CARNEGIE MATH PATHWAYS WestEd

1.4

Random Sampling & Experiments

INTRODUCTION

In Lesson 1.3, we learned that in order to generalize results from a sample to the population, the sample must be **representative** of the population. A sample is representative if it does not differ in any important way from the population. This lesson explores sampling techniques that yield representative samples.

1 Suppose our college is thinking of ways to raise money. Many students like parking spaces close to their classes. The administration is thinking of selling reserved parking spaces for $100. The college wants to know the percentage of students who would support this fee.

One way to find the percentage of students that support this fee would be to conduct a **census**. A census is a **survey** of an entire population. The college would ask every student on campus if she or he would support the fee.

Is this a reasonable plan? Why do you think so?

2 Read the following ways to sample students at our college. For each method:

- Tell whether the method would produce a sample that represents the student population.
- Explain why or why not.

A Choose four 8:00 a.m. classes at random. **Survey** all the students in each class.

B Put a **poll** on the front page of the college website. A poll is an opinion survey. Use the students who answer the question as the sample.

C Talk to students as they enter the Student Center.

3 None of the sampling methods above will produce a sample that is representative of the college's population of students. Suggest a better method to gather a representative sample.

NEXT STEPS

When we sample, our goal is for every population member to have the same chance of being selected. We want to avoid **selection bias**. This is the term used to describe a situation in which a sample differs from the population in some systematic way, so that some individuals are more likely to be selected than others. For example, imagine that we are surveying students and we choose only students entering the library. Not all students go to the library. The students in the sample may not be representative of the population of all students.

One good way to avoid *selection bias* is to select a **simple random sample.** In a simple random sample, all samples (of a given size) have the same chance of being chosen.

4 **Biased samples** result when sampling methods tends to leave out certain types of population members. The three previous sampling methods all produced biased samples.
There are different types of biased samples.

A One type of biased sample is a **voluntary response sample**. Good samples are chosen by researchers. In a voluntary response sample, the participants are self-selected. In other words, each participant *chooses* to participate.

Which sample from Question 2 above is a voluntary response sample?

B Another biased sample is a **convenience sample**. Convenience sampling does not use random selection. It involves using an easily available or "convenient" group to form a sample. Many samples have convenience sampling problems.

Which sample from Question 2 above is the best example of a convenience sample?

5 Suppose our college has 13,000 students. The college has the names and email addresses for all students in its database.

A Suggest a way that the administration could choose a simple random sample of 150 students to survey about the parking fee proposal.

B After the administration has chosen a sample, how could they actually conduct the survey?

C The administrator is told that the college can dedicate additional time and money to conducting this study. Would it make sense for the administrator to double the sample size and survey 300 students? Explain.

NEXT STEPS

Conducting an Experiment

Today we are going to conduct an experiment. In past lessons we discussed the four step process in a statistical study. The first step in this process was to *ask a question that can be answered by collecting data*. In today's experiment we ask the following question: "Does listening to music make a difference in how much students remember while studying?"

To prepare for our experiment we first need to write a **hypothesis**. A **hypothesis** is a **conjecture**—an informed guess based on some information you have about the topic you want to study. For example, an informed guess about music and studying may be based on the fact that you and your friends listen to music when you study and it seems to help you remember things better.

6 State a hypothesis about the effect that listening to music has on the amount of material a person remembers when studying. Do you think a person remembers *more*, *less*, or about *the same* as a student who studies in silence? Explain why you think this?

Your answer to Question 6 is a hypothesis. This is your best guess of the answer to your research question. It is "informed" by your prior knowledge and experiences. We can test a hypothesis by collecting and analyzing data.

7 Design an experiment to test your hypothesis. Explain how you will collect and analyze data to make a decision about your hypothesis. Clearly describe the explanatory variable and the response variable.

To test your hypothesis, we will now gather data. Your instructor will provide you with a short history lesson. Some students will read this lesson while listening to music and the rest of you will read in silence. When finished, you will take a short quiz to test how well you remember what you read.

TRY THESE

Your instructor will now help you grade your quiz. Once this is done, answer the question below.

8 In your group, did the person with the highest score on the quiz study in silence or study while listening to music? Can you make a decision about your hypothesis solely based on this? Explain.

We should not draw conclusions based on the person who got the highest score. That result represents only one person. One result is not enough evidence. This is an example of **anecdotal evidence**. An anecdote, or story, may be persuasive, but it only gives information about one individual.

Another example of anecdotal evidence would be observing a person who smokes a pack of cigarettes every day and does not get cancer. One person does not provide sufficient evidence. It would not be valid to conclude that people who smoke a pack of cigarettes every day do not get cancer.

9 We want to see if there is a difference between the students who studied while listening to music and the students who studied in silence. Compute the average score of the students who studied in silence, as well as the average score of students who listened to music. Describe any difference between the scores of those who studied with music and those who studied in silence.

10 Think about what you know about a well-designed study. What are two possible reasons you might observe a difference between the students who studied while listening to music and the students who studied in silence? Write down the two possible reasons.

Significance

When a difference between groups of a variable is so large that it is *unlikely* to be caused by chance, we say it is **significant.** In this case a significant difference would mean that we believe that listening to music was the reason for the difference in quiz score averages between the two groups. We do not yet have the tools to determine if the difference between averages is due to music listening or chance variation.

For now, we need to be careful. We shouldn't draw final conclusions yet. We need to rule out other possible causes. Think about other factors, or reasons, that could explain differences in quiz scores. For example, if some students took American History recently, they might get higher scores on the quiz.

NEXT STEPS

Direct Control, Random Assignment, Blinding, and the Placebo Effect

The goal of an experiment is to determine the effect of changing a treatment on the response variable. To do this, we want to rule out other possible explanations for differences in responses to treatments. Two common strategies to help with this are **direct control** and **random assignment.**

Direct control means that if you notice that there are other variables, *besides the explanatory variable*, that might affect the response, you try to manage those variables. **Random assignment** helps us create groups that are similar.

TRY THESE

11 Consider the following description of an experiment.[1] In this experiment, the researchers wanted to investigate whether the way people dry their hands after washing them has an effect on how clean their hands are.

> An experiment was conducted to compare bacteria reduction for three different hand drying methods: using paper towels, using a hot air dryer, and evaporation. In this experiment the participants handled uncooked chicken for 45 seconds. The participants then washed their hands with a single squirt of soap for 60 seconds and finally used one of the three hand drying methods. After participants dried their hands, the researchers measured the bacteria count on the participants' hands.

A One variable that might affect the response is the length of time that people handled the raw chicken. The researchers controlled this by having everyone handle the chicken for the same amount of time—45 seconds. Two other variables were controlled in this experiment. What are they?

B A **control group** is a group in an experiment that does not get a treatment. Including a control group in an experiment provides a basis for making comparisons. Is there a control group in this experiment? Explain.

[1]A.M. Snelling et al., "Comparative Evaluation of the Hygienic Efficacy of an Ultra-Rapid Hand Dryer vs. Conventional Warm Air Hand Dryers," *Journal of Applied Microbiology* 110, No. 1 (January 2011): 19-26. Details of this experiment were also reported in the online newsletter of the journal *Infectious Disease News* (September 22, 2010).

Blinding

Sometimes people already have ideas about whether the treatments in an experiment will be effective. These beliefs might influence the response. When participants in an experiment do not know the type of treatment they are receiving, they are said to be **blinded**. Blinding participants is a way to prevent prior beliefs about the treatment from influencing their response.

12 A shoe company wants to compare two different products for making hiking boots waterproof. Researchers create an experiment to compare two different products. In the experiment, researchers randomly give participants hiking boots. The hiking boots were waterproofed using one of the two waterproofing products. Participants then wear the boots on a hike along a mountain stream. They have to cross the stream and walk in the water in several places. After the hike, the boots are left to dry. Then the boots are sent to a lab where a technician evaluates the boots for water damage.

Do you recommend blinding only the participants, only the lab technician, or both the participants and lab technician? Explain your answer.

Placebo Effect

In experiments that use human participants, use of a control group may not be enough to establish whether a treatment really has an effect. Studies have shown that people sometimes respond in a positive way to treatments that have no active ingredients. These non-treatments that have no active ingredients may be colored water or sugar pills, and are called **placebos**. People often report that such non-treatments relieve pain or reduce other symptoms such as dizziness. Thinking that a non-treatment has helped the pain or made you feel better is an example of what is called the **placebo effect**.

Because of the placebo effect, experiments often include a control group that receives no treatment and another control group that receives a placebo. The placebo is identical in appearance (and taste, etc.) to what the people in other experimental groups receive. In such situations, participants should be blinded—we do not want people to know that they are receiving a placebo.

13 A researcher included a control group (with no treatment) and a placebo group in her experiment. How does including both groups allow her to decide when a placebo effect occurs?

14 How does including a placebo group in her experiment allow her to decide whether a particular treatment has a real effect on the response variable?

STUDENT NAME _____DATE _____

TAKE IT HOME

1 Imagine that you want to learn about the average number of hours, per day, that students at your college spend online. You want to select a simple random sample of 75 students from the full-time students at your college. You have a list of all full-time students, whose names are arranged in alphabetical order.

How would you select a simple random sample of 75 students from this population? Describe your process.

2 You want to estimate the average amount of time, per week, that students at a particular college spend studying. For each method, determine if the method is reasonable and why.

A Method A: Select 50 students at random from the students at the college.

B Method B: Select 100 students as they enter the library.

C Method C: Select 200 students at random from the students at the college.

D Method D: Select the 300 students enrolled in English literature at the college this semester.

E Which of the four sampling methods above do you think would be best? Explain your answer.

3 A researcher at King's College in London found that infomania (information overload) has a temporary, negative effect on intelligence quotient (IQ).[2] Imagine that a group of researchers wished to further test this conclusion. These researchers divided volunteers into two groups. Each subject took an IQ test. One group had to check email and respond to instant messages while they were taking the test. The other group took the IQ test without any distractions. Researchers found that the distracted group's average IQ test score was 10 points lower than the average IQ test score for the group that was not distracted.

A What is the explanatory variable, and what are its values (the individual treatments)?

B What is the response variable in this experiment?

C Explain why it would be good for the researchers to use random assignment to put each volunteer in one of the experimental groups. Why should the researchers do this rather than letting the volunteers decide which group they wanted to be in.

D Identify the control group in this experiment.

E Is it possible for the subjects of this study to be blinded? Explain your answer.

[2]The results of this study were never published, but the researcher himself eventually published a follow-up note about his work: "The 'Infomania' Study," http://www.drglennwilson.com/Infomania_experiment_for_HP.doc, accessed July 8, 2014.

4 Researchers recently designed an experiment to investigate whether dogs can be trained to recognize cancer by smell. In the experiment, dogs were trained to distinguish between people with and without cancer by sniffing exhaled breath. Dogs were trained to lie down if they detected cancer. After training, the dogs' ability to detect cancer was tested using breath samples from an unfamiliar group of people. According to the article that presented the study's results, "The researchers blinded both [the] dog handlers and [the] experimental observers to the identity of [the] breath samples."[3]

A Explain what was meant by the last sentence.

B Explain why this blinding is important for this experiment.

[3]Michael McCulloch et al., "Diagnostic Accuracy of Canine Scent Detection in Early- and Late-Stage Lung and Breast Cancers," *Integrative Cancer Therapies* 5 (2006): 30. For a summary of the study and its findings, see Linda Goldstein, "'Doctor Dogs' Diagnose Cancer by Sniffing It Out," *Knight Ridder Newspapers*, January 3, 2006, accessed July 5, 2014, http://www.azcentral.com/health/news/articles/0203doctordogs.html.