**Section 4.1 - Sampling & Surveys (Part 1) pp. 206-223**

**1. Populations and Samples.** Statistics is largely practiced in order to make inferences about populations of individuals based upon a sample chosen to represent the population. In this section, we are going to explore how to sample populations.

|  |
| --- |
| **Definitions:** 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. |

**Example:** Identify the population and sample in each of the following examples:

(a) The student government at a high school surveys 100 of the students at the school to get their opinion about a change in the bell schedule.

(b) The quality control manager at a bottling company selects a sample of 10 cans from the production line every hour to see whether the volume of the soda is within acceptable limits.

**2. The Idea of a Sample Survey**

The first step in a **sample survey** is to say exactly what *population* we want to describe. The second step is to say exactly what we want to *measure*, that is, to give exact definitions to our variables. The term sample survey is reserved for studies that use an *organized plan* to choose a sample to represent a population. The final step in planning a survey is to decide *how* to choose a sample from the population.

It should be noted that a survey or sample survey does *not* only refer to studies where people are asked questions. Choosing the cans in the example above is a type of sample survey.

**3. How to Sample Badly**

a. **Convenience Samples** - A convenience sample is choosing individuals who are easiest to reach.

**Example**:

* Convenience samples often produce *unrepresentative data*.
* Convenience samples are almost guaranteed to be *biased*.

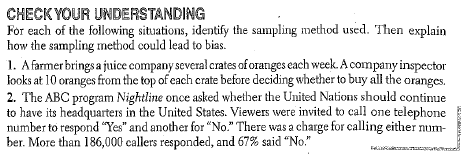
|  |
| --- |
| **Definition**: The design of a statistical study shows **bias** if it *systematically* favors certain outcomes.  Note: when asked to identify bias in the design of a statistical study, you are expected to identify the *direction* of the bias. |

**Example**:

b. **Voluntary Samples** - A voluntary response sample consists of people who choose themselves by responding to a general appeal. Voluntary response samples show bias because people with strong opinions (often in the same direction) are most likely to respond.

**Example**:

* Write-in and call-in opinion polls are almost sure to lead to a strong bias.
* Another problem is that people often times respond more than once.



**4. How to Sample Well**

A sample chosen by *chance* rules out favoritism by the sampler and self-selection by respondents. **Random sampling**, the use of chance to select a sample, is central to the principle of statistical sampling.

|  |
| --- |
| **Definition**: A **simple random sample (SRS)** of size *n* consists of *n* individuals from the population chosen in such a way that every set of *n* individuals has an equal chance to be the sample actually selected. |

**Example**:

* An SRS not only gives each individual an equal chance of being chosen but also gives every possible sample an equal chance of being chosen.
* Often times a Table of Random Digits or a Random Number Generator are used to choose SRSs.

**How to Choose an SRS Using Technology**

Step 1: **Label** – Give each member of the population a distinct numerical label from 1 to N.

Step 2: **Randomize** – Use a random number generator to obtain n different integers from 1 to N.

**How to Choose an SRS Using Table D**  
Step 1: **Label** - Give each member of the population a numerical label of the *same length*.

Step 2: **Table** - Read consecutive groups of digits of the appropriate length from Table D.

* The sample contains the individuals whose labels you find.
* Always use the shortest labels that will cover your population.
* Ignore any group of digits that was not used as a label or that duplicates a label already in the sample.
* Digits can be read in any order but it is recommended to read rows from left to right.

**Example**. The management company of a local mall plans to survey a random sample of 3 stores to determine the hours they would like to stay open during the holiday season.

a. Use a random number generator to select an SRS of size 3.

b. Use Table D at line 101 to select an SRS of size 3.

Aeropostale Just Sports  
All American Burger Mrs. Fields  
Arby’s Nike Factory Store  
Barnes & Noble Old Navy  
Carter’s for Kids Pac Sun  
Destination Tan Panda Express  
Famous Footwear Payless Shoes  
Forever 21 Star Jewelers  
GameStop Vitamin World  
Gymboree Zales Diamond Store  
Haggar



**5. Other Sampling Methods**

Unfortunately it is usually very difficult to actually obtain an SRS from the population of interest. It is often costly in time and money.

|  |
| --- |
| **Definition:** To select **stratified random sample**, first classify the population into groups of similar individuals, called **strata**. Then choose a separate SRS in each stratum and combine these SRSs to form the full sample. |

**Example**:

* Choose the strata based upon facts known before the sample is taken.
* If the individuals in each stratum are less varied than the population as a whole, a stratified random sample can produce better information about the population than an SRS of the same size.

|  |
| --- |
| **Definition**: To take a **cluster sample**, first divide the population into smaller groups. Ideally, these **clusters** should mirror the characteristics of the population. Then choose an SRS of the clusters. *All* individuals in the chosen clusters are included in the sample. |

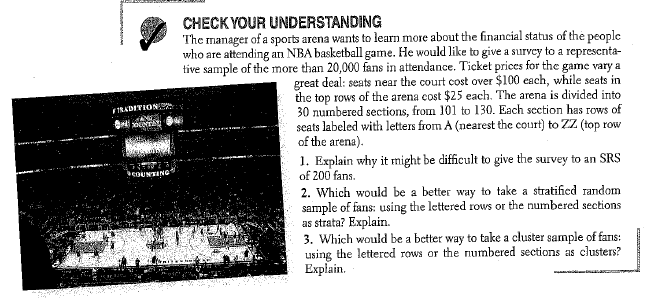
**Example:**

* Cluster samples are often used for practical reasons.
* They do not offer the statistical advantage of better information about the population that stratified samples do.

**Team Work** - With your team mates, read and discuss the *Sampling at a School Assembly* Example on pp. 221-222. Then list the advantages and disadvantages of each plan.

|  |  |  |
| --- | --- | --- |
| Plan | Advantages | Disadvantages |
| SRS |  |  |
| Stratified Random Sample |  |  |
| Cluster Sample |  |  |

**Multistage Samples** - Most large scale sample surveys are multistage samples that combine two or more sampling methods. A good example is on page 222.



HW: 1-11 odd, 17-25 odd.