Sampling Methodology: Design versus Process

This paper was written to explain the difference between a sampling design and the process for selecting a sample under that design. The explanation of the differences among Design, Algorithm, and Process in the Methodology Pattern in the DDI-4 model is a guide. This can be found in the document titled "Methodology Pattern Explained: Design vs. Algorithm vs. Process".

One of the main problems with understanding sampling as a design or a process is the language used to name what is being described. People talk about sample selection, when they really mean the design. Other times, they really mean the process. It is important to distinguish the two aspects.

Based on the earlier paper, then, a sample design contains the criteria needed to select a sample. It is not the selecting process itself. For brevity, we will assume the sample is probability based. A similar argument applies to other kinds of samples.

Here is a list of things that need to be known (i.e., execute a process) to select a single stage sample:

- Sampling method –
    - Simple random sampling (SRS)
    - Stratified simple random sampling (SSRS)
    - Systematic random sampling (SysRS)
    - Probability proportional to size sampling (PPS)
    - Others
- Sample size
- Population size
- Strata (if necessary)
    - Number
    - Sample size for each
- Probabilities
    - Inclusion probabilities for PPS
    - Probabilities for each stratum
    - Start, interval, and "take every" for SysRS
    - Relative inclusion probabilities for variance estimation
- Frame
    - Access
    - File format

For multi-stage samples, one needs to know

- Number of stages
- Sampling information for each stage
- Frame information necessary for selection at each stage

Providing the details for each of these criteria above constitutes specifying a sampling design. The key to understanding what the design addresses is what needs to be known in advance. Note that none of this indicates <u>how</u> a sample should be selected. The <u>how</u> is the algorithm and process. The information above just answers <u>what</u> – what constraints need to be satisfied.

The process contains the steps carried out to select a particular sample. For instance, statistical packages contain routines to carry out a wide range of statistical operations. Sampling is included. In the SAS statistical package, for example, a sample can be selected using PROC SURVEYSELECT. The parameters necessary to drive the PROC to select a sample under specified conditions are given in the design. Running the SAS PROC SURVEYSELECT in this case is the process. The algorithm is the steps the PROC is built to execute, depending on the parameters from the design.