TIPSHEET – SAMPLING

One of the most important aspects of survey research is selecting a population of interest and developing a method to sample units from that population to survey. There are a number of common sampling techniques, and researchers should carefully consider which approach best suits their needs. This tipsheet discusses sampling and some common sampling methods.

Why sample?

First, some terms:

- Census – a survey of the entire population
- Sample – a subset of the population (ideally randomly selected and representative) from which researchers can draw inferences about the population

Sampling is common because a census is often

- Too expensive
- Too time consuming
- Not sensible

Imagine that a researcher wants to conduct a simple public opinion poll to find out what Americans think about some issue of national importance. A census would provide the survey to all 300 million plus Americans. This would be next to impossible. The costs would be immense, the survey would take an incredible amount of time to administer, and reaching every citizen would be impractical. Instead, the researcher can select a sample of a few hundred Americans and make reliable estimates from that subgroup.

That being said, sampling introduces sampling bias. If it is practical and possible to survey an entire population, then performing a census might be desirable. For example, if a researcher wants to survey professors at a university, the entire population size might only be a few hundred individuals. A web-based survey could easily deliver a survey invitation to all of their email accounts. In this instance, a census is easy. A census, though, will eliminate sampling error, but this is just one of many possible sources of error in a survey.

Sampling terms

- Target population – the population to which you want to generalize your findings
- Unit of analysis – the individual members of the target population
- Sampling frame – a complete list of all units from which a sample is drawn
- Noncoverage bias – a unit has a 0% chance of being sampled because it is not in the frame
    - Systematic – there is some underlying reason for sampling frame exclusion
    - Non-systematic – the bias is completely random

Types of samples

Non-probability – no random selection is used in choosing sampling units
- Convenience samples
- Snowball

Probability – random selection is used in choosing sampling units
- Simple random samples
- Systematic samples
- Stratified samples
- Cluster samples

Simple random sampling

- Every unit has the sample probability of selection and every combination of units has the same probability of selection

- Probability of selection: n/N
  - n = sample size
  - N = population size
- Simplest sampling method
  - Random number table
  - Random number generator

Systematic sampling

- Every element has the same probability of selection, but not every combination can be selected
- Use when SRS is difficult
- Procedure
  - Determine N and n
  - Calculate sampling interval (N/n)
  - Pick random start point
  - Take every ith case
- Periodicity is a problem

Stratified sampling

- Proportionate
  - To ensure the sample resembles some aspect of the population
  - Population divided into subgroups called strata
  - SRS taken from each stratum
- Disproportionate
  - A strata is oversampled to make comparisons between groups or to explore a subgroup in greater depth
  - Population divided into subgroups with the probability of selection higher for smaller stratum
  - But then weighting becomes an issue

Cluster sampling

- Divide population into stratum called "clusters," then sample only some stratum
  - Clusters are almost always geographical
  - Typically a second stage in which SRS occurs within the sampled clusters
- Typically used in face-to-face surveys
  - Reduce cost
  - When no satisfactory sampling frame is available

Author: Patrick R. Miller, DISM Survey Research Associate