STRATIFICATION - HOW CAN WE LIVE WITHOUT IT?

by Lee Brown

Can you imagine trying to use an estimate of corn acreage from an area sample that completely disregarded land-use? Can you imagine using a hog estimate where the list sample was selected with no reference to control data but was simply a random sample from the entire list frame? Can we provide data users with reliable crop and livestock estimates based on these types of sampling plans?

Of course, the answer to these questions is a resounding NO! We cannot allow crop estimates to be based on area samples where a sizable portion of the sample could fall in nonagricultural land like cities or rangelands. Likewise, we cannot rely on livestock estimates based on samples where we would have no idea how many livestock producers would be interviewed or how large their operation might be. These types of designs give very imprecise survey results and, thus, are poor investments for the money spent on them.

One method of controlling these problems is to stratify the sampling frame (either list or area) before the sample is selected. Stratification is the process of subdividing the population into mutually exclusive categories, called strata. In the first example, all the land in each State is divided into different categories creating land-use strata, such as cultivated land, agri-urban, and rangelands. In the second example, hog operators on the list frame are classified by size of operation.

Many probability samples that NASS employs have a stratified design. There are several reasons why stratified sampling is used, along with several tradeoffs that we need to remember as well.

The biggest gain from stratification comes from the improvement in precision of the indications. Here, we mean smaller standard errors from the stratified design versus larger standard errors from a simple random sample over the entire population. Let's discuss how this improvement happens by using corn acreage as an example.

The standard error from a simple random sample will depend on how much the reported corn values, including both positive and zero, vary over all units in the sample. For a stratified sample, the size of the standard error will depend on how much the reported corn values vary within each stratum. As the stratification becomes more efficient at putting "like" operations into the same strata, the smaller the standard error will get. Thus, when the stratification is well-designed and the sample size is properly allocated to the strata, large reductions can be made in the standard errors.

Another advantage of stratified sampling comes from targeting different subgroups of the total population. Extreme operators and specialty crop growers are examples of how targeted

groups can be identified and grouped into separate strata. This capability insures that these types of operations can be "special handled" for certain situations. For example, if a large number of positive rice reports are needed for a general crop survey, then specialty strata based on rice acreage could be used.

An important requirement for creating a stratified sample is to have relatively up-to-date size information about each unit in the population. NASS maintains a profile of each unit in the population, called control data, which tells us what commodities are produced and how big the operation is. If an operation has changed considerably, but the control data available are not current, that operation will be misrepresented in the wrong stratum. Any survey data collected for that operation will definitely be different from the other "like" operations enumerated in that stratum. Consequently, the standard error would increase. Also, item imputation would also be affected since imputed values for list records are computed within the stratum level.

Another stratification problem concerns prioritizing control items so an operation possessing several control items used to create strata can be assigned to the "best" possible stratum. If an operation has both a large number of beef cows and a large number of dairy cows, to which stratum should that operation be assigned? Decisions concerning all these possible situations are made prior to constructing the survey strata, so this assignment can be made when the population is created.

From these discussions concerning the importance for control data, you can see that the basic requirement for a stratified design is good quality control information. A basic relationship exists: the better the control data, the better the stratified sample will perform. If poor control data are used for sampling, more problems associated with misclassification will later be found during summary and analysis.

If we have designed efficient strata for your surveys and you have maintained your list with quality control data, the net result will be reliable survey results for commodity estimation work. Stratification is a powerful and useful tool that affects many facets of your job. Stratification - you can't live without it!