MR. SAM PULLER PROVIDES SOME TIPS ON SAMPLING METHODS

by Bill Iwig

Recently, an energetic young NASS statistician, named Howdy Duit, wanted to learn more about survey sampling methods. He decided to visit that legendary sampler, Mr. Sam Puller.

Howdy Duit: Sam, when you talk about "sampling methods," what are you really talking about?

Mr. Sam Puller: Well, Howdy, "sampling methods" refer to how you identify the sample units that will be in your survey. If you are conducting a nonprobability survey, then in theory you can create a sample any way you want. There are really no constraints. If you are interviewing farmers, you could just start driving down a country road and stop at the first 20 homes. Of course, I know NASS usually sets its own rigorous requirements, even on nonprobability samples. If you are conducting a probability survey, then you must use "fancier" random sampling techniques where all units in your population have a known chance of selection.

Howdy Duit: What is fancy about random sampling? It sounds pretty straightforward to me! What can you tell me about these "fancier" techniques?

Mr. Sam Puller: Well, first of all, there are two common sampling methods: simple random selection or systematic selection.

Howdy Duit: I thought all our samples were random samples. What is this business about a systematic sample?

Mr. Sam Puller: A systematic sample is still a random sample, Howdy. It is just a more structured method of selection. First, you sort all of the units in the population into some type of order, perhaps by county. Then you figure out your sampling interval, which is the number of units in the population divided by the number of units in the sample. If your interval is 10, then you randomly select one of the first 10 units in the population and then take every 10th unit after it. A systematic selection assures that the sample is representatively dispersed across the whole population. In this case, every county would be represented in about the right proportions in the final sample. For other applications, you might want to sort the population by the size of the unit or perhaps just randomly.

Howdy Duit: Okay, I'm with you, Sam. I can use either simple random selection or systematic selection. That still doesn't sound too fancy! What's next?

Mr. Sam Puller: Well, Howdy, the application of these selection processes is what can

get kind of fancy. First, you need to decide if all the units will have an equal chance to be in the sample or whether the probability of selecting each unit will be related to some information you have, such as the size of each unit. I know that NASS uses "size" in at least a couple of different ways to affect the probability of selection. You select fields for your objective yield surveys based on the size of the fields. You want the bigger fields to have a greater chance of being in the yield survey since they represent a larger portion of the acreage for that crop. Second, NASS uses stratification to control the probability of selection.

Howdy Duit: I know we use stratification, but I don't know exactly how it works or how it controls the probability of selection.

Mr. Sam Puller: Stratification refers to the process of dividing the entire population into groups usually based on the size or type of farm. Each group is called a stratum. For example, you might put all farms with over 500 head of hogs into one stratum and sample a large proportion of them since these farms have a large number of hogs. You might create another stratum of farms with less than 100 head of hogs and only sample a few of them. The chance of selection is equal for each farm within each stratum. However, farms in the 500+ hog stratum have a much larger chance of selection than farms in the small hog stratum.

Howdy Duit: Okay, so there is a different probability of selection for each stratum. Then I guess you are saying that the records within each stratum can be selected with either simple random or systematic sampling. Is that correct?

Mr. Sam Puller: That's exactly right, Howdy! Now there are two other applications that you should know about. First, there is replicated sampling where you identify a number of subsamples within a large sample. Each independent subsample is called a replicate. Replicated sampling allows you to have some overlap in your samples from survey to survey by rotating the replicates. A second application is called cluster sampling where you sample clusters or groups of sample units rather than the individual sample units. This technique can help reduce costs.

Howdy Duit: Wow! Things are starting to get fancy now! Does NASS use both of these applications?

Mr. Sam Puller: You sure do, Howdy! You create replicates within strata for many of the samples you draw from your list frame. Your area frame group in Fairfax, Virginia uses both methods for selecting the segments used in the June Area Survey. For example, your segments are clusters of separate farming operations which you call "tracts."

Howdy Duit: So many options, Sam! How do you know which sampling method to use?

Mr. Sam Puller: Good question, Howdy. It depends on a number of different factors, including what information you have about the size of each unit in the population, how the population is organized, what your survey needs are, and what your costs are.

Probably the best advice I can give you is to do some reading on sampling techniques and then come back. We'll talk some more. I recommend the book, " A Sampler on Sampling," by Bill Williams.

Howdy Duit: Thanks, Sam, I'll try to find them.

As you might imagine, the energetic Howdy did find the books and did some studying. Before long people were saying: "How'd he do it? He's starting to sound like a Sam Puller!"