

# Unit 17: Samples and Surveys



## PREREQUISITES

This unit continues the discussion of sampling begun in Unit 16, Census and Sampling, which is a prerequisite.

## ADDITIONAL TOPIC COVERAGE

Additional coverage on sampling designs can be found in *The Basic Practice of Statistics*, Chapter 8, Producing Data: Sampling.

## ACTIVITY DESCRIPTION

In this activity, students design a short survey questionnaire. Then they develop a sampling design for selecting a sample of students to complete the questionnaire. If possible, they go out and collect the data. Once they have collected the data, encourage students to analyze the responses using techniques from Unit 13, Two-Way Tables or Unit 29, Inference for Two-Way Tables. Some other units, such as Unit 28, Inference for Proportions, may also be useful.

Students should work in small groups on this activity. In question 1 they are asked to create questions for a survey questionnaire. Each group should come up with several questions. After groups have completed drafting their questions, the class should compose a 4 – 7 question survey using a subset of the questions written by the groups. To make the questionnaire interesting, encourage students to select questions on a variety of topics. (Or, to save time, you can select the questions for the class survey questionnaire yourself.)

In question 2, students are asked to describe in detail their sampling design for selecting a sample of at least 100 students from their school or campus. (If your school/campus is small, you may want to involve another school in this activity.) The goal is to select a representative sample. After groups have completed their sampling plans, allow them to share them with the class. Then select a design for the survey.

Among instructors who have tried this activity, there is a difference of opinion of the value of spending a lot of time on the careful collection of the survey data. Much depends upon the seriousness of the questions in the survey. If the conclusions have the potential of affecting life

at your school or campus, then there is some intrinsic value in spending the time apart from experiencing the difficulties in getting reliable information through opinion polls.

# THE VIDEO SOLUTIONS

1. The sample was drawn from telephone lists and lists of people who owned cars. In 1936 only people who had considerable wealth had phones and bought cars. So, no data were collected from people who were not wealthy – an overwhelming majority of those people happened to be Roosevelt supporters.
2. A simple random sample might miss counties that are rural, suburban or urban, particularly if one of these types of counties is sparse among the various types of counties in a state.
3. Sometimes the entire population is divided into groups with similar characteristics. For example, the population could be divided into males and females. These non-overlapping groups are called strata.
4. The interviewer should not react any differently to those responses than he/she does to “normal” responses. In other words, there should be no reaction on the part of the interviewer.

# UNIT ACTIVITY SOLUTIONS

1. Sample questionnaire from Pattonville High School:

Q1. Students at Pattonville High School are required to do 50 community service hours to graduate. Do you agree or disagree that this is a reasonable requirement?

Agree          Disagree

Q2. Do you think that there is mutual respect between the students and teachers at our high school?

Yes          No          Not Sure

Q3. Some public schools begin the school year late in August. Other schools do not open until after Labor Day. Do you think that public schools should begin their year after Labor Day?

Yes          No          No Opinion

Q4. During this year have you participated in an organized sport?

Yes          No

2. a. Sample answer: The population is all students at the school, college or university.

Students could decide that they only wanted to get information from seniors or that they were only interested in students majoring in STEM fields, or that they only wanted information from students in a certain dorm. So, the target population can vary.

b. The designs could be highly variable.

Sample answer from high school: We decided to use a multistage sampling plan. We started by taking a random sample of 12 homerooms and asked for a list of students in each of these homerooms. Then we took a random sample of 10 students in each homeroom. This gave us a sample of 120 students. We asked the homeroom teachers to administer the surveys to the students selected from their homeroom. We selected a total of 120 students in case some students were out of school on the day the surveys were administered.

Sample answer from college/university: We decided to use a multistage sampling plan. We randomly selected 6 dorms and then randomly selected two floors from each dorm. Then we got a list of the room numbers and randomly selected 10 rooms. We got the names of

the students in these rooms and plan to e-mail them the questionnaire. We will work with the RAs in each dorm to get the word out that we need this for our class and that students should respond. For students who fail to respond via e-mail, members of the class will go to their dorm rooms and try to track them down.

# EXERCISE SOLUTIONS

1. Many African-American respondents would be unwilling to make negative comments about the police department to a police officer. To get trustworthy information, the poll should be taken by an independent agency that can preserve the confidentiality of the answers. Because the survey specifically seeks the opinions of African Americans, it would be a good idea to employ African-American interviewers.

2. a. The alternatives offered are slanted. We are asked to choose between a strong statement using the word “confiscation” versus a phrase from the Constitution. Forced to make this choice, many respondents will choose (ii) even though they might be sympathetic to a more reasonable gun control measure in light of concern over gun violence. One neutral question is “Some people have proposed greater restrictions on owning firearms. Do you agree or disagree that greater restrictions are needed?” However, a great variety of rewordings are possible and students may suggest even better alternatives.

b. The wording is impossibly complex. A translation into simpler English might be slanted because the question suggests reasons why recycling is desirable. Here’s one possible rewording: “Would you be willing to pay more for the products you buy if the extra cost were used to save resources by encouraging recycling?”

3. Sample answer: Because we want to ensure that both men and women are part of the sample (and we don’t know the percentage of women employees), it is best to use a stratified sampling plan. Particularly if there are fewer women than men, stratifying can ensure that we can draw conclusions about the women separately. Get a list of all employees separated by gender. Draw a random sample from the men and another random sample from the women.

Some students may explore multistage sampling. Perhaps the company has employees in many geographic locations, for example. We might sample locations, and then sample individuals at each location. Either or both stages could employ stratified samples.

4. a. Sample answer: The sample is biased because it will not include students who never eat breakfast or students who were out late Thursday night and didn’t show for Friday breakfast.

Alternative sampling plan: On Tuesday and Friday hand out questionnaires during breakfast, lunch, and dinner. Make sure to ask students whether they had already filled out the questionnaire so that they don't fill it out twice. This way the sample will include students from each mealtime and also students who show up on a Monday/Wednesday/Friday schedule or on a Tuesday/Thursday schedule.

b. Sample answer: The sample is biased because it does not include voters from any of the other 49 states. (Also the east and west coast tend to be more liberal than the interior states.)

One possibility is to use a multistage sampling plan in which a random sample of states are selected, then a random sample of counties within those states, and then a random sample of voters from each of the selected counties.

# REVIEW QUESTIONS SOLUTIONS

1. The question is slanted in favor of a freeze by suggesting desirable outcomes.
2. The question has no single correct answer, but the problem almost demands a stratified sampling of colleges to ensure representation of each type of institution.

Sample answer: For the first stage, we need to pick some colleges and universities. We would start by looking at the number of colleges in the state – if there are only a few Class I universities, all of them large, we would include all of them in our sample of colleges and universities. Then we would sample a few from each of the other classes. We plan to select 10 institutions in the first sampling stage making sure, if possible, to have at least two from each class. For the second stage, we would take a simple random sample of 20 faculty members from each of the selected institutions.

Some students may decide to stratify the faculty by discipline (business, engineering, science, etc.), though that will quickly use up the 200 available places. However, they could also stratify faculty by rank or stratify disciplines by STEM and non-STEM fields.

3. a. Multistage sampling design.
- b. Stratified random sample.
- c. This could be a convenience sample. The resident assistants could select friends, or students who just walked through the door at a certain time – whatever was easiest.
- d. Voluntary sampling.

4. Sample answer: The Ann Landers' poll did not ask the question in a neutral fashion. The letter that prompted the poll included the concern that the couple had friends who appeared to resent having had children. Including that statement may make it more likely for others who felt the same way to volunteer a response to the poll – thus, resulting in an unusually low estimate for the “Yes” response. The sample of those who responded to Lander's poll was clearly not representative of the population of parents in the U.S. The *Good Housekeeping* poll was also a voluntary poll. However, the sidebar introducing the question was quite neutral



and their response rate was much closer to the response rates of the *Kansas City Star* poll and *Newsday* poll. Even though the *Kansas City Star* poll randomly selected their participants, the sample was only from one state and shouldn't be used to make a statement about U.S. parents in general. Since the *Newsday* poll used random selection techniques to select a sample from the nation, the results from that poll are the most trustworthy.