DISCUSSION

Seymour Sudman, University of Illinois at Urbana-Champaign

In contributed paper session, it is sometimes difficult to find a common theme. Here either by luck or through the wisdom of the organizer of the session, five of the six papers share a common theme-finding optimum sampling designs for stratified samples. The contribution of these papers is primarily empirical and methodological, rather than theoretical. Two of the four papers also describe the use of mini-computers for this purpose.

The paper that I found most unusual was by our Canadian neighbors Chanut, Pellettier and Begin who, I hope, will forgive me for my lack of knowledge about oceanographic methods. Their paper, as you have heard deals with optimum sampling of various depths of a water column. Thus they are going back to the original use of the word stratum which was a layer of the earth and by analogy of the sea or air. The major new development of the authors is a program to be used with an on-board mini-computer to determine optimum strata boundaries as well as allocation within the strata.

Because of my lack of oceanographic background, it was not totally clear why such instant evaluation was required. If, in given locations, the characteristics of the water are fairly similar over time, then pre-stratification, using the results of the last survey would be possible. If the water characteristics are changing very rapidly, pre-stratification is not very satisfactory, but it is not then clear what use one can make of the data. Perhaps the authors could explain.

In the current study, the authors found no major differences in efficiency between using a fixed and an optimum allocation. One wonders if there would ever be situations where substantial improvement would be found over fixed or pre-stratified samples. If yes, what would characterize such situations. If not, then is there really a need for instant stratification?

It is almost universally found in examining the effects of optimum stratification that there is a broad range in which small changes from optimality hardly affect the variances. It is difficult to justify costly additional procedures once one is already near optimum. To summarize, this is a very interesting paper. I would have found it helpful if there were more discussion of basic issues of stratified sampling of water. This could have made it clearer what the special benefits were of instant stratification.

The paper by McCarthy and Clickner clearly has frequent applications. Given skewed populations, very typical in studies of organizations of all kinds, what is the optimum cutoff point for certainty sampling if one has a continuous distribution. The example used in the paper is of utility users. An iterative procedure is used until variances are minimized.

As the authors point out, relatively few iterations are required because variances are very flat in the broad area around optimality. Thus, the method is fast and efficient. As soon as one sees an increase in variances, one stops.

One could not use the method where, as in the example of total assets and as is often as the case, exact values are not available, but only a lower bound such as strata of households with incomes over $50,000 or firms with over 1,000 employees. Then, all members of the group must be sampled or some fraction, but no disproportionate sampling within stratum is possible. Thus, stratifying on a continuous variable theoretically always gives one more power than stratifying on grouped variables. In practice, it may make little difference, especially if there are measurement errors in the items used for classification.

The paper by Bethel provides an algorithm for optimum allocation when several criteria must be met and costs must be minimized. The program seems to be highly efficient taking less than 30 seconds of CPU time. Basically, the method is analagous to solving a convex linear programming problem.

I wonder if this method is adaptable to more general situations that seem to be common in the real world. A typical situation is that the resources available are fixed, unfortunately at a level lower than the minimum cost to satisfy all the criteria. While one still wishes to satisfy them, they are not all of equal value. That is, one has a regret function that one wishes to minimize relative to a fixed cost. Intuitively, one hopes that the methods described above could be adapted to the more difficult situation.

What are real world examples of uses of this method? Is the NAMCS paper by Tompkins and Shimizu such an example? That is, it is clear that the method proposed here is efficient. Are there problems to which it can be put?

It is not obvious from the title but the Tomkins-Shimizu paper is also about determining optimum stratification for the National Ambulatory Medical Care Survey (NAMCS). The decision was made to stratify by 14 physician specialties and for the six most common specialties by metropolitan status. If I understand this properly, for each of these strata precision was fixed at four levels (.3, .15, .1 and .05) and costs were minimized. Sounds very much like the Bethel paper, doesn't it?

The maximum samples required were found. Some of these results were fixed. It is still not clear exactly how this fixing occurred. It is also not obvious whether data will or will not be reported separately by metropolitan and not metropolitan for the six most common specialties. I have also heard that there are large regional differences in types of ambulatory care that patients receive. Is this taken into account in planning for this study. Finally, I really wonder if the decision was based on minimizing costs or whether there was not a cost constraint and decisions were made of acceptable variance levels based on these cost constraints.

Kaufman's paper describes and estimates generalized variances using replication and regression techniques to approximate variances as a function of sample size and other variables. These generalized variances are then used for optimum allocation as we have heard from the earlier papers.

The draft that I had was difficult for someone outside the area to follow. This was because there were a lot of initials used that were not explained. For example what is MOG. Also some terms are not explained. What is a quote. The methods, however, have been used for a broad range of applications where it is simply not feasible to present or use individual variance estimators. Readers in other fields might well be interested in what is being said in this paper, although in its present form it is really more an internal working paper.

The paper by Inglis, Groves and Heeringa relates to efficient methods of finding Black households--and by extension, other geographically clustered groups--using telephone screening procedures. This is a problem of growing importance since screening households may consume a major part of the survey resources. The empirical findings from Michigan are very encouraging in suggesting that there is little if any loss in efficiency in using banks of 200, or even 300 or 400 numbers to find enough cases of special rare populations, at least Black households.

A few comments for those why may not know these methods too well. The saving of such procedures is in quickly eliminating zero clusters. The need for large banks of numbers is for areas where the cluster has a small, but non-zero fraction of households who are in the required population. This is common and the SRC results address a real issue. One way of improving the efficiency of these procedures is to use the screening results for more than one study. Thus, in an election study with several waves of different households, the screening could be done only once and used for the subsequent studies. Of course, at some point in time, screening results become out-dated.

Any clustering method increases variance, but it is very encouraging to see that cluster effects are moderate relative to the major cost reduction. It would be nice to see some more detailed discussion of actual costs if these were available. Those of you interested in this topic might want to look at a paper of mine in the February 1985 Journal of Marketing Research which discusses optimum designs for screening of special populations.