DISCUSSION

Judy A. Bean, University of Iowa

First, I wish to thank the authors for presenting a set of interesting papers dealing with a range of problems that survey statisticians encounter. In order to allow time for discussion from the audience and on account of my own research interests, I intend to restrict myself to making specific comments on the three papers concerning the use of the balanced half-sample technique.

I am sure everyone here is aware of the increasing use of surveys to collect data. As more research is accomplished with surveys, fundamental and philosophical issues are being raised regarding the validity of inference. As the first step in the process of making statements about population parameters, variances of the estimates of the parameters are necessary. One such technique that has been employed to estimate variances is the balanced half-sample method. Today, we have heard the results of three investigations concerning the properties of the estimates of the procedure. For brevity I will refer to the balanced half-sample technique as BHS.

I. Estimating the Variance of the Slope of a Linear Regression in a Stratified Random Sample with the Balanced Half-Sample Technique

This paper provides us with more evidence on the behavior of the BHS technique for estimating the variance of a slope along with exposing us to a new form of the BHS method and to three ways of estimating the slope itself. Because of its flexibility the BHS method has been employed by survey statisticians, for example, Leslie Kish of the Survey Research Center at the University of Michigan, to estimate the variance of the slope but just what sample size is needed to yield an adequate stable estimate is not known. These results for the particular sample design used are the beginnings of guidelines needed by practicing statisticians. Naturally, more work in the area is needed.

As far as the comparison between the two forms of the BHS estimator, it would have been interesting to have included the estimator which is an average of the estimate obtained from the half-samples and the estimate from the complement half-samples since other investigations have shown this average to be better than the estimate obtained from the half-samples only. However, I suspect that in this situation the complement estimate would almost be identical to the usual half-sample estimate. Also, I think it may have been helpful to the reader if the expected value of this different form of the BHS method for the linear case would have been given. Unfortunately, as often happens, the results of the sampling experiments do not give a definite answer to the question of which form should be used. I wonder, for the "full-matrix" case, if the conclusion can be made that this different form will yield the same or smaller variances and mean square errors than the usual BHS estimator.

II. The Behavior of Balanced Half-Sample Variance Estimates for Linear and Combined Ratio Estimates When Strata Are Paired to Form Pseudo-Strata

When designing sample surveys, practicing statisticians often wish to select one primary unit per stratum in order to take full advantage of possible stratification gains. Thus, one does not have a satisfactory method for estimating variance from the sample itself; on the other hand, selecting two or more units from a stratum may obliterate potential gains in stratification. I am therefore delighted to see a study in which the problem is approached both mathematically and empirically.

As one would expect, the simulation results indicate that, when the method of collapsed strata is used, the BHS estimator of the variance of a ratio estimator is biased with the magnitude of the bias depending upon the formation of pairs of strata. The results are useful in establishing a direction for more research effort. The next step would be to examine the magnitude of the bias as the number of strata increases.

The other aspect of the problem discussed is whether or not for a large scale survey actually using both the method of collapsed strata and the balanced half-sample technique, different schemes of pairing strata has a practical effect. For this limited case, the answer was no. However, further study of the problem needs to be done. George Schnack of the National Center for Health Statistics and I have recently finished a feasibility study of the application of the BHS method for estimating variance components [Bean and Schnack (1977)]. As a substudy of that investigation, we noted that the formulation of the pairs of strata does affect the estimates of variance components for the Health Interview Survey. Thus, collapsing strata may seriously influence the estimates of components of variance necessary for designing purposes.

III. Evaluation of the Balanced Half-Sample Estimates for Linear and Combined Ratio Estimates for Non-Normally Distributed Populations

This paper correctly points out that the BHS technique is used to estimate variances from sample variables that are known to have non-normal distributions. Prior to this work, no study of the BHS estimator using Monte Carlo sampling from specified non-normal distributions has been performed. Thus, Dr. Hislop's work will add to the growing body of knowledge about the behavior of BHS estimators.

The results for the linear case together with the findings of other investigations have shown conclusively, I think, that when the estimator of the population parameter is linear, regardless of the underlying distribution of the variable, BHS estimates are satisfactory.

I was glad to learn that when the sample size is at least ten per stratum (so that the total sample size is at least 30) the results for the combined ratio estimate support the use of the balanced half-sample technique. The fact the estimates are extremely biased when the sample size is only two per stratum for three strata

does not alarm me since in practice for this situation one would not use the method. However, because of the magnitude of bias, a survey statistician should be cautious in using the technique when studying subdomains having small sizes as the paper indicates.

It would have been helpful if specific examples of variables having distributions studied here were given. Also, it is important to know if the numerator and denominator of the ratio for the various combinations of distributions are independent.

References

Bean, Judy A. And Schnack, George A., "A Feasibility Study of Estimating Variance Components Using the Balanced Repeated Replication Method." University of Iowa (Internal Memorandum).