

8.2: Statistical Inferences Notes

Suppose someone wanted to do a study about all sophomores, juniors, and seniors currently in high school in the United States. They might want to know the number of students who work after school, own a cell phone, or participate in extracurricular activities. Through statistical methods, we can gather and analyze information from a smaller sample population which allows us to make inferences about the much larger entire population. In Jordan School District there are currently 11,376 sophomores, juniors, and seniors in high school. It would be extremely costly and time consuming to interview every student; however, by taking a random sample of students we can calculate statistics which will allow us to draw conclusions about all 11,376 students in high school in Jordan School District.

Vocabulary

A **population** consists of all people or items which we wish to describe or draw conclusions about. A **sample** is a small group of people or items taken from the larger population.

The population characteristic that we are interested in is called the **parameter of interest**. In the case of our high school example, the parameter of interest could be the number of students who work after school.

If it were possible to gather data from an entire population, then the parameter of interest would be **the population parameter**.

It is often difficult to gather data from an entire population so we use **statistics**, or data that we gather from a sample of the population, to make an **inference** or conclusion about the parameter of interest for the population.

In obtaining a sample from a population it is important to use **random sampling** to ensure the sample is representative of the population. Each individual is chosen entirely by chance and each member of the population must have an equal chance of being included in the sample.

Example 1:

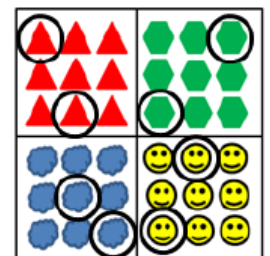
The Utah State Legislature wants to know what percentage of teen drivers text while they drive. They decide to survey 250 randomly selected teen drivers across the state.

Identify a) the population, b) the sample population, and c) the parameter of interest.

Obtaining a random sample is not as simple as you might think. In fact, there are a few different methods for sampling. Some of the methods can be biased. A **bias** occurs when part of the population is overrepresented or underrepresented. For example, if you wanted to know how many students support the school's athletic programs, you wouldn't interview only the cheerleaders or football players, because they regularly attend athletic events and would be overrepresented in the study.

Sampling Methods:

- In a **simple random sample** every member of the population has an equal chance of being selected to be part of the sample group. Drawing names from a hat is an example of this type of sampling. Another example would be assigning every member of the population a number and then using a random number table or generating random numbers through technology to randomly select members. The key is that you must have a list of all the members of the population.
- In a **systematic sample** it is assumed that the entire population is naturally organized in a sequential order. Using a random number generator, you select a starting point and then select every *n*th member to be part of the sample. For example, homes in a neighborhood are already in an order. You could randomly select a starting point and then select every third home.
- In a **stratified sample**, members of the population that share the same characteristic are grouped together. Then, members of that subgroup are randomly selected to make up the sample group. Each member of the subgroup has an equal chance of being selected. There are times when the subgroups are not equal in size. When this happens, members are chosen in proportion to their actual percentages in the overall population. For example, if you wanted to study all high school students who are involved in extracurricular activities you would probably want to divide them into their particular extracurricular activity and then select randomly from those groups so that each extracurricular activity is represented in the sample. The football team would have more members than the basketball team so you would select more football players than basketball players to participate.



- In a **cluster sample** the population is divided into smaller groups that are representative of the entire population and then groups are randomly selected. For example, if you wanted to make inferences about your entire school, you could randomly select 1st periods to survey.



- In a **convenience sample** members are randomly selected from a population that is readily available. For example if you wanted to ask shoppers what they think of a local store, you would survey every 5th person who exits the store on a given day. This method of sampling has a bias because people who like to shop at this particular store are more likely to be at the store that day.
- In a **volunteer sample (also called self-selected sample)** members of the population self-select to be included in the sample. Filling out a survey and returning it is an example of a volunteer sample. This is prone to bias because generally people who respond have strong opinions about the topic while others who are more neutral may not respond at all. An example of a volunteer sample is when you buy a pair of shoes at your favorite shoe store and the cashier asks you to complete an online survey about your experience that day. You decide whether or not you want to complete the survey.

Example 2:

The school newspaper wants to know the percentage of students who drive to school each day. For each method described below, determine what type of sampling method it is and justify whether or not the method is biased.

- a) The newspaper staff posts signs all over the school asking students to take a short survey online.
- b) The newspaper staff interviews every 5th person who walks into the school cafeteria.
- c) The newspaper staff randomly selects 20 fifth period classes to survey.

Example 3:

You want to know if students at your school prefer fast food or sit-down restaurants. What would your survey question look like to eliminate any bias? Explain the sampling method you would use and why?