PROBABILITY OR NONPROBABILITY: A SURVEY IS A SURVEY - OR IS IT?

by Vince Matthews

What is the difference between a probability survey and a nonprobability survey? NASS uses both types, but each has its advantages. A textbook definition of a probability survey is that every element in the population has a chance of being selected. This article will expand the definition and contrast probability and nonprobability surveys used by NASS.

A population is a well-defined collection of all the items to be surveyed. In the population of all licensed grain elevators in a State, each elevator is an element of the population. Statisticians try to be specific about who belongs to a population, and "licensed" achieves that for the population of grain elevators. For separate surveys of catfish producers and rice farmers, a grower who raises both belongs to two populations.

In a probability survey, each operation must have a chance of selection. When data are obtained from every operation, a census of the population results. In other words, for a census every operation in the population *is in the survey*. For a probability sample, every operation in the population *has a chance to be in the survey*. The probability survey will estimate the same farm characteristics as the census but will only question a small fraction of the population chosen by chance.

With only a small part of the population chosen for a probability survey, each interview is vital because many other elements of the population are represented by that one interview. The expansion factors are used to expand the individual responses up to an estimate for the entire population. An expansion factor of 293 means that one respondent in the probability survey represents 293 operations in the population.

What is a nonprobability survey? It is any survey which does not conform to the definition of a probability survey. For example, NASS usually tries to pretest new procedures before their adoption into the operational program. Rather than use a random sample for the pretest, NASS will often use a preselected set of farmers in a few specified States because interviews with those farmers are likely to expose as many potential problems as possible in the proposed procedures. NASS uses nonprobability surveys for needs such as crop weather and end of season crop yields.

Now, a comparison of the advantages of probability and nonprobability surveys:

Interpretation of Results. If I tell someone that 4.7 million acres of corn for grain were harvested in Indiana in 1987, that person should reply, "Are you sure?" The person questions how much confidence I have in the estimate. A unique feature of an estimate from a probability survey is that we can measure the precision of that estimate. In other words, we can measure how much that estimate might "bounce around" because we used a sample rather

than a census. The precision of probability estimates is measured by the standard error. Some statisticians feel that the need to measure the precision of estimates is reason enough to use a probability survey for every estimate. Their attitude is that since a nonprobability survey has no standard error, it should have a standard warning, "Let the user beware."

Types of Indications. NASS uses the term “indications” to refer to the statistical point estimates computed from the survey data. We do this to distinguish survey results from the official published “estimate”. The major indication from a probability survey is usually the direct expansion of the data reported by each respondent. Although NASS usually incorporates several indications before releasing an estimate, a direct expansion could be published as an estimate. Data users could then draw their own conclusions in comparisons with previous indications.

The indication from a nonprobability survey is usually judged in relation to a previous month's or year's indication before a figure is published. The indication is not expected to stand alone but instead to show the change that has occurred. Thus, there is a great reliance on seasonal cycles or changes from a base period. An example is NASS's monthly Potato Stocks Survey; those producers who return the December questionnaire become the group which is tracked from month to month as long as they have stocks. Thus, nonprobability surveys rely heavily on being able to model the relationships from one time period to another. The probability surveys tend to rely on direct expansions while nonprobability surveys tend to rely on ratios or percent changes.

Complexity of Procedures. The definition of a NASS probability survey is more stringent than the simple textbook definition. First, the population is usually surveyed simultaneously with list and area sampling frames to overcome list incompleteness. Second, a complex set of procedures must be used to make sure that NASS exactly defines each operation and that NASS avoids or adjusts for duplicate reporting. Third, probability surveys usually require stringent follow-up to farmers who do not respond by mail or telephone. An effort must be made to convert refusals so that response rates meet desired levels.

Nonprobability surveys may be difficult and complex also, but they do not have to obey the three requirements in the above paragraph. Sometimes there is little or no follow-up required, and the survey process might be complete as soon as the questionnaires are returned by mail. Sometimes stringent follow-up is required - it is more a subjective decision of how much effort NASS wishes to put on the survey. Probability surveys, however, are always required to have fairly stringent follow-up.

Consistency of Procedures. Probability surveys demand that procedures are followed exactly from statistician to statistician and from State to State. The surveys that NASS conducts nationwide tend to be probability surveys. We want to state confidently that the same procedures are used in all the States. In contrast, a nonprobability survey may or may not have strict consistency requirements. Again, it is more a matter of how much NASS demands for a particular need. Theoretically, nonprobability surveys do not have any requirements to obey, but NASS may place strict demands on a nonprobability survey because of its importance.

Costs. Nonprobability surveys have a clear advantage in this respect when they require less follow-up. Large costs are incurred in probability surveys because of telephone follow-up to overcome the low response rate to the mailing and, if it is important enough, field enumerators usually follow up on those operations which were inaccessible by telephone. A nonprobability survey may or may not incur these costs - it depends on what NASS demands for a particular situation.

Conclusion. A probability survey is usually more expensive and complex than a nonprobability survey. Nonprobability surveys have no requirements theoretically, and so their costs and complexity vary from one survey to another depending on what NASS has decided to require from each situation. The outstanding feature of a probability survey is that it has a built-in measure of how precise its indications are. This one feature is often enough to tip the balance in favor of conducting a probability survey.

In addition to being able to compute measures of precision for probability indications, an equally important advantage is that these indications are independent, i.e., inferences concerning population characteristics can be made without dependence on any other source of data. This advantage is not generally true for making inferences from nonprobability surveys. Although NASS has implemented probability surveys for most of our statistical program, nonprobability surveys are still useful in certain situations.