University of Nevada Cooperative Extension

Fact Sheet-11-48

Identifying Priority Rural Extension Outreach Needs: How Many Responses Do I Need from My Survey?

Marilyn Smith, Area Youth Development Specialist

George C. Hill, Associate Professor, College of Education

In rural Nevada, the populations are small and varied. Even within very small counties, Extension program needs are sometimes polarized by a variety of issues. Obvious divisions include geographic location (town vs. rural lifestyle), jobs (mining, tourism, government, agriculture), age (youth, family, senior citizens), and gender. So one of the first challenges in trying to identify the highest priority needs for Cooperative Extension community outreach programs in a rural community is to decide how to sample groups of people to survey and how many participants need to be included in the survey.

The purpose of this publication is to provide some basic guidance regarding sampling communities. In statistical terms, the sample refers to the group of persons selected for a study. The population designates the larger group of people from which the sample is drawn. For example, in a community needs assessment survey it is not usually practical to send a survey to every person in that community. More typically, a random sample is selected in a way that provides everyone an equal chance to be selected for the survey so that the sample is representative of the entire community population. Sampling provides the ability to obtain information from fewer respondents to describe characteristics of an entire population.

The purpose of sampling is to include enough respondents to satisfy the statistical probabilities that the results are valid while not having to include every citizen in the survey. Sampling can save both time and money over trying to survey an entire community. With appropriate and accurate sampling procedures, we do not need a response from every person in the community, but rather a sample representative of that population.

Accurate Sampling: Guidance from a statistician is always a recommended approach. Before making that appointment, an individual that is new to survey methodology or has not been actively engaged in the discipline for some time will want to be prepared with some basic knowledge about the county to be surveyed. The discussion with the statistician can be more confusing than helpful unless individuals

are prepared to provide the statistician with some basic information about the study design.

> A discussion with the statistician might go something like this: "I want to find out the highest priority needs for Extension education in the six programming areas in the county where I work. This will be a mailed survey using a 5-point Likert type scale for most questions. The county has 5,000 people in two geographically isolated communities. I plan to use the voter registration list for adult participants for a general survey of all people in the county. The biggest employers in the county are in mining and tourism."
>
> "The number of people employed in agriculture is small in comparison to the total population, but I need to be able to identify the specific needs of this subsample. There are 100 agricultural producers in this county and their contribution to the local economy is great. I think I should use the mailing list from Nevada's State Agriculture Statistician who meticulously maintains a database of agricultural producers to survey this group so that I can make sure I have a good mailing list for this small subset of people."

This Extension professional has thought through several issues to be considered and has gotten to know this county really well. He/she has talked to many people to get this kind of insight. The statistician will be pleased that this Extension professional has done this kind of homework because two of the three most important steps in sampling are completed: (1) identification of target population(s) and (2) availability of a population list(s) (Dillman, Smyth, and Christian, 2009). With a little more discussion, this Extension professional is ready to select a sample.

For the purposes of this example, let us assume that the sample size calculations will be determined for the general survey of the 5,000 people in the county. Let us assume that a separate and more specific survey will be conducted with the 100 agricultural producers.

Sample size calculations: Calculating a sample size for the 5,000 people example provided above can be complicated, but a variety of tools are available to simplify this calculation. Several online calculators are available (do an Internet search for "sample size calculators") and commercial calculators are available for more sophisticated analyses (for example, SPSS SamplePower). Many statistics books and survey research books contain tables and discussions regarding sample size (for example "How Many Subjects" by Kraemer and Thiemann or "How to Conduct Your Own Survey" by Salant and Dillman).

Several decisions must be made when considering sample size even when using one of the computer programs, tables or calculators. In general, though, sample size calculations are non-intuitive because as the population increases, the sample size does not increase proportionally. The following example demonstrates this point. Because it is a non-intuitive phenomenon of statistics, this is one of the issues that are difficult to understand for people who are new to statistical sampling. Look at this example.

In the community of 5,000 adults, roughly 880 responses are needed. In larger populations of 50,000 adults, 1,045 responses are needed, and from a community of 1 million, about 1,067 (all estimates are based upon 3 percent sampling error at a 95 percent confidence level). Paradoxically, using the sample size table provided by Salant and Dillman (1994, page 55) an estimated sample

size of 92 respondents is needed from a population size of 100 farmers.

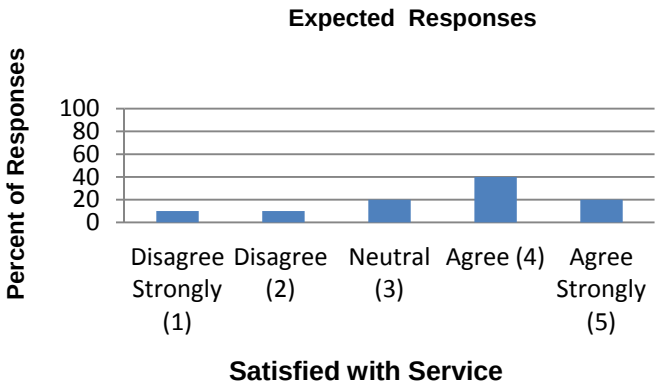
This example demonstrates one of the primary characteristics of probability sampling. It is the size of the sample, not the proportion of the population sampled, that affects precision. In the example of the 100 farmers, little is gained by using a sample of 92. However, when larger communities are involved, sampling is an important tool because it would be a waste of time and resources for a county Extension professional to send a survey to 5,000, 50,000 or even 1 million people when a sample of that population can estimate the results at a confident level.

One guideline that many professionals use is that if the population is less than 500, it is usually more feasible to survey the entire population rather than try to estimate from a sample unless there are special issues or situations where this is not possible. It should also be noted that sample calculations are estimates of the number of completed, usable questionnaires, the actual sample size must also allow for ineligibles and non-respondents.

SPSS SamplePower: One of the computer programs that can help us make decisions about sample size numbers is SPSS SamplePower. This program does the calculations based upon guided decisions to determine the margin of error that we are willing to accept. If the sample size is too small, the margin of error will be too wide. If the sample size is too large, the estimate will be more precise than necessary, and the study will be wasting resources. Three major decisions must be made by the survey researcher to use this computer program: (1) the distribution of response with a calculation of the expected mean, (2) the margin of error, and (3) the expected percent missing.

(1) Let us start with the distribution of response to calculate the expected mean: Suppose you plan to ask the people in your county how satisfied they are with the availability of fresh fruits and vegetables in local stores and shops. Each person will respond to the statement "I am satisfied with this service" using a scale that ranges from "Disagree Strongly" to "Agree Strongly" (corresponding to scores of 1 to 5). Since you expect that most people will agree with the statement, we might expect a logical sequence (low to high) of responses. This is one of the guided decisions that SPSS SamplePower can calculate based upon the user's knowledge of the population to be surveyed. In the example shown in Figure 1, the survey researcher is expecting a distribution where 10 percent disagree strongly, 10 percent disagree, 20 percent are neutral, 40 percent agree, and 20 percent agree strongly. This corresponds to a mean of 3.5 with a standard deviation of 1.20.

Figure 1: Example: Estimated responses regarding satisfaction of the availability of fresh fruits and vegetables on a scale from disagree strongly to agree strongly.



(2) A selection of the margin of error is usually based upon how critical it is that precise measurements are reported. For example, a margin of error (0.1 points) used in

the SPSS SamplePower calculator means that in most cases (95 in 100) the estimated mean (from figure 1) will fall within the range of the true mean plus/minus 0.1 points.

(3) Finally, percent missing (not returning a completed survey) also needs to be estimated. The larger percent missing would mean that a larger number of surveys would need to be mailed. In the example of this survey, we are using a 70 percent missing estimate. This means that for every 100 people enrolled in the study, 70 will not provide data and will be excluded from the analysis.

Sample Size from SPSS SamplePower Calculator: Based upon the assumptions described above, the study will enroll 1,865 people of the 5,000 member population. With this sample size, there is a 95 percent likelihood (confidence level) that the sample mean will fall within 0.1 points (confidence interval) of the true mean. If we observe a mean of 3.50 we will be able to report that the true mean probably falls in the range of 3.50 plus/minus 0.1 points. (In national surveys, this calculation will be shown as (95% +/- 1%). Reporting these confidence levels provides the reader with assurance of the accuracy of the survey.

Summary

A variety of tools are available to help estimate the sample size for a study. The survey researcher must have information about the community to be able to estimate that sample size for that population. The dispersion of scores or the missing data rate can be estimated by looking at other surveys done in the county or in neighboring counties. In the references provided below discussion about how much error can or should be allowed must be considered by the researcher. However, the suggestions provided in this fact sheet do not take the place of consultation with an expert statistician prior to implementation of a survey. Using the suggested tables and/or computer programs will help the survey researcher think through the study design and answers to questions he/she will need to provide the statistician for a productive meeting.

References

Cohen, J. (1989). *Statistical power analysis for the behavioural sciences, 2nd Ed*. Hillsdale, NJ: Erlbaum.

Dillman, Don A., Smyth, J.D., and Christian, L.M. (2009). *Internet, mail and mixed-mode surveys, 3rd Ed*. New Jersey, John Wiley & Sons, Inc.

IBM SPSS SamplePower 3.0.1 (2010). Developed by Michael Borenstein, Hannah Rothstein, David Schoenfeld, Larry Hedges and Jacob Cohen.

Kraemer, H.C. and Theimann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage.

Salant, Priscilla and Dillman, Don A. (1994). *How to conduct your own survey*. New York: John Wiley & Sons, Inc.