Comments on Superpopulations and Superpopulation Models

c00ed64.doc (8/21/00 version), Voss and Letters

Comments by John Buonacorssi 9/7/00
Response by Ed Stanek 9/7/00

1. (P.4) I think Cochran's comment that "further, it is far removed from reality to regard the population as a fixed batch of known numbers" is only applicable in some settings. Quality control sampling from an incoming shipment, is clearly a case where his comment is false.

Response (EJS): Is the idea behind this remark that the population size may be unknown? It seems like two ideas are involved here. Consider quality control sampling where shipments are always coming in. Then the population size is unknown and changing. We can 'fix' the size (by focusing on shipments that have been received), and then the population size is potentially known. This is like the 'conventional' definition of a population referred to by Cochran. Alternatively, we can consider the past shipments to be a sample from an infinite 'superpopulation' of shipments, and let this sample size increase (as new shipments are received). The comment of Cochran above would apply to this idea. This idea hasn't been encorporated in the description of superpopulations, and should be added.

2. (p.8) In "Questions that arise with the infinite population concept". In a nutshell, it seems what you are saying is that the troubling feature is the question of how the population relates to the superpopulation; i.e., is there a plausible model relating the two. But, the assumption that the population "represents" the superpopulation is only relevant if one is trying to make inferences about the superpopulation (i.e., Case 3 in Table 1).

Response (EJS): Consider the discussion of farm yield (page 12-14). There is a plausible model between farm acerage and total yield (which can be captured in a finite population model, or a superpopulation model). The focus may be on yield in the finite population. It does seem to me that unless the population 'represents' the superpopulation, there is a problem inferring that the superpopulation model holds in the population, even if the finite population is the target of inference.

(p. 13) Under "Superpopulation Models."

3. There is suddenly a distinction being made between superpopulation models and superpopulations. This doesn't seem to be dwelt on earlier in the paper and the distinction should be stated more clearly here.

Response (EJS): Good point. In the next version of this document, this distinction is made more clear (I hope).

4. Page 16. I'm still working to understand in what sense it is "not possible to project this vector back to the population ..." I'm rereading parts of ed63.

Response (EJS): I'm currently working on c00ed63.doc. A simpler way to see what I mean

is to look at c00ed27.doc as a background document, but more specifically to look at c00ed48.doc (which I have attached).

5. **In addition to what is in Julio's comment 1.

In my view another main use of superpopulation models beyond the two he gives is to provide a different framework in which to evaluate estimators (or other inferential techniques). This is closely related to his comment to "... explain the data generation process ..." but the main use of the process is to use both stages (the superpopulation model and the sampling) in evaluation of a method (e.g., the notion of model unbiasedness instead of just unbiasedness based on sampling). From a Bayesian perspective, the population could be truly static, but the Bayesian would use a superpopulation model.

Response (EJS): There is a lot in this comment, especially when combined with Julio's comment that I would like to understand clearly.

One of the uses of superpopulations cited by Cassel, Sarndal and Wretman (1977), and the focus of Cochran (1946) was to use superpopulations to evaluate estimators (Cochran was comparing systematic sampling and stratified sampling). In the newer draft of c00ed64.doc, I included a section entitled "Superpopulations: a Framework for Comparing Statistics from a Population" which discusses this. This does provide a framework for evaluating estimators. I hope this addresses this point.

I'll try to re-state the second part of the comment about the process. It seems that the idea is very similar to that of Hartley and Sielken (1975).

1. You assume that there exists a possibly infinite superpopulation.
2. You assume a model holds in the superpopulation.
3. You assume some connection between the population and the superpopulation (such as the population is a simple random sample of the superpopulation).

The 'population' can be represented as a set of random variables in the sense that it is potentially observable. Once realized, the population is a fixed set of numbers with labels. It is in this sense that the population is static. I think a part of the issue (and controversy) in sampling has to do with distinguishing between the labels for the units, and the values for the units. Are the labels fixed (even in the superpopulation)? For different realizations of the population from the superpopulation, do we get a different set of labels? The superpopulation model of Hartley and Sielken appears to allow the labels to change for different realizations of the population. Since much of the focus of what I'm trying to understand is estimation for a given unit (label), these differences are relevant.

To me, the main function of the superpopulation is to represent the population by a set of random variables from which statistics can be used. Without random variables, there aren't any statistics. [More comment needed]

6. ** Regarding the Voss article.

I read it when it first came out and then read at least one letter and Voss's response when it came out. I haven't gone back to dig out either, but I remember my initial reaction being that 1) I didn't like most of what the letter writer (or writers) was saying and 2) I

though Voss wimped out in his response and seemed to lend credence to what I view as primarily the wrong analysis.