DISCUSSION

Robert E. Fay, U.S. Bureau of the Census

Through a bit of rearranging, the papers in this session group themselves into three pairs. All six concern complex samples. Two papers discuss statistical tests of models for categorical data, two describe or illustrate computer software for analysis of data, and two focus on possible applications of components of variance models. The discussion will follow this grouping.

The paper by Jeffrey R. Wilson, "A Simulated Comparison of Chi-Square Tests for Comparing Vectors of Proportions for Several Cluster Samples," considers three approaches to testing the equality of vectors of proportions when the data are drawn from a clustered sample:

i. Wald tests, defined in the paper to include the familiar Pearson chi-square as well as the version based upon the estimated sample covariance matrix;
ii. Model-based approaches, including work by Brier and by Koelher and Wilson;
iii. Other design-based approaches, using "partial information", such as the delta-bar adjustment to the Pearson chi-square test, as suggested by Rao and Scott, $X^2/\delta_.$.

Both the Type I behavior and the powers of these tests are compared in the simulation. I would have been interested in the consideration of at least one of two additional alternatives: the Satterwaite correction to chi-square based upon the estimated covariance matrix, as proposed by Rao and Scott (1981, 1984), and the jackknifed chi-square (Fay 1985). These tests exhibited the best overall performance in the study of Thomas and Rao (1985) and are based only upon the sample design instead of a specific model. (The jackknifed test would not have been competitive for the situation of 10 clusters considered in part of the simulation, but could have been expected to performed well at 25 and 50 clusters.)

Given the current availability of effective design-based tests, one might ask the question "Why use models at all for this testing problem?" One answer is that the process of modeling frequently provides additional information or insights about the population. A second answer is that general experience would support a presumption that under some situations, such as when the number of clusters is small, the model-based solutions might prove more effective than design-based tests requiring somewhat more data. Before using a model-based test, however, one might additionaly want to know the extent to which the performance of the model-based test might be degraded by departures from the specific model.

The paper defines:

p_{jk} = the proportion in the kth sample cluster, jth stratum,
π_j = the proportion in the jth stratum.
N_{jk} = the number in the kth sample cluster in the jth stratum.

An implicit consequence of the assumptions made in the paper is that $E(p_{jk}|N_{jk})$ is equal to πj. Yet, examples of where the proportion might depend upon cluster size come readily to mind; for example, households frequently constitute clusters in demographic surveys, and the average characteristics of persons within households almost always varies by household size. The models in the paper do not accommodate this source of variation, while the design-based methods do.

This work on modeling clustering effects is welcome. At the same time, further work appears necessary to reach a more complete conclusion with respect to the relative merits of these new proposals compared to design-based alternatives.

In their paper, "Categorical Data Analysis for Complex Surveys," A. C. Singh and S. Kumar propose a modification to the Wald test. The Wald test represented the first general design-based solution to testing categorical data models for complex samples. As noted by these authors and others, the Wald test often becomes unstable in these applications. Later design-based alternatives, such as $X^2/\delta_.$ and the jackknifed test, adopted different strategies to avoid this loss of stability. Singh and Kumar return to the Wald test and repair the instability directly. This approach appears quite promising. Aspects of the specific form seem ad hoc, however, and the authors should consider experimenting with their recipe further. For example, following in the footsteps of Rao and Scott, one possibility would be to base the modification on the eigenvalues of P^-V, where P^- is a generalized inverse of the estimated covariance matrix for multinomial sampling. Should further evaluation prove as favorable, one may hope that analysts who now prefer the Wald test may adopt some form of modification as a standard.

David Morganstein, Adam Chu, Leyla Mohadjer, and Mike Rhoads describe three related software products in their paper, "Estimation and Analysis of Survey Data Using SAS Procedures WESVAR, NASSREG, and NASSLOG." WESVAR computes variances for simple statistics, NASSREG for linear regression, and NASSLOG for logistic regression. All three compute variances based on BRR, balanced repeated replication, implemented through assignment of replicate weights to each record. Since replicate weights may be used to represent other replication methods, such as the jackknife, as well as more complex methods (e.g., Dippo, Fay, and Morganstein 1984, Fay 1984), I'd like to encourage the authors to adapt this more general perspective. Certainly, their contribution will be welcomed as a useful tool for the analysis of survey data.

The paper "An Application of Logistic Regression Methods to Survey Data: Predicting High Cost Users of Medical Care," by Lisa LaVange, Vincent Iannacchione and Steven Garfinkel also addresses the question of logistic regression for complex samples. The paper discusses the estimation of standard errors through Taylor series methods and describes computer software, again based upon SAS, to implement this method. The analysis in their paper provides a helpful example of the use of these methods. Again, the existence of RTILOGIT will undoubtedly be good news to those who have not yet heard of this software.

The remaining two papers, "The Analysis of Survey Data Using Stochastic Regression Coefficients with Application To NHANES Data," by Danny Pfefferman and Lisa LaVange, and "Complex Sample Design for Estimating Regression Parameters," by Thomas J. Tomberlin, both discuss the application of components of variance models to sample surveys. The two papers complement each other; the first provides a thorough discussion of the use of such models in analysis, while the second employs these models to formulate the problem of survey design for purposes of analysis with linear regression.

The paper of Pfeffermann and LaVange contains a careful discussion comparing design-based to model-based inference. In general, I highly recommend their paper for its presentation of this topic. I would only add the comment that the estimation of variance components can be addressed by different methods, some of which are far more complicated than others. Although simplicity is (usually) a virtue, my own experience is that the increase in efficiency from the more complicated methods, for example, maximum likelihood, may be worth the effort when the final result is important. A two-phase strategy may be of use here, namely to use simpler methods for exploratory efforts and to follow with maximum likelihood or similar efficient alternatives for the final product.

Tomberlin makes an important contribution to what is a fairly limited literature: the design of complex sample surveys when modeling is the primary objective. Since the primary purpose of the paper is design, I believe that the author focused on simple estimators to make the problem tractable. In general, I would favor the methods presented by Pfeffermann and LaVange at the analysis phase. I will also note that the author restricted the choice of alternatives to proportionate sampling, whereas a design-based perspective allows the consideration of more aggressive designs that could disproportionately sample Y on the basis of X. Nonetheless, his paper makes a significant contribution to the problem of design.

References

Dippo, C., Morganstein, D. and Fay, R. (1984), "Computing Variances from Complex Samples with Replicate Weights," Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., 489-494.

Fay, R. (1984), "Some Properties of Estimates of Variance Based on Replication Methods," Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., 495-500.

______ (1985), "A Jackknifed Chi-Square Test for Complex Samples," Journal of the American Statistical Association, 80, 148-157.

Rao, J.N.K., and Scott, A. J. (1981), "The Analysis of Categorical Data From Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables," Journal of the American Statistical Association, 76, 221-230.

______ (1984), "On Chi-Squared Tests for Multiway Contingency Tables With Cell Proportions Estimated From Survey Data," Annals of Statistics, 12, 46-60.

Thomas, D.R., and Rao, J.N.K. (1985), "On the Power of Some Goodness-of-Fit Tests Under Cluster Sampling," Proceedings of the Section on Survey Research Methods, American Statistical Association, 291-296.