

# Unit 17: Samples and Surveys



## SUMMARY OF VIDEO

Listen to a news broadcast, read a newspaper, open a magazine and you'll find an item related to the results of a poll on some topic. But how do pollsters survey a population as large and diverse as the United States and wind up with a complete and unbiased picture of attitudes on a particular topic? Before selecting the sample to be surveyed, pollsters need to identify the population – the group of interest to the study.

For example, that could be all registered voters, people in a particular age group, or households with a certain income. The characteristic of a population that we are interested in is called a parameter. We can't determine the true value of a parameter unless we can examine the entire population, which isn't usually possible. However, we can estimate the unknown parameter based on information collected from a sample, a subset of the population. Such an estimate is called a statistic. (Remember **P**arameters are for **P**opulations and **S**tatistics are for **S**amples.)

The pollsters at the University of New Hampshire's Survey Center conduct everything from academic research to political polls. They are experts at selecting samples to represent the attitudes and opinions of the whole population. For a public opinion survey they use random digit dialing to select households; they start with a random sample of households. However, another sampling stage is required because they talk to an individual at each house and not to everyone in the house. So, for example, they might ask to speak to the person in the house who has had the most recent birthday.

Convenience sampling, where pollsters survey a convenient group such as their friends, or voluntary sampling, where data are collected from those who volunteer for the survey, often create an unrepresentative sample and produce biased results. The same issue arises when a sample draws from a list that excludes a portion of the population. This happened in the 1936 presidential election when a *Literary Digest* poll predicted Alf Landon would be the winner. However, Franklin Roosevelt went on to win in a landslide. The *Literary Digest* drew their samples from lists of car and telephone owners – items that, at the time, were indicative of wealth. The poll had omitted the largely pro-Roosevelt poor from their survey, causing bias in favor of Landon.

Getting a representative sample is the cornerstone of accurate sampling. But just as important as a representative sample is the design of the questionnaire. Some practical advice on questionnaires includes using simple words, not asking people about things they are not likely to know about, and keeping the questions short. Even small things such as changing the order in which you read the choice of responses to survey participants can change how they answer a question.

In a simple random sample, each individual of the population has an equal chance of being selected. This can be hard to achieve in a real-life survey since it can be nearly impossible to get a complete list that includes every single member of a large population. Another way of ensuring a representative sample is by doing a multistage sample. For example, the Survey Center might begin with a random sample of counties in New Hampshire. Then they would take a random sample of towns within those counties. Finally they would select random households within those towns.

The problem with multistage sampling is that it could leave out groups of interest merely by chance. To solve this problem, the Survey Center might decide to use a stratified random sample. For this type of sampling design, the entire population is divided into groups with similar characteristics called “strata.” For example, census tracts might first be classified as urban, rural, or suburban, and then a separate random sample is selected from each stratum. In New Hampshire this ensures that cities are represented in the sample even though most counties are rural.

# STUDENT LEARNING OBJECTIVES

- A. Realize that most national social and economic data are produced by large-scale sample surveys.
- B. Know that samples of large, geographically dispersed human populations rarely use a simple random sample; be familiar with multistage samples and stratified samples.
- C. Recognize that non-statistical aspects of sampling such as training of interviewers and wording of questions can have strong effects on the results of a sample survey.

# CONTENT OVERVIEW

This unit describes methods related to conducting surveys. Particularly when populations are large, geographically-dispersed human populations, it would be nearly impossible to include everyone in a survey. So, one aspect of conducting a good survey is the **sampling design** – the method used to choose a **sample** that is representative of the **population**. Equally important are the design of the questions and interviewer training.

**Convenience sampling** and **voluntary sampling** are two methods for choosing a sample that may not produce a **representative sample**. In convenience sampling, a sample is chosen in a way that makes it easy to obtain. For example, the pollster could stand outside a grocery store on some weekday morning and interview people as they enter the store. That would be an easy way to get a sample, but the sample probably won't be representative of the opinions of the population – for one thing, most likely there will be more women in the sample than men, and the sample won't contain people who work weekdays 9 to 5. So the sample will be **biased** toward the views of women who are not working weekday mornings. Voluntary sampling is equally hazardous. A television show might ask people to call or text in their responses. Generally people who feel strongly about a topic are more likely to volunteer.

Using random sampling techniques as part of the sampling plan produces samples that are more likely to be representative of the population. In a **simple random sample**, every person in the population has an equal chance of being chosen for the sample. However, for large populations, a simple random sample can be difficult to conduct. Here are two new concepts of sample design: **multistage samples** and **stratified samples**. For a two-stage sampling process, a sample of clusters is first selected and then random samples within each cluster are chosen. For a stratified sampling process, two or more strata are defined and then random samples are taken from each stratum.

Questionnaire design concerns the wording of questions and the overall order and length of the questionnaire. In terms of wording, consider the following:

- Don't use long words when a shorter word would mean the same thing.
- Stay clear of words that might be unfamiliar to respondents.
- Be sure that questions are neutral and do not lead the respondent in a particular direction.
- Keep sentences relatively short and simple.

- Avoid asking two questions in one – for example, the question “Have you argued with your friends or parents this month?” is really two questions in one.
- Be specific and avoid terms that are vague. For example, words such as “often” or “sometimes” should be replaced by specific terminology such as “every day” or “once a week.”
- Finally, interviewers need to be trained not to show their own opinions and not to suggest answers, but to encourage people to respond. In addition, the gender or race of an interviewer needs to be taken into account. For example, people may give different answers about racial issues depending on the race of the interviewer.

# KEY TERMS

The **population** is the entire group of individuals about which information is desired.

A **sample** is a subset of the population from which information will be extracted. A

**representative sample** is one that accurately reflects the members of the entire population. A

**biased sample** is one in which some individuals or groups from the population are less likely to be selected than others due to some attribute.

A **sampling design** describes how to select the sample from the population. There are many sampling designs, including the following:

- **Simple random sampling** is a sampling design that chooses a sample of size  $n$  using a method in which all possible samples of size  $n$  are equally likely to be selected.
- **Convenience sampling** is a sampling design in which the pollster selects a sample that is easy to obtain, such as friends, family, co-workers, and so forth.
- **Voluntary sampling** or **self-selecting sampling** is a sampling design in which the sample consists of people who respond to a request for participation in the survey.
- **Multistage sampling** is a sampling design that begins by dividing the population into clusters. In stage one, the pollster chooses a (random) sample of clusters. In subsequent stages, random samples are chosen from each of the selected clusters.
- **Stratified sampling** is used to ensure that specific non-overlapping groups of the population are represented in the sample. The non-overlapping groups are called **strata**. In a **stratified random sample**, the sample is obtained by taking random samples from each of the strata.

# THE VIDEO

Take out a piece of paper and be ready to write down answers to these questions as you watch the video.

1. Why was the *Literary Digest* poll so far wrong in predicting the outcome of the 1936 presidential election?
2. Why would a simple random sample of counties in a state give results that might not represent the entire state?
3. In sampling, what are strata?
4. You are an interviewer for an opinion poll. How should you react to answers that seem anti-social or immoral?

# UNIT ACTIVITY:

## CONDUCTING A SURVEY

In this activity, you will write a short survey questionnaire to learn something about students at your school or campus. Then you will develop a sampling plan for selecting a sample of students who will complete the survey.

1. In your group, discuss what you would like to know about students at your school/campus (for example, their study habits, personal tastes, opinions on some topic, etc.). Then create a set of questions designed to gather information on the topics you have discussed. Provide choices for the responses to each question. (For example, the response could be Yes or No; the response could be to rate something on a scale from 1 to 5; recall in the Somerville Happiness Survey in Unit 13, the choices for Happiness were Unhappy, So-so, and Happy.)
2. Next, you need to plan how and to whom you will administer your survey. Your sample should include at least 100 students. However, the quality of the information you get from your sample survey will depend on the sampling plan.
  - a. First identify your target population.
  - b. Describe in precise detail exactly how the sample will be selected and the survey questionnaire administered. Be prepared to present your plan to the class and explain why you think it will produce a representative sample.

Go out and collect your data!

# EXERCISES

1. A big-city police department wants to know how African-American residents of the city feel about police service. They prepare a questionnaire with several questions about the police. A sample of 300 mailing addresses in predominantly African-American neighborhoods is chosen, and a police officer is sent to each address to administer the questionnaire to an adult living there. Do you think that this sample survey will produce trustworthy information? Why or why not?
2. Comment on each of the following as a potential sample survey question. If either question is unclear, slanted, or too complicated, restate it in better words.
  - a. Which of these best represents your opinion on gun control?
    - i. The government should confiscate our guns.
    - ii. We have the right to keep and bear arms.
  - b. In view of escalating environmental degradation and predictions of serious resource depletion, would you favor economic incentives for recycling of resource-intensive consumer goods?
3. A large company has been accused of not promoting women as quickly as men. You want to take a sample survey among the company's 20,000 employees to see if they believe that promotion policies are fair. Briefly describe how you would design the sample.
4. Explain why each of the following samples might be biased. Select an alternative method for choosing a more representative sample.
  - a. The campus food service wants to know how students feel about their food. They hand out a survey during Friday morning breakfast between 7 a.m. and 9 a.m.
  - b. The President of the United States wants to check his/her approval rating after two years in office. A sample of 1000 voters is selected from California.

# REVIEW QUESTIONS

1. Do you think that this is a good question to ask in a sample survey? Explain your answer.  

A freeze in nuclear weapons should be favored because it would begin a much-needed process to stop everyone in the world from building nuclear weapons now and reduce the possibility of nuclear war in the future. Do you agree or disagree?
2. You want to study the attitudes of college faculty members toward undergraduate teaching. These attitudes appear to be different depending on the type of college. The American Association of University Professors classifies colleges as follows:  

Class I: Offer doctorate degrees and award at least 15 per year.

Class IIA: Award degrees above the bachelor's but are not in Class I.

Class IIB: Award no degrees beyond the bachelor's.

Class III: Two-year colleges.

Suggest a sampling design for collecting a sample of faculty from colleges in your state, with total sample size about 200.

3. A large university wants to conduct a focus group on campus satisfaction. Below are descriptions of sampling plans the university might use for selecting a sample. Identify the type of sampling design.
  - a. The university randomly selects five dorms, then randomly selects 10 rooms from each dorm, and then randomly selects a student living in each room.
  - b. The university gets a complete list of all full-time students enrolled at the university along with their class (freshman, sophomore, junior, senior). A random sample of 20 names is selected from each class.
  - c. Ten resident assistants were recruited and asked to find 10 students from their residence halls to participate in the focus groups.
  - d. A poster was hung in the Student Union inviting students to participate (with the promise of free food!).

4. Advice columnist Ann Landers received a letter from a young couple who were thinking about whether or not to have children. The letter stated their main concern was that many of their friends appeared to resent having had children and they wanted Landers' advice. Landers put the question to her readers: "If you had to do it over again, would you have children?" She reported the results of her survey in the June 1976 issue of *Good Housekeeping* magazine: 70% of the respondents replied "No."

In a sidebar, *Good Housekeeping* responded: "All of us at Good Housekeeping know that no mother will be able to read Ann Landers' report without passionately agreeing or disagreeing. We would like to know what your reaction is." Then they asked their readers to respond to Landers' question. In the October 1976 issue, *Good Housekeeping* reported that 95% of the respondents said "Yes."

Two newspapers, The *Kansas City Star* and *Newsday*, conducted their own polls. The *Kansas City Star* randomly selected a sample from the Kansas City area. *Newsday* conducted a national poll using random selection techniques. The percent of "Yes" responses was 94% to the *Kansas City Star* poll and 91% to the *Newsday* poll.

Explain why there could be so much discrepancy in the results from the four surveys. Which poll would you most trust to give an accurate estimate of the percent of United States parents who would respond "Yes"? Support your answer.