



# Harvard University Program on Survey Research

## TIP SHEET ON SURVEY SAMPLING, COVERAGE AND NONRESPONSE

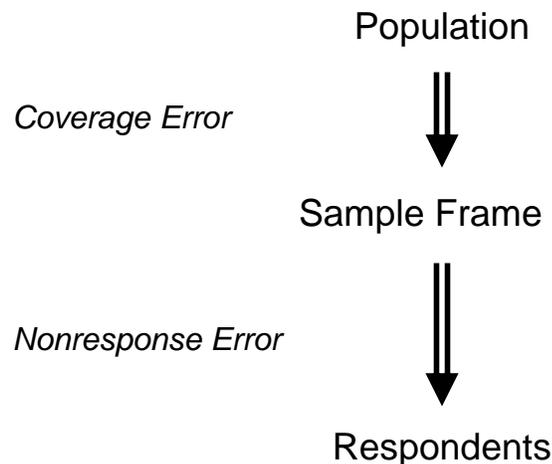
Surveys can be a strong tool because, when done properly, they allow you to generalize your findings to a larger population. This tip sheet is designed to provide an overview of how surveys can generalize to larger populations.

- A **census** attempts to collect data from all members of a population.
- **Random samples** let you use collect data from a portion of a population and use sampling statistics to generalize your findings to a large population.

### Two problems can get in the way of this:

- **Coverage Error** occurs when some members of a population are excluded from the sample frame you use for your study.
- **Nonresponse Error** occurs when some of the respondents you select in your sample don't respond.

### Generalizing from Respondents to a Population in a Sample Survey



### Sample Error

Sample error, calculated for samples based on statistical theory, describes the variability of a sample statistic across multiple hypothetical samples which could be drawn.

- Sample statistics calculate error caused by *variation* between different potential samples of respondents and the entire sample frame.
- Coverage error and nonresponse error are not accounted for in sampling statistics. If coverage error or nonresponse error are systematic, they can introduce *bias* into survey measures.

- **Tip:** Sample error is much easier to calculate than other sources of error, but the numbers commonly reported assume there are no other sources of error.

## Types of Samples

Sampling statistics make assumptions about the type of sample you use. There are many ways or drawing a sample, but only random (probability) samples let you generalize to a larger population.

### **Random (Probability) Samples:**

- Based on probability theory
- Allow generalization
- Sample statistics can be calculated
  - Sample records are drawn from a well-specified frame
  - Sample records are drawn according to random procedures
  - Each sample record has a known probability of selection

### **Non-Random Samples:**

- Do not meet the above criteria
- Inference can only be made to itself
  - Sample error **cannot** be calculated
  - Frame is not well specified
- Benefits:
  - Convenient and cost effective
  - Can be used for idea generation
- **Tip:** Convenience samples, opt-in samples, snowball samples, and other types of non-random samples cannot be generalized to a larger population using sampling statistics.
- **Tip:** Surveys that are “representative” on a number of demographic characteristics are not necessarily random.

**Example:** Opt-in Internet samples are becoming a popular method of collecting data. The only people who have a chance to be included in these samples are people who have web access and visit the web sites that post ads. Folks who are eager to click on ads are more likely to be included than others. *From a sampling standpoint, the researcher only has data from Internet ad clickers who saw and clicked on particular ads, and can't generalize the sample beyond this.*

## Sample Frames

The sample frame is the source of the sample. A well-designed, well-implemented random sample can be generalized to a sample frame.

- A sample frame can be a list, or a set of procedures that could generate a list if needed.
- Not all records in a sample frame will lead to eligible survey respondents. As long as you can identify the records that aren't part of your sample, this isn't a problem for your data.

- If a sample frame might contain the same individuals multiple times, it's important to develop a way to identify this and account for it.
- If you are using a sample of addresses or telephone numbers to reach people, you need to randomly select one person at the address or telephone number.
- **TIP:** Sample frames aren't perfect. It is usually a good idea to verify that the person you contact is the eligible for the study.

## Coverage Error

Coverage Error refers to people excluded for sample frame. Typical sampling statistics are calculated assuming there is no coverage error. If some people are systematically excluded from your sample frame, then the sampling statistics you calculate won't account for coverage error.

- Common sources of coverage error:
  - Telephone surveys exclude people without telephones
  - Most telephone surveys exclude cell-phone users
  - Internet surveys exclude non-Internet users
- Bias: Depends on both magnitude of undercoverage AND differences between those covered and not covered on specific statistics. Bias can be different for different items in the same survey

**Example:** People who use cell-phones exclusively may not differ significantly in vote choice but might have big differences on attitudes toward technology.

**TIP:** If your sample frame excludes a large percentage of your target population, you need to think about how the people who are not covered might differ from those who are included in the sample frame.

## Nonresponse Error

People who do not respond to survey constitute nonresponse. The percentage of people who do not respond to surveys is increasing dramatically.

- Common reasons for nonresponse:
  - *Unable:* People with hearing problems, speak English poorly, etc., might not be able to answer some surveys.
  - *Unavailable:* People who work long hours are often unavailable to complete a survey.
  - *Unwilling:* Refusals to surveys are increasing
- Nonresponse bias is different from the level of nonresponse. Similar to coverage bias, this depends both on the magnitude of nonresponse AND differences between responders and others on specific statistics. Even with the same level of nonresponse, bias can be different for various questions. For example, people who refuse to participate in surveys may not differ significantly in vote choice but might have big differences on attitudes toward privacy and confidentiality.

- There are two different strategies to addressing nonresponse in surveys:
  - *Up Front: Surveys can be designed to reduce nonresponse. Typical measures include advance notification, long field periods, and refusal conversions.*
  - *Back-End: Weights adjust for differences between responders and non-responders based on demographic characteristics. Increasingly sophisticated methods of adjusting for differences between responders and non-responders are being developed and used.*
- **TIP:** If you have high levels of nonresponse in your study, try to think about how people who did not respond to your survey might have differed from those that did. How would this impact your data?

## Sources of Further Reading

### **Sampling and Coverage**

Fink, A. *How to sample in surveys*. Thousand Oaks, CA: SAGE Publications, 1995. A how-to guide.

Kalton, G. *Introduction to Survey Sampling*, Sage University Paper 35: Quantitative Applications in the Social Sciences, 07-035. Sage Publications, Newbury Park, CA, 1983. A basic overview of sample methods that covers a lot of territory in less than 100 pages.

Henry, G. *Practical Sampling Applied Social Research Methods Series*, Vol. 21 Sage Publications, Newbury Park, CA, 1990. A good discussion of practical sampling methods that pays particular attention to non-scientific techniques.

Kish, L., *Survey Sampling*, Wiley, New York, NY: Wiley, 1965. The classic practitioners guide for scientific sample design, with an emphasis on area frames.

Levy, P.S. and Lemeshow, S. *Sampling of Populations*, 3<sup>rd</sup>. ed. New York, NY: Wiley, 1999. A solid and well-rounded book with formulae summarized in readable boxes. Good basic coverage of current software for analyzing complex sample designs. Chapter 15 (written by R.J. Casady and J.M. Lepkowski) is the best overall guide to telephone sampling available.

Dorofev, S. and Grant, P. *Statistics for Real Life Sample Surveys*. Cambridge, UK: Cambridge University Press, 2006. A book that provides practical advice for analyzing and adjusting imperfect survey data.

### **Coverage and Nonresponse**

Weisberg, Herbert. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago, IL: University of Chicago Press, 2005. A strong and readable overview to the many types of errors that you can find in survey data.

Groves, R. M. *Survey Errors and Survey Costs*, 3<sup>rd</sup>. Edition New York: Wiley Interscience, 2003. The bible on survey errors. Now in its third edition.

Biemer, P. B. and Lyberg, L. E. *An Introduction to Survey Quality*. New York: Wiley Interscience, 2003. An approach to survey quality that balances variability with bias.

Lessler, J. T. and Kalsbeek, W. D. *Nonsampling Error in Surveys*. New York: Wiley Interscience, 1992. Comprehensive discussion of the topic.

Updated: November 19, 2006  
Chase Harrison (charrison@gov.harvard.edu)