# Chapter 4.1 Lecture Notes and Examples Part 1

**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**: A parent of an athlete is concerned about cuts to the athletic budget. She asks spectators at a game what they think. Why is this a biased method? What is the direction of the bias?

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

---

**CHECK YOUR UNDERSTANDING**

For each of the following situations, identify the sampling method used. Then explain how the sampling method could lead to bias.

1) A farmer brings a juice company several crates of oranges each week. A company inspector looks at 10 oranges from the top of each crate before deciding whether to buy all the oranges.

2) The ABC program *Nightline* once asked whether the United Nations should continue to have its headquarters in the U.S. Viewers were invited to call one telephone number to respond "Yes" and another for "No." There was a charge for calling either number. More than 186,000 callers responded, and 67% said "No

---

**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**: Put names of students on slips of paper in a "hat" and draw out 10 slips. Every sample of size 10 has equal likelihood of being chosen.

- 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. Explain how you would use technology to choose 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 | |

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

c. Explain how you would use a line of Table D to choose an SRS of size 3.

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

| Line | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 101 | 19223 | 95034 | 05756 | 28713 | 96409 | 12531 | 42544 | 82853 |
| 102 | 73676 | 47150 | 99400 | 01927 | 27754 | 42648 | 82425 | 36290 |
| 103 | 45467 | 71709 | 77558 | 00095 | 32863 | 29485 | 82226 | 90056 |

**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.

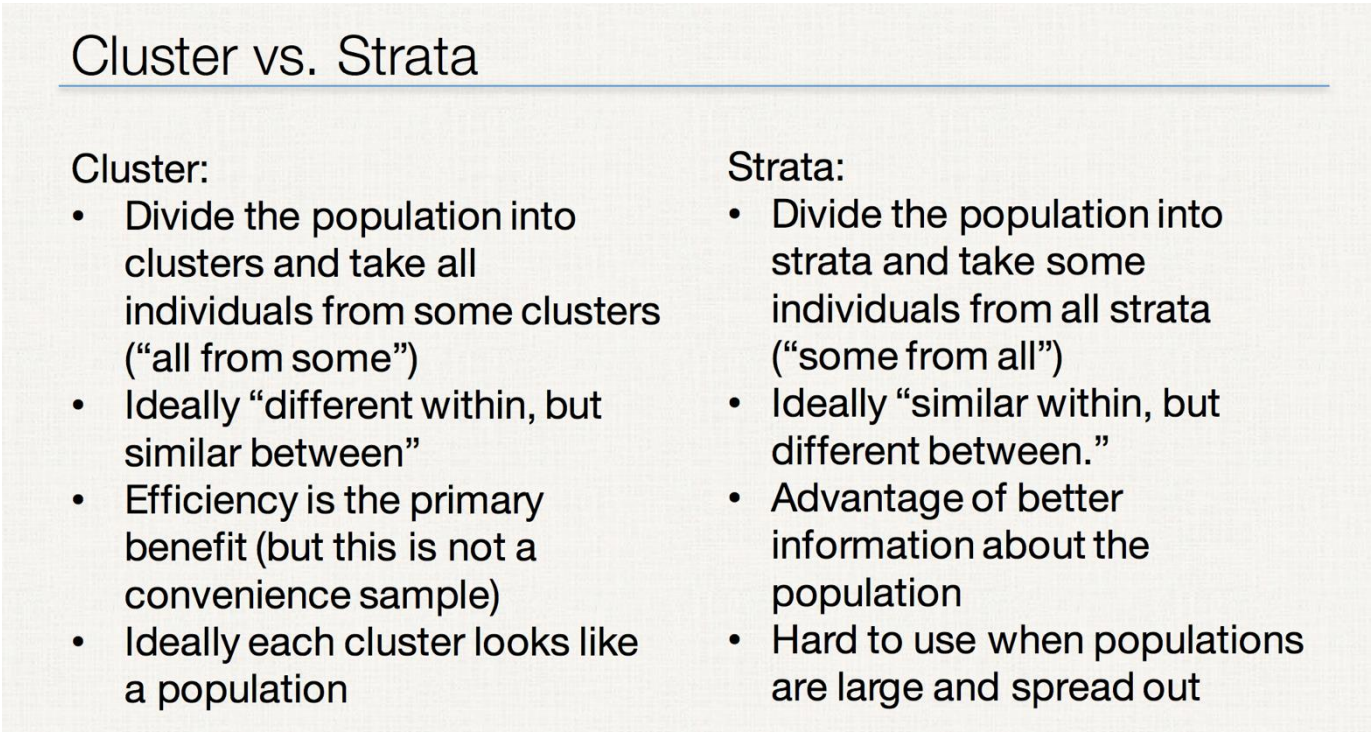
---

**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.

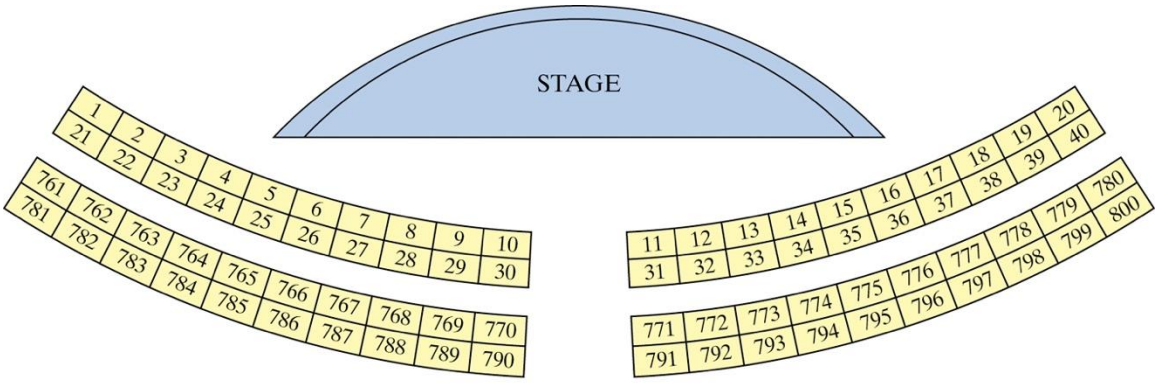
**Example:**

- 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.

---

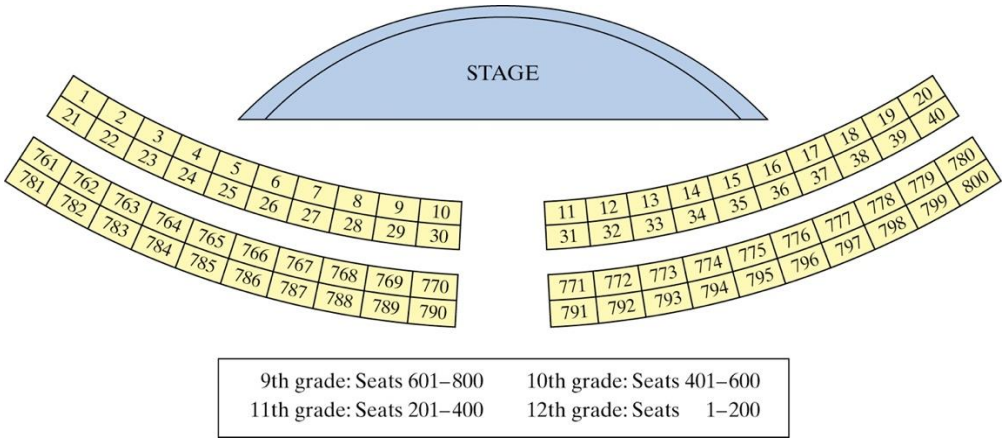


**Example 4: The student council wants to conduct a survey during the first five minutes of an all-school assembly in the auditorium about use of the school library. They would like to announce the results of the survey at the end of the assembly. The student council president asks your statistics class to help carry out the survey. There are 800 students present at the assembly. A map of the auditorium is shown below. Note that students are seated by grade level and that the seats are numbered from 1 to 800. Describe how you would each sampling method to select 80 students for the survey. Give advantages and disadvantages for each.**




| | |
|---|---|
| 9th grade: Seats 601–800 | 10th grade: Seats 401–600 |
| 11th grade: Seats 201–400 | 12th grade: Seats 1–200 |

**A) SRS**




**B) Stratified Random Sample**

**C) 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 219.

**CHECK YOUR UNDERSTANDING**

The manager of a sports arena wants to learn more about the financial status of the people who are attending an NBA basketball game. He would like to give a survey to representative sample of the more than 20,000 fans in attendance. Ticket prices for the game vary in a great deal: seats near the court cost over $100 each, while seats in the top rows of the arena cost $25 each. The arena is divided into 30 numbered sections, from 101 to 130. Each section has rows of seats labeled with letters from A (nearest to the court) to ZZ (top row of the arena).

a) Explain why it might be difficult to give the survey to an SRS of 200 fans.

b) Which would be a better way to take a stratified random sample of fans: using the lettered rows or the numbered sections as strata? Explain.

c) Which would be a better way to take a cluster sample of fans: using the lettered rows or the numbered sections as clusters? Explain.

HW 4.1 Part A: 1,3,5,7,9,11 page 226
HW 4.1 Part B: 17, 19, 23, 25 page 227 - 228