

## Population and Sampling

- The **entire set of individuals or objects in which we are interested in** is the **population**.
- **Subset of population is a sample** and the **number of objects in a sample** is called a **sample size**.
- The **process of selecting a sample that is representative of the total population** is called **sampling**.

We need to ask ourselves:

- 1) How should the sample be collected?
- 2) How large is our sample size?
- 3) How reliable are our conclusions?

Example: Senior citizens are 20% of Littleton’s voting population. In a poll of 100 citizens, half of whom were senior citizens, 30 senior citizens voted yes and 20 non-senior citizens voted yes. What is the population, sample and sample size?

**Sampling Methods:** The goal in sampling is to obtain individuals that will participate in a study so that accurate information about the population can be obtained. We want the sample to provide as much information as possible, but each additional piece of information has a price. So the question becomes “How can a researcher obtain accurate information about the population through the sample while minimizing the costs in terms of money, time, personnel, and so on?”

When doing a survey, it usually is not practical to get the opinion of every member of a population. You can get a fairly accurate picture of the opinion of a population by surveying a sample of the population. A sample is a smaller group that represents the whole population.

There are five good ways to choose a sample: **Simple random sampling, stratified sampling, systematic sampling, cluster sampling and convenience sampling.**

### 1. Simple random sampling (often abbreviated as random sampling):

- Every possible sample has an equally likely chance of occurring.

Example: Sophia has 4 tickets to a concert. Six of her friends, Yolando, Michael, Kevin, Terri, Annie, and Casey, have all expressed an interest in going to the concert. Sophia decides to *randomly* select three of the six. (There are 20 subsets of size 3, YMK, YMT, YMA, YMC, YKT, YKA, MKT, ..., TAC)

How do we actually select the individuals in a simple random sample? Simple random sample is just like drawing the names out of a hat. We could write the six names on different sheets of paper and then select three from the hat. It’s that easy! But, often the size of the population is so large that performing simple random sampling in this fashion is not practical. Each person could be assigned a number and then you can use a calculator or computer to randomly select the number you need in your sample. Each person is equally likely to be chosen.

### **THINK ABOUT:**

No particular way of choosing

## 2. Stratified sampling:

- A stratified sample is obtained by separating the population into non-overlapping groups called strata and then obtaining a simple random sample from each stratum (or group).

The individuals in each stratum should be similar in some way. An advantage of stratified sampling over simple random sampling is that it may allow fewer individuals to be surveyed while obtaining the same (or more) information. This occurs because individuals within each subgroup have similar characteristics, so opinions within the group do not vary much from one person to the next.

**In other words, a stratified sample is a simple random sample of different divisions of the population.**

### THINK ABOUT:

Grouping → taking small sample from each group

## 3. Systematic Sampling:

- A systematic sample is obtained by selecting every  $n$ th individual from the population.

The first person selected is a random number between 1 and  $n$ , and then survey every  $n$ th person after that random number. For example, you want to survey every 8th person. Randomly choose a number between 1 and 8, such as 5. This means you survey the 5th,  $5+8 = 13$ th,  $13+8 = 21$ st,  $21+8 = 29$ th and so on.

**In other words, systematic sampling is like selecting every 5th person out of a line.**

### THINK ABOUT:

Every  $n$ th #, person, etc.

## 4. Cluster Sampling:

- A cluster sample is obtained by selecting all individuals within a randomly selected collection or group of individuals.

For example, a quality control engineer wants to verify that a certain machine is filling bottles with 16 ounces of liquid detergent. To obtain a sample of bottles from the machine, the engineer could use systematic sampling by sampling every  $n$ th bottle from the machine; however, it would be time consuming waiting next to the filling machine for the bottles to come off the line. Instead, suppose that as the bottles come off the line, they are placed into cartons of 12 bottles each. Then the engineer could randomly select a few cartons and measure the contents of all 12 bottles. This would be cluster sampling. It is good in this situation because it speeds up the data collection process.

**In other words, imagine a mall parking lot. Each subsection of the lot could be a cluster - Section F-4, for example.**

### THINK ABOUT:

Grouping → sampling all in the group

## 5. Convenience Sampling: (Sometimes called “Self-selection”)

- A convenience sample is a sample in which the individuals are easily obtained.

The most popular convenience sample is one in which the individuals in the sample are self-selected (the individuals themselves decide to participate in a survey).

Examples: 1) a radio DJ asks his/her listeners to phone the station to submit their opinions

2) Dateline will present a story on a certain topic and ask its viewers to “tell us what you think” by going on-line to complete a questionnaire.

**CAUTION:** Convenience sampling is generally not a good design because the individuals who decide to be in the sample generally have strong opinions about the topic. A typical individual in the population will not bother phoning or logging on to their computer to complete a survey. Therefore, convenience sampling has limitations or is biased.

### THINK ABOUT:

Groups include only those who are listening, watching, etc.

### Errors in sampling that cause bias:

- 1) nonresponse of individuals selected to be in the survey
- 2) inaccurate responses
- 3) poorly worded questions (“Do you oppose the reduction of estate taxes?” would be better if written as “Do you favor or oppose the reduction of estate taxes?”) The question should be balanced.
- 4) bias in the selection of the individuals

**Example:** A news program reports on a proposed school dress code. The purpose of the program is to find out what percent of the population in its viewing area favors the dress code. Identify the type of sampling and any bias in each sampling method.

- a) Viewers are invited to call the program and express their preferences.

Convenience

- b) A reporter interviews people on the street near the local high school.

Simple Random

- c) During the program, 300 people are selected at random from the viewing area. Then each person is contacted.

Stratified

## Unit 6 Day 5 HW: Population and Sampling

1. An important part of employee compensation is a benefits package, which might include health insurance, life insurance, child care, vacation days, retirement plan, parental leave, bonuses, etc. Suppose you want to conduct a survey of benefits packages available in private businesses in Hawaii. You want a sample size of 100. Some sampling techniques are described below. **Categorize each technique as a simple random sample, stratified sample, systematic sample, cluster sample or convenience sample.**
  - a. Assign each business in the Island Business Directory a number, and then use a random number table to select the businesses to be included in the sample.
  - b. Use postal ZIP codes to divide the state into regions. Pick a random sample of 10 ZIP code areas and then include all the businesses in each selected ZIP code area.
  - c. Send a team of five research assistants to Bishop Street in downtown Honolulu. Let each assistant select a block or building and interview an employee from each business found. Each researcher can have the rest of the day off after getting responses from 20 different businesses.
  - d. Use the Island Business Directory. Number all the businesses. Select a starting place at random, then use every 50th business listed until you have 100 businesses.
  - e. Group the businesses according to type: medical, shipping, retail, manufacturing, financial, construction, restaurant, hotel, tourism, other. Then select a random sample of 10 businesses from each business type.
2. Modern Managed Hospitals (MMH) is a nation for-profit chain of hospitals. Management wants to survey patients discharged this past year to obtain patient satisfaction profiles. They wish to use a sample of such patients. **Categorize each technique as a simple random sample, stratified sample, systematic sample, cluster sample or convenience sample.**
  - a. Obtain a list of patients discharged from all MMH facilities. Divide the patients according to length of hospital stay (2 days or less, 3 – 7 days, 8 – 14 days, more than 14 days). Draw simple random samples from each group.
  - b. Obtain a list of patients discharged from all MMH facilities. Number these patients, and then use a random number table to obtain the sample.
  - c. Randomly select some MMH facilities from each of five geographic regions, and then include all the patients on the discharge lists of the selected hospitals.
  - d. At the beginning of the year, instruct each MMH facility to survey every 500th patient discharged.
  - e. Instruct each MMH facility to survey 10 discharged patients this week and send in the results.

**Determine the population and sample, if possible, then determine the sampling used. Are there any errors in the sampling that may cause bias? Explain.**

3. An interviewer in a mall is told to survey every 5<sup>th</sup> shopper, starting with the 2<sup>nd</sup>.
  
4. A researcher randomly selects 5 of the 70 hospitals in a metropolitan area and then surveys all of the surgical doctors in each hospital.
  
5. A researcher segments the population of car owners into four groups: Ford, General Motors, Chrysler, and foreign. She obtains a random sample from each group and conducts a survey.
  
6. A list of students in elementary statistics is obtained in which the individuals are numbered 1 to 540. A professor randomly selects 30 of the students.
  
7. In order to estimate the percentage of defects in a recent manufacturing batch, a quality control manager at Intel selects every 8<sup>th</sup> chip that comes off the assembly line starting with the 3<sup>rd</sup> chip, until she obtains a sample of 140 chips.
  
8. In order to determine the average IQ of ninth-grade students, a school psychologist obtains a list of all high schools in the local school system. She randomly selects five of these schools and administers an IQ test to all 9<sup>th</sup> grade students at the selected schools.
  
9. In an effort to determine customer satisfaction, United Airlines randomly selects 50 flights during a certain week and surveys all passengers on the flights.
  
10. In an effort to identify whether an advertising campaign has been effective, a marketing firm conducts a nation-wide poll by randomly selecting individuals from a list of known users of the product.

11. A school official divides the student population into four classes: freshman, sophomore, junior, senior. The official takes a random sample from each class and asks the members' opinions regarding student services.
  
12. A survey regarding download time on a certain web site is administered on the internet by a market research firm to anyone who would like to take it.
  
13. A lobby group has a list of the 100 senators of the United States. In order to determine the Senate's position regarding farm subsidies, they decide to talk with every seventh senator on the list starting with the third.
  
14. A manufacturing company would like to determine the approximate market share of a certain product. A representative of the company is asked to stand in front of a certain grocery store and ask the first 100 people who go into the store whether they use their product.