General comments and questions on superpopulations and superpopulation models based on the papers

RE: c00ed64.doc
Voss (1999), Hinkelmann (2000) and Wolfinger and Stroup (2000)

Comments by Julio Singer– 09/01/00
Response by Ed Stanek- 09/06/00

1. I am a little confused about the objectives of using superpopulation models. It seems that there are two situations when they may be employed: the first is when you are interested in making inferences about a finite population which is dynamic (as in the first example given in c00ed64.doc) and the other when you want to use it as an alternative to explain the data generation process (as in an experimental situation where you want to make inferences about an~~na~~ industrial process).

Resonse by EJS: First, I have been distinguishing superpopulations, from superpopulation models. This distinction is not of much importance when the superpopulation model simply represents the units in the population as random variables, but in other settings, the distinction may be important. Underlying nearly every superpopulation model is a conceptual superpopulation. My comments are directed to superpopulations, not the models.

Some of the early discussion of superpopulations in the literature (such as that by Cochran and Deming) was only of an abstract concept of a superpopulation. Although the abstract concept (like a dynamic population) was of interest, assumptions were not seriously reviewed, and as a result, the link between the a method that used superpopulations, and the abstract concept was pretty vague. Only a small amount of the literature that uses superpopulations discusses estimating parameters of the superpopulation. This was done by Fuller, and Konijn, for example, but the superpopulation was linked to the population in a simplistic manner (such as the population being a random sample from the superpopulation). [Note: I have not looked at Fullers and Konijn's papers, and am relying on Hartley and Sielken's remarks about them.] My feeling is that when the superpopulations are ill defined (such as an infinite collection of subjects), the value in their use is also ill conceived. Therefore, I would say that superpopulations have little to contribute as a concept when the finite population is dynamic.

The second example that you give is an experimental setting. If we write the potentially observable population as a single large population, then the superpopulation can be viewed as a set of random variables connected (one by one) to the potentially observable population. This is the idea in c00ed54.doc, which I also refer to in c00ed66v1.doc. I think superpopulations can be used here, and actually underly the old randomization based inference.

2. I think that there are situations where superpopulation models do not make sense: consider for example, inferences about the proportion of defective automobiles produced by a certain company during a specified period in time.

Response by EJS: Suppose the target parameter is the true proportion of defective automobiles produced by the company during a specified period of time. We can list the automobiles made (assuming the time period is in the past), and select a simple random sample without replacement from the list. Then, we can see whether each automobile selected is defective (y=1), or not defective (y=0). How do we use the sample data to draw inference to the population parameter? We do so by representing each observation in the sample as a random variable, and in fact, by

representing the population as a collection of random variables, some of which are observed. This is the basic superpopulation. The superpopulation model is simply the mean model.

3. If superpopulations are essentially finite populations, then saying that the elements of a random sample drawn from it follows a Normal distribution must be viewed as an approximation.

Response by EJS: I agree with this.

4. I tend to agree with Wolfinger and Stroup with respect to Voss's paper. The major problem is related to interpretation of the random factor main effect and its interaction with the fixed effect. Essentially, these two terms are used to define the covariance structure associated to the response variable (since they have null expected values). Therefore we are dealing with two different models and it seems reasonable to have different answers. The hypothesis of interest here (common to both models) seems to be that the covariance term is null and thus I believe that the comment by Schwarz (1993) is appropriate (middle of page 353).

Response by EJS: The comments by Wolfinger and Stroup depends on having different models- whereby each model is a representation of a set of random variables. In the superpopulation settings, the random variables can be linked to potentially observable values in a popualtion. In the model described by Wolfinger and Stroup in their first rebuttal point (p228), they say, "Suppose instead that we simply consider the covariance structure of the observed data (with no additional hypothetical unobserved data of infinite dimension". To form the covariance structure, they have to consider the observed data as a realization of some random variables. How are these random variables to be interpreted? I think if this angle is pursues, they will end up back at the superpopulation model, and therefore, there are not really two models. Wolfinger and Stroup comment that the parameters of the UP model have natural interpretations as 'covariances between certain sets of observations,'. By this, they mean covariances between certain sets of random variables- and the random variables are not defined. Hence, while if you ignore the random variable definitions, the covariances seem interpretatable, this interpretation is based on a basic set of random variables that are metaphysical.

I do think that the question as to whether there are two different interpretable models is the basic question. I only see one of these models as interpretable.