IPDET

Module 9: Choosing the Sampling Strategy

Introduction

- Introduction to Sampling
- Types of Samples: Random and Nonrandom
- Determining the Sample Size

Sampling

- Is it possible to collect data from the entire population? (census)
  - If so, we can talk about what is true for the entire population
  - Often we cannot (time/cost)
  - If not, we can use a smaller subset: a SAMPLE

Concepts

- Population
  - the total set of units
- Census
  - collection of data from an entire population
- Sample
  - a subset of the population
- Sampling Frame
  - list from which to select your sample

More Sampling Concepts

- Sample Design
  - methods of sampling (probability or non-probability)
- Parameter
  - characteristic of the population
- Statistic
  - characteristic of a sample

Random Sampling

- Lottery, each unit has a equal chance of being selected
- Can make estimates about the larger population based on the subset
- Advantages:
  - eliminates selection bias
  - able to generalize to the population
  - Often cost-effective

Types of Random Sampling

- Simple random sample
- Random interval sample
- Random-start and fixed-interval sample
- Stratified random sample
- Random cluster sample
- Multistage random sample

Simple Random Samples

- Most common and simplest
- Establish a sample size and proceed to randomly select units until we reach the sample size
- Uses a random number table to select units

Random Interval Samples

- Use when there is a sequential population that is not already enumerated and would be difficult or time consuming to enumerate
- Uses a random number table to select intervals

Random-Start and Fixed-Interval Samples

- Starting point is random, but the interval is NOT random
- Systematic sampling, starting from a random place and then selecting every n^{th} case

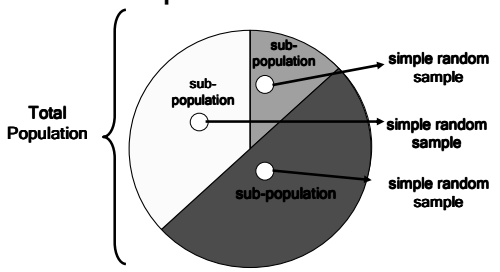
Random-Start and Fixed-Interval Samples

1. Estimate the number of units in the population
2. Determine desired sample size
3. Divide step (1) by step (2) = interval
4. Blindly designate a starting point in the population
5. Count down the interval and select that unit for the sample
6. Continue counting down the same interval and selecting the units

Stratified Random Samples

- Use when specific groups must be included that might otherwise be missed by using a simple random sample
  - usually a small proportion of the population

Stratified Random Samples



IPDET © 2009 13

Random Cluster Samples

- Another form of random sampling
- Any naturally occurring aggregate of the units that are to be sampled that can be used when:
  - you do not have a complete list of everyone in the population of interest but have a list of the clusters in which they occur **or**
  - you have a complete list of everyone, but they are so widely distributed that it would be too time consuming and expensive to send data collectors out to a simple random sample

IPDET © 2009 14

Multistage Random Samples

- Combines two or more forms of random sampling
- Most commonly, it begins with random cluster sampling and then applies simple random sampling or stratified random sampling

IPDET © 2009 15

Drawback of Random Cluster and Multistage Random Sampling

- May not yield an accurate representation of the population

IPDET © 2009 16

Nonrandom Sampling

- Can be more focused
- Can help make sure a small sample is representative
- Cannot make inferences to a larger population (you cannot generalize if you do not randomize…)

IPDET © 2009 17

Types of Nonrandom Samples

purposeful (judgment)	set criteria to achieve a specific mix of participants
snowball (referral chain)	ask people who else you should interview
convenience	whoever is easiest to contact or whatever is easiest to observe

IPDET © 2009 18

Forms of Purposeful Samples

- Typical cases (median)
- Maximum variation (heterogeneity)
- Quota
- Extreme-case
- Confirming and disconfirming cases

Snowball

- Also know as chain referral samples
- When you do not know who or what to include in sample
- Often used in interviews – ask interviewee for suggestions of other people who should be interviewed
- Use cautiously

Convenience

- Based on evaluator convenience
  - visiting whichever project sites are closest
  - interviewing whichever project managers are available
  - observing whichever physical areas project officials choose
  - talking with whichever NGO representatives are encountered

Shortcomings of Nonrandom Sampling

- Can be subject to all types of bias
- Are they substantially different from the rest of the population?
  - collect some data to show that the people selected are fairly similar to the larger population (e.g. demographics)

Combined Sampling Strategies

- Example:
  - Nonrandomly select two schools from poorest communities and two from the wealthiest communities
  - Select a random sample of students from these four schools

Determining the Sample Size

- Statistics are used to estimate the probability that the sample results are representative of the population as a whole
- Evaluators must choose how confident they need to be

Confidence Level

- Generally use 95% confidence level
  - 95 times out of 100, sample results will accurately reflect the population as a whole
- The higher confidence level, the larger sample needed

Sample Size and Population

- By increasing sample size, you increase accuracy and decrease margin of error
- The smaller the population, the larger the needed ratio of the sample size to the population size
- Aim for is a 95% confidence level and a ±5% confidence level

Sample Sizes for Large Populations

Margin of Error	Confidence Level		
	99%	95%	90%
± 1%	16,576	9,604	6,765
± 2%	4,144	2,401	1,691
± 3%	1,848	1,067	752
± 5%	666	384	271

Summary of Sampling Size Issues

- Accuracy and precision can be improved by increasing the sample size
- The standard to aim for is a 95% confidence level and a margin of error of +/- 5%
- The larger the margin of error, the less precise the results will be
- The smaller the population, the larger the needed ratio of the sample size to the population size

A Final Note….

"The world will not stop and think – it never does, it is not its way; its way is to generalize from a single sample" -- Mark Twain

Questions?