SM3 HW 14.1 Sampling and Collecting Data for Statistical Inference

OBJECTIVES:

- Understand that statistics allows inferences to be made about population parameters based on a random sample from that population.
- Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
- Evaluate reports based on data.

Suppose someone wanted to do a study about all sophomores, juniors, and seniors currently in high school in the United States. They want to know the number of high school students who work after school all across the country. 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 Alpine School District there are currently 14,376 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 500 students from the schools in Alpine School District, we can calculate statistics which will allow us to draw conclusions about all 14,376 students in high school in Alpine 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 representative group of people or items taken from the larger population.
- Proper selection of a sample is important to avoid **bias** which 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 students on a team, because they regularly attend athletic events and would be overrepresented in the study.
- The population characteristic that we are interested in learning about is called the **parameter,** sometimes called the **parameter of interest**. In the case of our high school example, the parameter of interest would be the number of students who work after school. The only way to find the true population parameter is to collect data from every member of the population and calculate the measure we are interested in. This is usually impractical or impossible to execute.
- It is often too difficult to gather data from an entire population so we use **statistics**, or data that we gather from a representative **sample** of the population, to make a conclusion about the parameter of interest for the population, this conclusion is called an **inference**.
- In obtaining a sample from a population it is important to use **random sampling** to ensure the sample is representative of the population. Random sampling is a technique where a group is selected from the population using some element of chance and each member of the population must have an equal chance of being included in the sample.
- A **survey** every member in the sample is asked to respond to a set of questions.
- **Experiments** require that a treatment is induced on the sample so that their response to that treatment can be measured. Often this includes at least two groups. One group receives the trial treatment, while the other, sometimes called the control group, does not receive the treatment or receives a fake treatment (placebo). At the end of an allotted period of time, the two groups are compared to determine if the treatment had an effect.
- **Observational studies** require you to observe outcomes without interacting with any members of the sample or inducing a treatment usually due to ethical issues or impractical circumstances.

EXAMPLE: In our example above about finding out how many high school sophomores, juniors and seniors work after school, what is the population?
What is the sample?
What is the parameter of interest?
If we wanted to make inferences about about all high school aged students across the country, what may be some concerns about our sample?

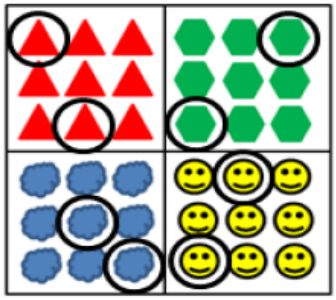
SAMPLING METHODS:

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 with the name of every person in the population in it 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. This method is the LEAST BIASED.

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 nth member to be part of the sample. For example, names in a phone book are already in alphabetical order. You could randomly select a starting point (the 73rd name for example) and then select every one-hundredth name until you reach the end of the phonebook. This method is generally unbiased.

Stratified Sample: members of the population that share the same characteristic are grouped together (called strata). Then a simple random sample of the members from that subgroup is selected. All of the samples from the different groups are combined together to make up the sample group. Each member of the subgroup has an equal chance of being selected. For example, we want to determine the effectiveness of a new pain medication and we think that the medication may act differently in men than it does in women. We divide the population into men and women (the strata) and then take a simple random sample of 50 men and a simple random sample of 50 women to make a stratified random sample of 100 people. This method is generally unbiased.

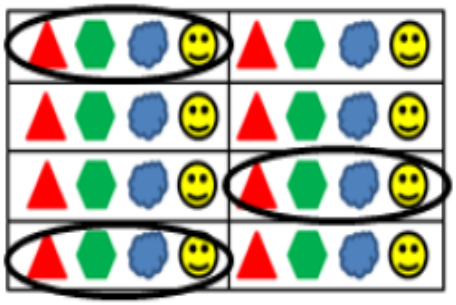
Each type of shape is one of the strata, within each strata a random sample of 2 is chosen and the 8 resulting selections make up the sample.



Cluster Sample the population is naturally divided into smaller groups that are representative of the entire population and then the entire groups are randomly selected. For example, if you wanted to make inferences about your entire school, you could randomly select 15 1st period classes to survey. Each classroom would be a cluster of students, randomly choosing 15 classes and including every student in each class makes up your sample. This method can be biased if the groups chosen are not representative of the population of interest. Generally it is considered an unbiased sampling method.

***This illustration to the right is a little misleading because it shows all the clusters (groups) as being exactly the same, this is most often not the case and is not necessary to do a cluster sample.

The population is separated into groups and then a random selection of the groups is made. The entire group is then included in the sample



Convenience Sample members are randomly selected from a population that is readily available. This method is susceptible to bias. 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.

Voluntary Response sample: members of the population self-select to be included in the 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 be motivated to respond at all. An example of a volunteer sample is when a radio station broadcasts an invitation to vote in a poll about gun control, only those that feel strongly either for or against the issue are inclined to respond.

EXAMPLE: The school newspaper wants to know the percentage of students who drive to school each day. 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 fifth person who walks into the school cafeteria.

c. The newspaper staff randomly selects 20 fifth periods to survey.

DATA COLLECTION METHODS:

Sample Surveys

The purpose of a sample survey is to gather information about the sample by means of a survey. There are several advantages to using a survey. Surveys are inexpensive and can collect a large amount of data representative of the population. They can be done in a variety of forms and about a variety of topics. Surveys also have the ability to focus only on the necessary information. However, surveys are flawed by non-responders since a survey is generally voluntary; people have the option not to participate. Additionally, people in a survey know that they are being studied and they may not be as honest in their responses as they would be if they were not being studied. Surveys are also open to interpretation and bias. Surveys can be written in a way that biases the responders. Also questions can be interpreted differently than intended by those responding to the survey. Surveys can be administered with to randomized samples, such as simple random sampling, cluster sampling, etc. all of which would ensure that the sample is random and representative of the overall population.

Experiments

The purpose of an experiment is to assign a treatment, using control over some of the conditions in order to gather data about the treatment's effectiveness. An experiment is the only way to establish causation. When an experiment is designed, all of the variables are controlled. This allows the experimenter to demonstrate that a change in one variable causes the change in another variable. There are drawbacks to experiments. They can be very expensive and time consuming. Ethics may be questioned especially if animals or people are used in the experiment. Experiments must not intentionally harm any of the subjects. The attitude and behavior of those conducting the experiment can also affect the results. It is imperative that randomization is used when assigning subjects to their treatment groups. Each group needs to be representative of the overall population.

Observational Studies

The purpose of an observational study is to observe subjects in their natural environment without their knowledge and without assigning treatments to the subjects. There are some advantages to using an observational study. It is simple and inexpensive to conduct. It provides deeper and richer information than a survey because the observer is seeing behavior firsthand and is able to observe the process not just the result. There are also some disadvantages. The results cannot prove causation nor can they be applied to the general population. It is only representative of those being studied. The results are subjective and open to interpretation by the observer. There may also be a question of ethics, especially if people are involved. People have a right to privacy and the observational study must not infringe upon the rights and expectations of people. If you are doing the study in the present, you can randomize the individuals involved.

EXAMPLE:

Which type of study method is described in each situation? Should the sample statistics be used to make a general conclusion about the population?

a. Researchers randomly choose two groups from 20 volunteers. Over a period of 6 weeks, one goup works on a computer for an hour right before going to sleep, and the other does not. Volunteers wear monitoring devices while sleeping, and researchers record their quality of sleep.

b. Students in an elementary class observe the growth of some newly hatched chickens.

c. Market researchers want to know if people like the new store at the local mall. They ask every fourth person who enters the mall if they like the new store.

12.1 EXERCISES NAME_____ PER._____

For each of the scenarios below, identify the **population**, the **sample** and the **parameter of interest**.

1. A team of biologists wants to know the average weight of fish in a lake. They decide to drop a net and measure all fish caught in three different locations in the lake.

Population: Sample: Parameter of Interest:

2. An AP Government class wants to know the percentage of eligible voters in the state of Utah who voted in the most recent election. There are 1,938,249 people in Utah who are 18 and older. The class randomly looks at 15 state house districts and discovers that 50.5% of the eligible voters actually voted.

Population: Sample: Parameter of Interest:

3. A local radio station has added an additional radio personality and is trying to determine what type of music to play during this person's air time. This time slot is geared towards teenage listeners. The station has decided to survey 300 randomly selected students from the ages of 13 to 19.

Population: Sample: Parameter of Interest:

Identify which of the six sampling methods (simple random, systematic, stratified, cluster, convenience or voluntary) were used in each of the following examples and then tell whether the method is biased or unbiased.

4. You are in charge of school activities. You want to know what activities students would prefer to participate in during the school year. You decide to put the name of each student in the school into a big bowl. You draw 100 names and ask those students to respond to a survey about the activities they prefer.

Method: Biased | Unbiased

5. You are in charge of school activities. You want to know what activities students would prefer to participate in during the school year. You assign each student in the school a number. You randomly select a starting number among the first 10 numbers and then select every tenth student in the list from that point forward.

Method: Biased | Unbiased

6. You are in charge of school activities. You want to know what activities students would prefer to participate in during the school year. You use the rolls from each homeroom class. You put the all the names from one class into the bowl and draw two names from the class. You go through each homeroom class, drawing 2 names from each class. You ask those students to respond to a survey about the activities they prefer.

Method: Biased | Unbiased

7. You are in charge of school activities. You want to know what activities students would prefer to participate in during the school year. You get the list of all the homeroom classes and randomly select 5 classes. You go to each of the classes selected and survey all the students in that class.

Method: Biased | Unbiased

8. You are in charge of school activities. You want to know what activities students would prefer to participate in during the school year. You stand in the cafeteria during your lunch break and ask students in they would be willing to participate in your survey as they walk by.

Method: Biased | Unbiased

9. You are interested in finding out if the crime rate in your city has changed in the past year. You decide to have the evening news broadcast an invitation to respond to a poll about crime.

Method: Biased | Unbiased

10. You want to know the average number of hours that high school seniors spend playing video games in your state. You randomly select 20 high schools in the state and then ask all the seniors at each of the 20 high schools about their video game habits.

Method: Biased | Unbiased

11. An auto analyst is conducting a satisfaction survey, sampling from a list of 10,000 new car buyers. The list includes 2,500 Ford buyers, 2,500 GM buyers, 2,500 Honda buyers, and 2,500 Toyota buyers. The analyst selects a sample of 400 car buyers, by randomly sampling 100 buyers of each brand.

Method: Biased | Unbiased

12. A shopping mall management company would like to know the average amount that shoppers in the mall spend during their visit. They post two survey takers near one of the exits who ask shoppers to tell them what they spent as they leave the mall.

Method: Biased | Unbiased

13. A restaurant owner wants to find out the average number of dishes ordered at each table served on Friday evenings, their busiest time. She decides to collect and analyze every fifth receipt of the night, starting at 6:00 p.m.

Method: Biased | Unbiased

14. In order to determine the average composite score on the most recent ACT exam, students were divided into groups based on whether they were enrolled in remedial, regular, or honors language arts. Individual scores were randomly selected from each group.

Method: Biased | Unbiased

Imagine that you want to know whether a new diet plan is effective in helping people lose weight. To collect data about this diet plan you might choose to conduct a survey, carry out an experiment or do an observational study to determine this.
If you used a survey, you could simply ask people that had tried the diet plan in they lost weight.

If you used an observational study, you might monitor volunteers that try the diet plan and measure how much weight they lost.

If you used an experiment, you might randomly assign participants to two groups. One group (the control group) eats as they normally would and the other group (the experimental group) eats according to the diet plan. At the end of two months, the two groups are compared to see the average weight gain or loss in each group.

Based on these three examples:

15. What are the possible advantages and disadvantages of conducting a survey? (give at least one of each)

16. What are some possible advantages and disadvantages of the experiment? (give at least one of each)

17. What are some possible advantages and disadvantages of the observational study? (give at least one of each)

Identify whether the following are examples of a survey(S), experiment(E) or observational study(O).

S E O 18. To determine whether drinking orange juice prevents colds, researchers randomly assigned participants to a group that drank no orange juice or a group that drank two glasses of orange juice a day. They measured the number of colds that each group had over the course of the year and compared the results of the two groups.

S E O 19. To determine whether exercise reduces the number of headaches, researchers randomly selected a group of participants and recorded the number of hours each participant exercised and the number of headaches each participant experienced.

S E O 20. To determine the effectiveness of a new advertising campaign, a restaurant asked every tenth customer if they had seen the advertisement, and if it had influenced their decision to visit the restaurant.

S E O 21. To determine if a new drug is an effective treatment for the flu, researchers randomly selected two groups of people that had the flu. One group was given a placebo (a sugar pill that has no physical effect) and one group was given the new drug. Researchers measured the number of days that participants experienced flu symptoms and compared the two groups to see if they were different.

S E O 22. To determine if higher speed limits cause more traffic fatalities, researchers compared the number of traffic deaths on randomly selected stretches of highway with 65 mph speed limits to the number of traffic deaths on an equal number of randomly selected stretches of highway with 75 mph speed limits.

23. Describe how you might select a simple random sample of 50 students from Orem High and use a survey to investigate which soft drink people prefer: Soda A or Soda B.

24. Describe how you might select a convenience sample of 50 students at Orem High and use an observational study to investigate which soft drink people prefer: Soda A or Soda B.

25. Describe how you might select a cluster sample of 3 clusters of students at Orem High and use an experiment to investigate which soft drink people prefer: Soda A or Soda B.

26. Describe the method you would use to determine if excessive texting causes bad grades. Explain why you chose that method and what conclusions could be drawn from the study.