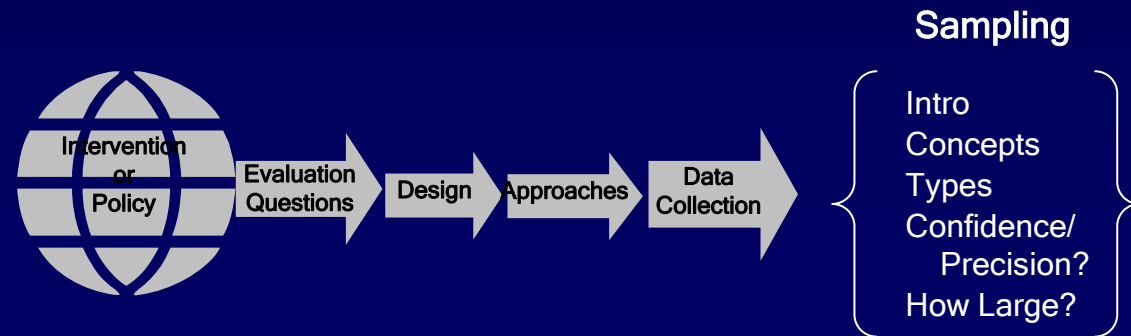
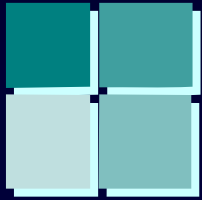


IPDET

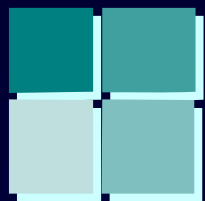
Module 9: Sampling





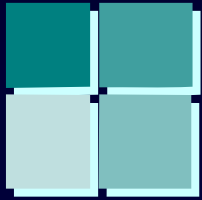
Introduction

- Introduction to Sampling
- Sample Concepts
- Types of Samples: Random and Non-Random
- How Confident and Precise Do You Need to Be?
- How Large a Sample Do You Need?



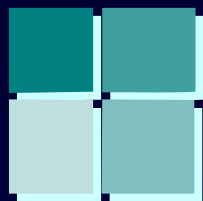
Sampling

- 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



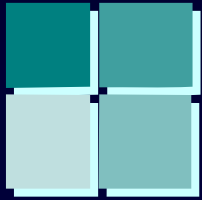
Concepts

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



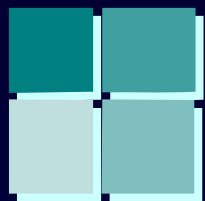
More Sampling Concepts

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



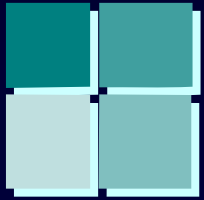
Random Sample

- A random sample allows us to make estimates about the larger population based on what we learn from the subset
- Advantages:
 - eliminates selection bias
 - able to generalize to the population
 - cost-effective



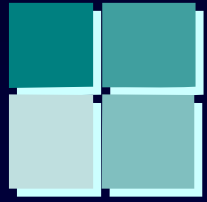
Types of Random Samples

- Simple random sample
- Stratified random sample
- Multi-stage random sample
- Cluster random sample



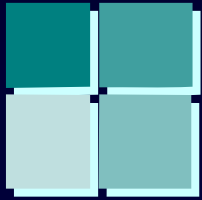
Simple Random Sample

- Simplest
- Establish a sample size and proceed to randomly select units until we reach the sample size

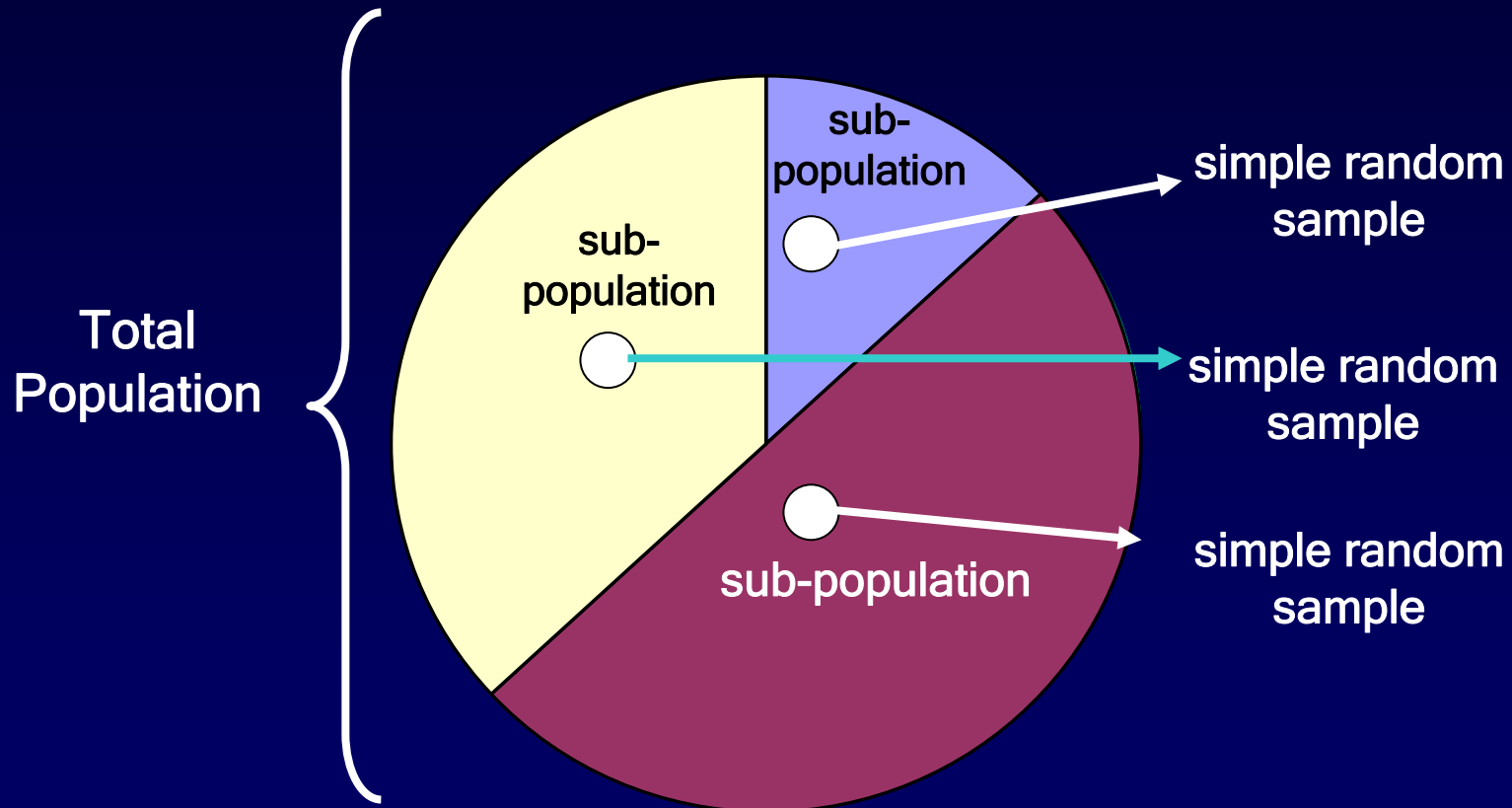


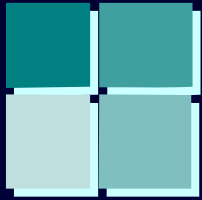
Stratified Random Sample

- 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



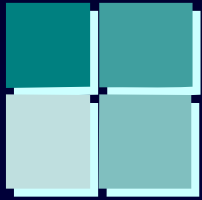
Stratified Random Sample





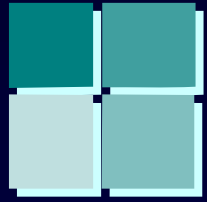
Cluster Sample

- Another form of random sampling
- Any naturally occurring aggregate of the units that are to be sampled that are 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



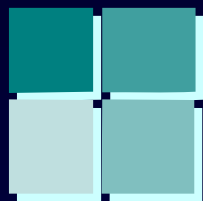
Multi-stage Sample

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



Drawback of Cluster and Multi-stage Sampling

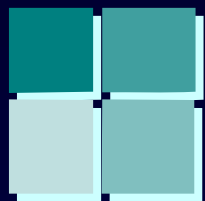
- May not yield an accurate representation of the population



Summary of Random Sampling Process

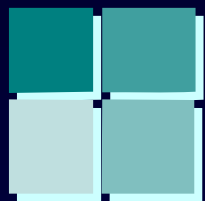
Step Process

- 1 Obtain a complete listing of the entire population
- 2 Assign each case a number
- 3 Randomly select the sample using a random numbers table
- 4 When no numbered listing exists or is not practical to create:
 - take a random start
 - select every n^{th} case



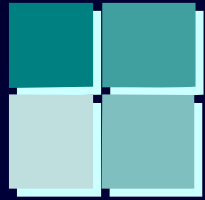
Non-Random Samples

- Can be more focused
- Can help make sure a small sample is representative
- Cannot make inferences to a larger population



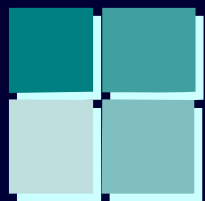
Types of Non-Random Samples

- quota set number from each subset
- snowball ask people who else you should interview
- judgmental set criteria to achieve a specific mix of participants
- convenience whoever is easiest to contact or whatever is easiest to observe



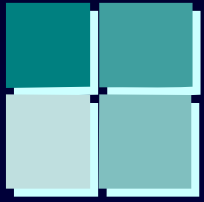
Non-Random Sampling Issues

- People selected in a biased way?
- 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)



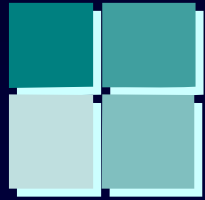
Combinations: Random and Non-Random

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



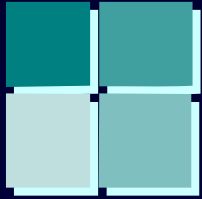
Possibility of Error

- Sample different from the population?
- Statistics: data derived from random samples



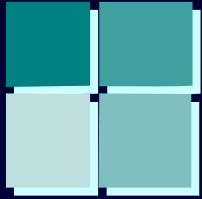
How Confident Do You Wish to Be?

- Confidence level
 - E.g., 90% (90% certain your sample results are an estimate of the population as a whole)
- The higher confidence level, the larger sample needed



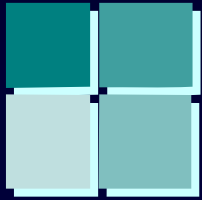
Confidence Standard

- Standard is 95%
 - 19 of 20 samples would have found similar results
 - we are 95% certain that the population parameter is somewhere between the lower and upper confidence interval calculated from the sample



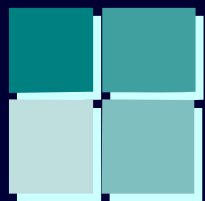
Confidence Interval

- Sometimes called sampling error or margin of error
- Example:
 - in polls 48% for, 52% against, with (+/- 3%)
 - actually means 45% to 51% for and 49% to 55% against
- Also known as precision or margin of error



Sample Size

- 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 $\pm 5\%$ confidence level



Sample Sizes for Large Populations

Precision

Confidence Level

| | 99% | 95% | 90% |
|-----------|--------|-------|-------|
| $\pm 1\%$ | 16,576 | 9,604 | 6,765 |
| $\pm 2\%$ | 4,144 | 2,401 | 1,691 |
| $\pm 3\%$ | 1,848 | 1,067 | 752 |
| $\pm 5\%$ | 666 | 384 | 271 |

 To continue on to the
 **Next Module** click here

 To return to the
 **Table of Contents** click here