DISCUSSION

J.N.K. Rao, Carleton University

The consequences of ignoring the effect of survey design, and analysing survey data by standard methods for simple random samples, are now well-understood. For instance, the standard chi-squared and likelihood ratio tests for categorical data analysis can yield unacceptably large type I error rates under cluster sampling. A number of alternative methods that take account of the survey design, therefore, have been proposed. The three papers in this session make excellent contributions to these methods. The papers by A.J. Scott and Georgia Roberts investigate respectively logistic regression models and a general class of polytomous response models with cross-classified categorical data from complex sample surveys, while the paper by C.A. Francisco and W.A. Fuller develops asymptotic confidence intervals and standard errors for quantiles and interquantile range constructed from stratified cluster samples.

If the full estimated covariance matrix of cell estimates or detailed cluster level data were available, we have at least three options: (1) the well-known weighted least squares method based on the Wald Statistic (Koch, Freeman and Freeman, 1975), (2) the Satterthwaite adjusted tests (Rao and Scott, 1984), (3) the jackknife chisquared test statistics based on a replication strategy (Fay, 1985). The investigations by Fay (1985) and Thomas and Rao (1984), however, have shown that the Wald tests can behave poorly as the number of cells in the table increases.

Alternative methods designed for use with published tables, containing information on cell design effects and/or design effects for marginal totals, have also been proposed recently. Scott's paper provides a simple, effective method for logistic regression models with binary response data. The sample sizes, n_i, in the cells (corresponding to factor combinations) are simply changed to "effective sample sizes" $\tilde{n}_i = n_i/\hat{D}_i$ and the standard methods are then applied to the data $(\tilde{n}_i, \hat{p}_i)$, where $\hat{p}_i$ is the survey estimate of i-th cell response proportion and $\hat{D}_i$ is the estimated design effect of $\hat{p}_i$. If the estimated covariance matrix $\hat{V}_p$ of the $\hat{p}_i$'s were diagonal, as in the case of teratological experiments involving animal litters, then Scott's method is asymptotically correct and provides an alternative to weighted least squares Wald statistic based on $\hat{V}_p$. In the context of sample survey data, however, $\hat{V}_p$ is often not diagonal, but Scott's empirical study based on the Canada Health Survey data showed that the method might work quite well in practice. For large tables, Scott's method might not work as well as those based on average design effects (Rao and Scott, 1987) due to instability in some of the $\tilde{n}_i$ caused by imprecise estimation of cell design effects or cell variances.

The generalized response models considered by Roberts cover many useful models, including binary response models, ordered polytomous response models of McCullagh and multinomial response models of Haberman. She makes a systematic study of the effect of design on standard chisquared tests and develops both first order and second order (Satterthwaite) adjustments to chisquared tests, along the lines of Roberts, Rao and Kumar (1987) for logistic regression models. The adjustments based on the average design effect, $D_.$, or the average of eigenvalues not depending on any hypothesis, are particularly useful with published tables since it should be feasible to report the eigenvalues or the design effects along with the tables of estimated proportions. These simple adjustments, however, could lead to overly conservative tests with small tables, especially for nested hypotheses, as can be seen from the Canada Health Survey example (3 × 3 table).

Woodruff's method of constructing large-sample confidence intervals for medians and other quantiles is well-known. This method can be used for general survey designs since it requires only the standard variance estimators for means or proportions. Francisco and Fuller provide asymptotic justification of Woodruff's method for single-stage, stratified cluster sampling, by extending the well-known Bahadur's representation of sample quantiles for simple random samples. Their key theorem 3 establishes the asymptotic multivariate normality of k selected sample quantiles. This result can be used to obtain asymptotic standard errors for the sample median (or any other sample quantile) and the sample interquartile range. These standard errors also use only the standard variance estimators for means or proportions, as in the case of Woodruff's confidence interval. In fact, the standard error for a quantile is proportional to the length of Woodruff's interval. A limitation of these standard errors, however, is the dependence on the confidence coefficient, $1-\alpha$. Perhaps an improved standard error can be obtained through calibration of the standard errors for selected values of α. Some preliminary work of mine (jointly with C.F.J. Wu) indicates that the balanced half-samples method or suitable modifications of the bootstrap might yield consistent standard errors for quantiles. These standard errors do not depend on α since they are not derived from any confidence interval.

It is gratifying that Francisco and Fuller have incorporated the proposed procedures into the PC CARP survey data analysis computer program.

REFERENCES

1. Fay, R.E. (1985), "A Jackknifed Chi-Squared Test for Complex Samples," *Journal of the American Statistical Association*, 80, 148-157.

2. Koch, G.G., Freeman, D.H., Jr., and Freeman, J.L. (1975), "Strategies in the Multivariate Analysis of Data from Complex Surveys," *International Statistical Review*, 43, 59-78.

3. Rao, J.N.K. and Scott, A.J. (1984), "On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data," *Annals of Statistics*, 12, 46-60.

4. Rao, J.N.K. and Scott, A.J. (1987), "On Simple Adjustments to Chi-Squared Tests with Sample Survey Data," *Annals of Statistics*, 15, in Press.

5. Roberts, G.R., Rao, J.N.K., Kumar, S. (1987), "Logistic Regression Analysis of Sample Survey Data," *Biometrika,* 74, in Press.

6. Thomas, D.R. and Rao, J.N.K. (1984), "A Monte Carlo Study of Exact Levels of Goodness-of-Fit Statistics Under Cluster Sampling," *Proceedings of the Section on Survey Research Methods,* American Statistical Association, 207-211.