-To-Size Samples. Journal of Official Statistics, 19, 99–117.

Received July 2012