DISCUSSION

David D. Chapman, U.S. Bureau of the Census

Variance Estimation For Sample Surveys

The four papers presented in this session are diverse. They touch on various aspects of applied survey research. The Proctor paper discusses variance estimation from a Design of Experiments point-of-view. The Schindler and Kulpinski paper discusses the estimation of variance components within the context of Balance Repeated Replication. The Shimizu paper discusses the adequacy of a variance approximation for an estimator of a percent in a complex sample survey. The Ghangurde paper deals with models for estimating sampling error. Three papers--Proctor, Shimizu, and Ghangurde--deal with estimating variances. Two papers--Proctor and Schindler and Kulpinski--deal with variance component estimation.

These papers are excellent examples of applied survey research: research aimed at producing either new information from a survey or understanding the quality of current information.

Proctor--Projecting Variances in Order to Compare Methods of Estimating Weed Abundances.

In a short space, a complex experiment is described, variance components estimated, the H. Fairfield Smith's empiracal law is used to relate plot size to plot variance, and a cost analysis made of alternative measurement methods. The paper addresses a variety of statistical questions. An entire session could be spent discussing any one. These questions include: (1) what to do about negative estimates of variance components, (2) iterative estimation of variance components, and (3) the use of cost functions in design of experiments. The paper is an excellent example of the analysis of a designed experiment.

The paper does have some problems. The sample design is never fully described. In the analysis and in the ANOVA tables, the author uses "SAS" program statements rather than standard mathematical notation for models and expected mean squares. Certain models are described in terms of "SAS" procedure commands.

Least squares is used to estimate variance components. While I am not familiar with this procedure, it apears to be used when there are more estimates of mean squares than variance components to be estimated. In many standard designs the number of mean squares from the ANOVA exactly equals the number of variance components estimated. It is not clear why you would have more mean squares than variance components.

Schindler and Kulpinski--Components of Variance by Replicated BRR (Balanced Repeated Replication).

This paper presents a simple technique for estimating variance components while using Balance Repeated Replication to estimate variances. The general principal is easily applicable to other surveys or variance estimation procedures. The paper presents a brief, easily understood, review of Balance Repeated Replication as a variance estimation procedure. Variance component estimation has its basis in the fact that changing the number of sample units in the last stage or in any stage of a sample design will change the variance.

The authors make the statement that:

> "If we now randomly take a fraction, such as one-half or three quarter and recalculate the variance using BRR for this subset of data, a large estimate of the variance will be obtained."

The expected value will be larger; but a single estimated variance may be either larger or smaller. If it is smaller, then negative estimates of variance components will result. It would be interesting to know how the authors would estimate confidence intervals on the variance components. In many variance components estimation procedures, the mean squares are independent under assumptions of normality. In this design, the mean squares or variance are not independent. It would be useful to know how this lack of statistical independence affects the estimation.

Ghangurde--Models for Determination of Sampling Errors.

This paper presents examples of the use of linear and nonlinear models to describe the relationship between an estimate and it's reliability. The paper is a good example of the logic of developing a model to relate the coefficient of variance and the estimate. Design effects are utilized to identify characteristics with similar reliability.

Two error models are considered:

$$CV = a \times {}^{b}\varepsilon \qquad (1)$$

$$CV = a' \times {}^{b'} + \varepsilon' \qquad (2)$$

The error for one model is assumed to be log normally distributed; the error for the other model is assumed to be approximated normal distributed. It is not realistic to make both assumptions at one time.

Shimizu--Accuracy of Difference Method for Approximation Sampling Errors for Percent Estimates from a Complex Sample.

This paper demonstrates that for some surveys variance approximation is satisfactory. The paper examines the adequacy of the difference of the relative variance of the numerator and denominator as approximation for relative variance of a percent--the numerator divided by the denominator. The paper examines the absolute and relative error of the estimate for a variety of estimates from the National Reporting System for Family Planning Services.

It is not obvious from the paper how the approximation is determined. I would suggest that in the final paper that the derivation be provided in an appendix or referenced.