

Unit 16: Census and Sampling

SUMMARY OF VIDEO

There are some questions for which an experiment can't help us find the answer. For example, suppose we wanted to know what percentage of Americans smoke cigarettes, or what percentage of supermarket chicken is contaminated with salmonella bacteria. There is no experiment that can be done to answer these types of questions. We could test every chicken on the market, or ask every person if they smoke. This is a census, a count of each and every item in a population. It seems like a census would be a straightforward way to get the most accurate, thorough information. But taking an accurate census is more difficult than you might think.

The U.S. Constitution requires a census of the U.S population every ten years. In 2010, more than 308 million Americans were counted. However, the Census Bureau knows that some people are not included in this count. Undercounting certain segments of the population is a problem that can affect the representation given to a certain region as well as the federal funds it receives. What is particularly problematic is that not all groups are undercounted at the same rate. For example, the 2010 census had a hard time trying to reach renters.

The first step in the U.S. Census is mailing a questionnaire to every household in the country. In 2010 about three quarters of the questionnaires were returned before the deadline. A census taker visits those households that do not respond by mail, but still not everyone is reached. Some experts favor adjusting the census to correct the undercount using information gathered by smaller but more intense samples.

There is an alternative to a census, and that is a sample. While a census is an attempt to gather information about every member of the population, sampling gathers information only about a part, the sample, to represent the whole. Because a sample is only part of the population, we can study it more extensively than we can all of the members of the population. Then we can use the sample data to draw conclusions about the entire population. However, for those conclusions to be valid, the sample must be representative of the population. To make sure that it is, statisticians often rely on what is called simple random sampling. That means the sample is chosen in such a way that each individual has an equal chance to be selected. This helps eliminate bias in the study design, which occurs if certain outcomes are systematically favored.

Sampling is widely used in a variety of areas such as industry, manufacturing, agriculture, and medical studies, to name just a few. For example, consider food manufacturing with a look at processes Frito-Lay uses in making potato chips. Here's just some of what happens. A truck carrying 45,000 pounds of raw potatoes arrives at the plant, but is not allowed to unload until a sample of its potatoes has been carefully tested. First, a 150-pound sample of potatoes is taken from different locations in the truck (some from the front, middle and back). Next, an inspector selects 40 pounds of those potatoes and punches a hole through the core. Those holes make it easy to spot the samples when they undergo a cooking test. In other potatoes, the inspector searches for internal defects, green edges, rot, and other flaws. Each defective potato is weighed, and if the sample percentage is too large, the whole load must be rejected. The cooking sample is peeled and tossed directly into the slicing machine. Then the sample chips with their telltale holes are plucked out and go to a mini laboratory for further testing. Once everything in the sample is found to be up to specifications, Frito-Lay will accept the multi-ton shipment, based on the 150-pound sample.

All along the production line, workers continue taking samples to ensure the chip-making process stays on track. Sample chips are measured for thickness, color, and salt content. Even the finished bags are sampled to check their weight, both before and after being packed into cartons. If Frito-Lay waited until the end of the line to inspect the finished product, problems that were minor to begin with could be greatly compounded. Instead, sampling at key points catches problems early, before they get out of hand.

STUDENT LEARNING OBJECTIVES

- A. Know that a census is an attempt to enumerate the entire population; understand that a census is needed for information about every small part of the population, but for information about the population as a whole, a sample is faster, cheaper, and at least as accurate (if not more accurate).
- B. Recognize the distinction between population and sample.
- C. Recognize the strong bias in voluntary response samples, and generally in samples that result from human choice.
- D. Know what a simple random sample is and how to use a random digits table or computer random number generator to select a simple random sample.

CONTENT OVERVIEW

A **census** is an attempt to gather information about every member of some group, called the **population**. This unit introduces the U.S. Census and its problems in collecting data on the entire U.S. population. One of the most serious problems is undercounting certain segments of the population. Unfortunately not all groups of people are undercounted at the same rates. For example, undercount rates for minority groups are higher than for whites and undercounted rates for renters are higher than for homeowners. Moreover, undercount rates for those living in poverty are higher than for the affluent. The U.S. government uses sampling to estimate undercount rates for various groups. However, it never changes the official headcount number based on the results from sampling.

A **sample** allows the researcher to gather information from only a part of the population. Sampling – collecting data from a portion of the population – is the general means of gathering information about a population when it is not possible to get information from each individual in the population. Sampling saves both time and money. In some cases, such as for Frito-Lay potato chips, both the whole potatoes in a sample and the chips in a sample are destroyed as part of the data collection process. In such cases a census would be out of the question or there would be no product left to sell.

In order for a sample to provide good information about a population, the sample needs to be representative of the population. A **simple random sample**, a sample in which each member of the population is equally likely to wind up in the sample, is one means of ensuring that the sample is representative of the population and not biased. A simple random sample can be selected from the population in the same way that a subgroup is randomly selected from a larger group to receive a certain treatment. Hence, you should refer to Unit 15, Designing Experiments, for directions on selecting a random sample.

Sampling bias occurs when a sample is collected in such a way that some members of the population are less likely to be included than others. A voluntary television poll is an example of a biased sample. Since it is voluntary, only those with strong views are likely to call or text in to vote. Furthermore, only those watching the particular station at the time the poll is given will participate. In this case, the entire segment of the population who do not watch that particular station will be left out of the sample.

KEY TERMS

The entire group of objects or individuals about which information is wanted is called the **population**.

A **census** is an attempt to gather information about every individual in a population.

A **sample** is a part of the population that is actually examined in order to represent the whole.

A **simple random sample** 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 of being the selected sample.

Sampling bias occurs when a sample is collected in such a way that some individuals in the population are less likely to be included in the sample than others. Because of this, information gathered from the sample will be slanted toward those who are more likely to be part of the sample.

UNIT ACTIVITY:

THE U.S. CENSUS

Use a search engine (such as Google) to find the 2010 (or most recent) Census homepage. Then answer the following questions.

1. What is the current U.S. population? (Note this number will change. Check back at the end of the assignment to see how much the population has changed during the time you worked on this assignment.)

2. Click the Population Finder. Select your state from the scroll-down menu to access the 2010 (or most recent) Demographic Profile for your state.
 - a. What was the population of your state in 2010?
 - b. What percentage of your state's population was male? Female?
 - c. Which was higher for your state, the percent under 18 or the percent 65 or over? (Give the percentages.)

3. Select another state that is close to your state. Write a brief paragraph comparing the demographics of the two states.

EXERCISES

1. A local television station takes quick polls of public opinion by announcing a question on the 6 o'clock news and asking viewers to call-in or text their opinion of "Yes" or "No" to the station. The results are announced on the 11 o'clock news. One such poll finds that 73% of those who called in or texted are opposed to a proposed local gun control ordinance.

- a. What do you think the population is in this situation?
- b. Explain why this sampling method is biased. Is the percent of the population who oppose gun control probably higher or lower than the 73% of the sample who are opposed?

2. The students named below are enrolled in a new statistics course. Use a random digits table (such as Table B in *The Basic Practice of Statistics*) at line 136 or a calculator/computer random number generator to choose five of these students at random to be interviewed in detail about the quality of the course. Explain how you chose your sample.

Agarwal	Dewald	Hixson	Puri
Anderson	Fernandez	Klassen	Rodriguez
Baxter	Frank	Mihalko	Rubin
Bowman	Fuhrmann	Moser	Santiago
Bruvold	Goel	Naber	Shen
Casella	Gupta	Petrucelli	Shyr
Cordero	Hicks	Pliego	Sundheim

3. On the Hudson Valley, NY *Patch* Facebook page, readers were asked to send in stories of awful Valentine's Day gifts. The following were selected:

- Leftover chocolate (and he had eaten one!)
- Flowers purchased the day BEFORE Valentine's because it was cheaper to buy them the day before
- A recycled card from an ex-boyfriend with an open box of chocolates

Readers were then asked to vote on the best "worst Valentine's Day gift ever" story.

- a. Describe the population.
 - b. Describe the sample.
 - c. Do you think the response to this poll is representative of the views of the residents of Hudson Valley, New York? Explain.
4. Identify the population and the sample in each of the following situations.
- a. A realtor is interested in the median selling price of homes in Worcester County, Massachusetts. She collects data on the selling prices of 50 homes.
 - b. A psychologist is concerned about the health of veterans who served in combat. She examines 25 veterans to assess whether or not they are showing signs of post-traumatic stress disorder (PTSD).
 - c. An educator asks 20 seniors from Eastern Connecticut State University whether or not they had taken an online course while at the university.

REVIEW QUESTIONS

1. The students listed below are enrolled in an elementary French course. Students are assigned to small conversation sections at random.

Arnold	Ashford	Bartkowski	Barrett
Beerbohm	Burns	Campbell	Chang
Colon	Deneuve	Dodington	Drummond
Elsevier	Erskine	Garcia	Fernandez
Flury	Hardy	Holmes	Hyde
Jones	Juarez	Kempthorne	Levine
Martinez	Moore	Munroe	Neale
Nguyen	Oakley	Orsini	Perlman
Poe	Prizzi	Putnam	Quincy
Randall	Rodriguez	Rostenkowski	Rowley
Schiller	Scott	Smith	Stevenson
Swokowski	Taylor	Vuong	Ward

- a. Choose a simple random sample of eight of these students to form Section 01. Explain how you obtained the names for the first section.
 - b. Assign the remaining students at random to the Sections 02, 03, 04, 05 and 06. Explain the process you used to make the assignments.
2. Identify the population and the sample in each of the following situations.
- a. A professor asks a sample of students during their college orientation whether they planned to take an online course their first semester at college.
 - b. A physical therapist is investigating a new exercise regimen to see if it could improve the function of arthritic knees. She chooses 10 of her patients and has them follow the new exercise regimen.
3. A university president wishes to know what types of activities and jobs graduates of the university are doing 5 years after graduation. You have been asked to deliver this information to the president.

- a. What is the population of interest?
 - b. State reasons for taking a sample rather than a census to obtain information for the president.
4. Suppose you want to know whether or not a population supports a certain measure. You have one month to find out.
- a. List some of the pros and cons for getting this information by conducting a census.
 - b. List some of the pros and cons for getting information about a population by taking a sample.