

Fundamentals of Survey Sampling

Lisa Klein

University of Wisconsin Survey Center
University of Wisconsin-Madison



UWSC

University of Wisconsin Survey Center



www.uwsc.wisc.edu

Outline

- What is survey sampling?
- Terms and Definitions
 - Sample Frame
 - Non-probability sampling
 - Probability sampling
 - Over-sampling
 - Stratification
- Sampling Applications



UWSC

University of Wisconsin Survey Center





What is Survey Sampling?

What makes a survey a survey?

- The purpose of a survey is to describe a *population* (Kalton, 1983)
- Population= “The totality of elements under study, where the ‘elements’ are the units of analysis”
- Populations must be carefully and deliberately defined, because study design and results depend entirely on the population
 - Who will be included?
 - Who will be excluded?
- Cost typically prohibits us from interviewing all members of a population. We therefore must select a **sample** from within the **population** to interview.

What is Survey Sampling?

- A study must meet two requirements to be used to describe a population:
 - ◆ It must be possible to describe the relationship between the population and the units studied
 - ◆ Comparable information must be obtained from all units studied
- In practice, this means
 - ◆ A census or probability sample
 - ◆ Standardized measurement



UWSC

University of Wisconsin Survey Center



Terms and Definitions *(Groves et al 2004)*

- **Sample Frame:** A list of all of the population members from which the sampled elements can be selected
 - either a list or an equivalent procedure for identifying the population elements (ie, geographic area sampling)
- **Probability Sampling:** Each population element has a known, nonzero chance of sample inclusion
- **Non-Probability Sampling:** The use of volunteers and the purposive choice of elements for the sample on the grounds that they are “representative” of the population



UWSC

University of Wisconsin Survey Center





Sample Designs *(Kalton)*

- A sample design is a set of rules that specifies how a sample of a given size is to be selected
- Sample designs are distinguished on the basis of methods of selection
- Non-probability or model
 - Interpreting results obtained with these samples requires broad assumptions (or models) about the distribution of survey variables in the population
- Probability
 - Probability samples enable one to make inferences based on statistical methods, rather than on assumptions about the distribution of survey variables in the population



UWSC

University of Wisconsin Survey Center





Types of Non-Probability Samples

- Convenience, haphazard, volunteer
 - Person on street, intercepts
- Purposive or expert choice
 - Counties in Wisconsin like those that implemented child support reforms
 - For some purposes, small purposive samples may be very useful, but they cannot describe the population
- Quota
 - Match criterion distribution on key characteristics
 - Stop dialing on RDD of 2000 when reach 1000 completes, leaving some numbers undialed
- Substitution
 - Replace non-respondents with another sample element



Uses for Non-Probability Samples

- Qualitative applications
 - Focus Groups
 - Question Testing
 - Cognitive Interviews
 - Pre-testing of Survey Questions



UWSC

University of Wisconsin Survey Center



Convenience Sampling and Biases

- Non-probability (convenience) samples are inherently subjective because they do not give all members of a population equal chance at inclusion
- Convenience samples therefore cannot represent a population, because they cannot be assessed by assumption-free statistical methods



Probability Samples

- Define population, elements, and selection process in such a way that each unit has a known nonzero probability of selection
- Known nonzero probability of selection enables sampling distribution of estimators to be described
- Probability samples have the property of measurability: the sampling variability of the design can be estimated
- A probability sample has two mutually defined components
 - Selection process
 - Estimation process or estimator
- In random samples each element has equal and independent chance of selection. Every sample of a given size is equally likely.

Probability Samples and Survey Error

Four Main Types of Survey Error *(Dillman, 2007)*

- **Sampling Error:** The result of sampling some, not all, elements of the survey population
- **Coverage Error:** The result of not allowing all members of the survey population to have an equal or known chance of being sampled for participation in the survey
- **Measurement Error:** The result of poor question wording or questions being presented in such a way that inaccurate or un-interpretable answers are obtained
- **Non-Response Error:** The result of people who respond to a survey being different from sampled individuals who did not respond in a way relevant to the study

Properly implementing probability sampling techniques decreases sampling and coverage error



University of Wisconsin Survey Center



Stratified Sampling

- If certain information is known about the elements of the population being studied, this information can be used in a supplementary manner via stratified sampling techniques
 - ◆ to improve sample design
 - ◆ for analytical purposes at the conclusion of the study
- **Stratification:** The classification of the studied population into mutually exclusive subpopulations (strata), and ***the selection of separate samples from each of these strata***



UWSC

University of Wisconsin Survey Center



Stratified Sampling

- Stratified sampling takes the total studied population and splits it into smaller units for analyses
- Examples of strata
 - Geographic
 - Age
 - Types of Treatment
 - Types of Illness
 - Race



UWSC

University of Wisconsin Survey Center



Stratified Sampling

- **Proportionate Stratification:** Strata sample sizes are made proportional to the strata population sizes using an estimator
- Stratified sampling design results in increased precision
- Sampling techniques are often very complex. A sampling statistician can be a helpful resource when designing a stratified sample



UWSC

University of Wisconsin Survey Center



Over-sampling

- A target sub-group within a population may be rare relative to the sample frame that covers it (*Groves, 2004*)
- Over-sampling is the “selective application of a higher sampling fraction to rare sub-groups of particular interest in the population studied, so as to ensure that the final sample includes a sufficient number of these rare cases to permit separate analysis”
- Screeners can be included within a survey design in order to facilitate location of the rare population within the frame population

Over-sampling

- Oversampling techniques are commonly used to achieve a sample large enough to make generalizations about data from a very small population within in a larger sample frame
- Examples include:
 - Racial and ethnic minority groups
 - Groups utilizing certain services
 - Groups of people who partake in specific behaviors



UWSC

University of Wisconsin Survey Center



Post-Survey Adjustments *(Groves 2004)*

- Due to...
 - Non-Response
 - Coverage error between sample frame and target population
- ...statistics based on respondents may not match those of the target population (under-representation of groups)
- Frequent sources of error
 - Lack of ability to contact certain populations (ie, respondents without telephones or addresses)

Post-Survey Adjustments *(Groves 2004)*

- Post-survey weighting adjustments are also made to adjust for over-sampling
 - Corrects the proportion of the over-sampled group back to their representative proportion of the actual population

Sampling Resources

- Sampling work can be exceptionally complex
- The training and background necessitated for sampling statistics is rigorous
- Many survey research organizations contract sampling work out, or employ a sampling statistician on staff



Conclusions

- Proper sampling is an essential component in the accuracy of data collection
- Random sampling techniques are required in order to make statistical generalizations about larger populations
- Convenience sampling has a role in scientific research; however, it's applications are limited to qualitative aspects of survey work
- Over-sampling and stratified sampling play a useful role in ensuring representation of sampled populations