Sampling Design

BUS 230: Business Research and Communication

1

1.1 Goals and Learning Objectives

Goals and Learning Objectives

- Goals of this chapter:
  - Learn methods for obtaining samples from populations.
  - Learn biases that can occur when using non-probability approaches.
- Learning objectives:
  - LO2: Recognize and use the appropriate techniques to collect or use survey data to address a research problem.
  - LO2.C: Identify sources of respondent and administrative error and develop the ability to construct and administer a survey instrument that minimizes these errors.

2 Sampling Design

2.1 Gallop Poll Example

Sampling Design

- Example: Every three months the Gallop poll attempts to discover the top financial concerns of households in the United States.
- Population: All U.S. households.
- Identifies top financial concerns in their sample, reports what percentage of households list each as one of their top concerns.
- With 95% confidence, report the percentages *for all U.S. households* within a 3% margin of error.
- Survey only 1000 households.
- Any doubters? Any concerns?

2.2 Population and Sample

Population Versus Sample

- **Statistics**: the study of how to use data to answer interesting questions.
- **Population**: the complete collection of all elements to be studied.
- **Census**: collection of data that includes *every* member of the population.
- **Sample**: a collection of data from a subset of members from a population.
- **Statistics**: method of using *sample data* to make *statements or inferences* about a population.
  - Confidence intervals: statements concerning the degree of confidence and margin of error.
  - Hypothesis testing: using sample estimates and margins of errors to test statements *about the population*.

Sampling Design

Simple Random Sampling

- Sampling technique when *every member of the population* has an equal probability of being selected into a sample.
- Implies sample should be representative of population (on average, by random chance it may not be).
- Even with an infinitely large sample, statistical theory can generate the accuracy seen in the Gallop poll in a sample as small as 1000 elements.

Potential Problems

- Did the sampling method truly allow all elements of the population an equal chance of being selected?
- Did the sampling method allow all subgroups a fair chance of being represented?
- All age groups? all racial and ethnic groups? all income groups?

2.3 Sampling Frame

Sampling Frame

Sampling Frame

- Aka "working population," list of elements from which a sample will be drawn.
- Example Population: UW-L undergraduate students.
- Sampling frame: E-mail directory of UW-L students constructed week before Fall 2011 semester.

Sampling frame error

- When sampling frame creates errors in generating samples.
- Some members of population are excluded from the sampling frame, or..
- Sampling frame includes members which are not part of the population, or..
- Sampling frame inadvertently alters the probabilities in which sample elements are chosen.

Sampling Frame Error Examples

- UW-L Student Example:
  - Some people on the e-mail list may have dropped out of UW-L.
  - Some e-mail addresses may have a typo.
  - Some e-mail boxes may be full.
- Potential Voters for next election:
  - Sampling frame: voter registration database.
  - Some potential voters may not have registered yet.
  - Some potential voters may have moved.

Multi-Stage Sampling

- **Multi-stage sampling:** when the means of taking a sample is broken into stages.
- Example: population = airline passengers.
  - Stage 1: Select airline companies.
  - Stage 2: Select major hub airports.
  - Stage 3: Select dates to gather data.
  - Stage 4: Select individual flights from each selected airline, airport, for a given date.
  - Stage 5: Survey everyone on the selected flights.
- **Sampling unit:** element or group of elements that is selected to a sample.
  - Above example: airline flight.
  - Simple random sampling: individual flier, individual member of population.
- Categories in each stage also referred to as a "sampling unit."

2.4 Sampling Error

Random Sampling Error

- **Random sampling error:** difference between sample statistic and population parameter.
- Unbiased error.
- Decreases with larger sample sizes.
- Easy to estimate the size of the error.
- We use the estimate of the error to construct confidence intervals, hypothesis tests.

Systematic Sampling Error

- **Systematic sampling error:** errors that are not due to chance, but are due to flaws in the way the sample is drawn.
- Size of the error cannot be estimated.
- Causes bias, cannot be estimated.
- Similar to biases caused by self-selection, poorly constructed survey questions.
- This bias is caused by the sampling frame.

Examples of Systematic Error

- Mail surveys: it has been found that people with more education are more likely to fill out mail survey forms than people with less education.
  - End up with a sample with a higher average level of education than the population.
  - What if your outcome variable is related to education attainment?
  - Willingness to buy a product, financial concerns, etc.
- Telephone surveys:
  - Do those with unlisted numbers, or on "do-not-call lists" have shared characteristics related to the outcome?
  - Do people with only mobile phones have shared characteristics related to the outcome?

Probability vs Non-probability Sampling

- **Probability sampling:** a-priori, every member of the population has a known, non-zero, probability of being selected into the population.
- **Simple random sample:** probability sample where every member has an equal probability of being selected.
  - Suppose Gallop poll is a simple random sample. Sample size=1000. Population size=112,611,029.
  - Each household in population has a 0.000889% chance of being selected.
- **Non-probability sampling:** Sampling technique in which elements of a population are selected based on personal judgment or convenience.
  - Since probabilities are not known, cannot fully rely on statistical theory to make accurate confidence intervals, hypothesis tests.
  - Most of the time, researchers say "So what? I don't care."

3 Non-Probability Sampling Methods

3.1 Convenience Sampling

Convenience Sampling

Convenience sampling: sampling technique where most convenient elements are drawn from the population.

Examples

- Point of contact samples (grab people on their way into a store)
- Website visitor survey.
- Store receipt survey
- Me! Teaching and learning techniques for BUS 230.

Important to keep in mind: can you think of reasons for why a particular example can cause **systematic sampling bias**.

3.2 Judgment Sampling

Judgment Sampling

Judgment sampling: researcher uses his or her own judgment for determining who is put in the sample.

Examples

- Test markets for new products: company chooses a set of "typical" cities it believes will most closely match the national market.
- Company with many clients may select key accounts to research customers' opinions.

3.3 Quota Sampling

Quota Sampling

- **Quota sampling:** put a minimum requirement on the number of observations that must be drawn from a number of subgroups.
- Used to assure all subgroups are adequately represented.

Examples

- Make sure to sample at least 100 customers from each age group.
- Make sure to sample at least 100 customers from each income group.

3.4 Snowball Sampling

Snowball Sampling

- **Snowball sampling process:**
  1. Initial sample is selected,
  2. Then friends or contacts of the members in the initial sample are selected (or self-selected)
  3. Then their friends or contacts are selected (or self-selected).
  4. Repeat step 3 and watch your sample size snowball.
- Benefit: it may be very hard to find any members of very unique populations. Eg: harp players.
- Problems:
  - Very non-random. Individuals in round i are likely to have similar characteristics as those in round $i-1$.
  - Often, every round of sampling is pure self-selection bias.

3.5 Systematic Sampling Bias

Systematic Sampling Bias

- Haphazard sampling using non-probability sampling methods can create biases.
- **Convenience Sampling:** why were these elements convenient? Is it related to the outcome variable.
  - Example: Study on academic performance, Sampling method = clock tower point-of-contact at 12pm.
  - Example: Store receipt survey, J.C. Penny offers 15% off next purchase if you complete the survey online.
- **Quota Sampling:** may cause some groups to be over-represented. Does being a member of a particular group influence the outcome?
- **Snowball sampling:** probably big systematic sampling bias.

4 Probability Sampling Methods

4.1 Systematic Sampling

Probability Sampling Methods

- **Simple random sampling:** every member has an equal chance of being selected, sample elements are randomly drawn.
- **Systematic sampling:** A starting point is selected, then every nth element is selected from the sampling frame.
  - This will yield random results if ordering of elements in the sampling frame is random.
- Periodicity: when outcome variable or characteristics of population follow a pattern.
  - Example: Don't collect retail sales information every 7th day.
  - Example: donor list ordered by size of the donations. Don't sample every 50th person if the 10 largest are first.

4.2 Stratified Sampling

Stratified Sampling

- **Stratified Sampling:** Has the same goal of quota sampling, make sure various subgroups of the population are adequately sampled.
- Process:
  1. Population is separated into groups, or *strata*.
  2. Random sampling is conducted within each strata.
- **Proportional stratified sample:** number of observations drawn from each strata is proportional to the population size of that strata.
- **Disproportional stratified sample:** some strata are over-sampled (relative to their population size), likely because their population is small, but the findings from that particular strata are interesting or important.

4.3 Cluster Sampling

Cluster Sampling

- **Cluster sampling:** an economical, multi-stage sampling method.
- Process:
  1. Population is broken into clusters that *do not likely differ* based on the outcome variable.
  2. Clusters are *selected randomly*.
  3. *Every item* of a particular cluster is put into a sample.
- If sampling a company's customers involved travel costs, economical not to fly to every city where there are customers. Pick cities randomly, survey all customers in these cities.