Chapter 4, Section 1

4.1: Samples & Surveys

Section 4.1
Samples and Surveys

After this section, you should be able to...

- IDENTIFY the population and sample in a sample survey
- IDENTIFY voluntary response samples and convenience samples
- DESCRIBE how to use a table of random digits to select a simple random sample (SRS)
- DESCRIBE simple random samples, stratified random samples, and cluster samples
- EXPLAIN how undercoverage, nonresponse, and question wording can lead to bias in a sample survey

Populations and Samples

The population in a statistical study is the entire group of individuals about which we want information.

A sample is the part of the population from which we actually collect information. We use information from a sample to draw conclusions about the entire population.

How do we gather data?

- Surveys
- Opinion polls
- Interviews
- Studies
  - Observational
  - Retrospective (past)
- Experiments

The Idea of a Sample Survey

Step 1: Define the population we want to describe.
Step 2: Say exactly what we want to measure.

A “sample survey” is a study that uses an organized plan to choose a sample that represents some specific population.

Step 3: Decide how to choose a sample from the population.

Sampling Design

- Sampling Design: method used to choose the sample from the population
- Types of Samples:
  - Simple Random Sample
  - Stratified Random Sample
  - Systematic Random Sample
  - Cluster Sample
  - Multistage Sample
Simple Random Sample (SRS)

- Consist of \( n \) individuals from the population chosen in such a way that
  - every individual has an equal chance of being selected
  - every set of \( n \) individuals has an equal chance of being selected

Methods of Selecting an SRS

- Draw names from a hat
- Assign each person in the group and randomly generate chosen numbers
  - Ways to randomly generate numbers
    - Computer
    - Random Table of Digits
    - Calculator

Table of Random Digits

A table of random digits is a long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with these properties:
- Each entry in the table is equally likely to be any of the 10 digits 0 - 9.
- The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.

Use Table D at line 130 to choose an SRS of 4 hotels.

<table>
<thead>
<tr>
<th>01 Aloha Kai</th>
<th>08 Cabina</th>
<th>15 Palm Tree</th>
<th>22 Sea Shell</th>
</tr>
</thead>
<tbody>
<tr>
<td>02 Anchor Down</td>
<td>09 Casa del Mar</td>
<td>16 Radisson</td>
<td>23 Silver Beach</td>
</tr>
<tr>
<td>03 Banana Bay</td>
<td>10 Coconuts</td>
<td>17 Ramada</td>
<td>24 Sunset Beach</td>
</tr>
<tr>
<td>04 Banyan Tree</td>
<td>11 Diplomat</td>
<td>18 Sandpiper</td>
<td>25 Tradewinds</td>
</tr>
<tr>
<td>05 Beach Castle</td>
<td>12 Holiday Inn</td>
<td>19 Sea Castle</td>
<td>26 Tropical Breeze</td>
</tr>
<tr>
<td>06 Best Western</td>
<td>13 Lime Tree</td>
<td>20 Sea Club</td>
<td>27 Tropical Shores</td>
</tr>
<tr>
<td>07 Cabana</td>
<td>14 Outrigger</td>
<td>21 Sea Grape</td>
<td>28 Veranda</td>
</tr>
</tbody>
</table>

| 69051 | 64817 | 87174 | 09517 | 84534 | 06489 | 87201 | 97245 |

Our SRS of 4 hotels for the editors to contact is: 05 Beach Castle, 16 Radisson, 17 Ramada, and 20 Sea Club.

SRS

- Advantages
  - Unbiased
  - Easy

- Disadvantages
  - Large variance/high variability
  - May not be representative
  - Must be able to identify entire population

A university’s financial aid office wants to know how much it can expect students to earn from summer employment. This information will be used to set the level of financial aid. The population contains 478 students who have completed at least one year of study but have not yet graduated. A questionnaire will be sent to an SRS of 100 of these students, drawn from an alphabetized list. Starting at line 135, select the first three students in the sample.

135 66925 55658 39100 78458 11206 19876 87151 31260 136 08421 44753 77377 28744 75592 08563 79140 92454 137 53645 66812 61421 47836 12609 15373 98481 14952
Chapter 4, Section 1

**Stratified Random Sample**

- Population is divided into homogeneous (alike) groups called strata
  - Strata 1: Seniors
  - Strata 2: Juniors
- SRS’s are pulled from each strata
- Helps control for lurking variables

**Advantages**
- More precise unbiased estimator than SRS
- Less variability
- Cost reduced if strata already exists

**Disadvantages**
- Difficult to do if you must divide stratum
- Formulas for SD & confidence intervals are more complicated

**Common Strata**

- What are some common stratas in the following areas?
  - Politics
  - School

**Systematic Random Sample**

- Pick a method of identifying subjects randomly before starting
- Requires strict adherence
- **Example:** Suppose a supermarket wants to study buying habits of their customers, then using systematic sampling they can choose every 10th or 15th customer entering the supermarket and conduct the study on this sample.

**Cluster Sample**

- Based upon location
- Randomly pick a location & sample all there
- **Examples:**
  - All houses on a certain block
  - All houses in a specific zip code
  - All students at specific schools in MDCPS
  - All students in specific homeroom classes
Chapter 4, Section 1

Cluster Samples

- Advantages
  - Unbiased
  - Cost is reduced

- Disadvantages
  - Clusters may not be representative of population
  - Formulas are complicated

Multistage Sample

- At least two separate levels/stages of SRS.
- Example:
  - Stage 1: Juniors vs. Seniors
  - Stage 2: Divide the above groups (Juniors and Seniors) by AP, Regular and Honors...select 10 for each of the groups for a total of 60.

Sampling at a School Assembly

Describe how you would use the following sampling methods to select 80 students to complete a survey.

- (a) Simple Random Sample
- (b) Stratified Random Sample
- (c) Cluster Sample

Identify the Sampling Design

1) The Educational Testing Service (ETS) needed a sample of colleges. ETS first divided all colleges into groups of similar types (small public, small private, etc.) Then they randomly selected 3 colleges from each group.
Identify the Sampling Design

2) A county commissioner wants to survey people in her district to determine their opinions on a particular law up for adoption. She decides to randomly select blocks in her district and then survey all who live on those blocks.

Identify the Sampling Design

3) A local restaurant manager wants to survey customers about the service they receive. Each night the manager randomly chooses a number between 1 & 10. He then gives a survey to that customer, and to every 10th customer after them, to fill it out before they leave.

Check Your Understanding:

A very bored FLVS student at ATM wants to know the average number of pages in all the books in the library. The library has 20,000 books, arranged by type (fiction, biography, history, etc.) in shelves that hold about 50 books each. You want to select a random sample of 500 books.

(a) Explain how to select a stratified random sample of 500 books. Justify your choice of strata. Why might the bored student want to conduct a stratified random sample?

The student should stratify by book type to ensure a representative sample of ALL type of books, since different types of book categories may have dramatically different lengths.

The student should proportionally sample from each type of book by placing all titles from one category/strata in a hat and selecting a proportional amount. The process repeats for all the stratas for a total of 500 books sampled.

(b) Explain how to select a cluster sample of 500 books. Justify your choice of clusters. Why might the bored student want to conduct a cluster sample?

Cluster sampling would be logistically easier. With a cluster sample you randomly select 10 shelves and count the pages in ALL 50 books on each selected shelf, for a total of 500 books.

Inference for Sampling

The purpose of a sample is to give us information about a larger population.

The process of drawing conclusions about a population on the basis of sample data is called inference.

Why should we rely on random sampling?

1) To eliminate bias in selecting samples from the list of available individuals.

2) The laws of probability allow trustworthy inference about the population
   • Results from random samples come with a margin of error that sets bounds on the size of the likely error.
   • Larger random samples give better information about the population than smaller samples.
Errors in Surveys

Sources of Error in Sample Surveys

Undercoverage occurs when some groups in the population are left out of the process of choosing the sample.

Nonresponse occurs when an individual chosen for the sample can’t be contacted or refuses to participate.

A systematic pattern of incorrect responses in a sample survey leads to response bias (wanting to look cool, not wanting to be a prude, etc.).

The wording of questions is the most important influence on the answers given to a sample survey.

Voluntary response bias occurs when participation is optional. Usually only people with strong opinions respond.

Errors?!

• How much do you weigh?
• Will you not vote for President Trump’s reelection?
• Why should guns be outlawed?
• How often do you exercise?
• How many cigarettes do you smoke each week?
• How often should Mrs. Daniel give quizzes?