

Section II: Surveys

THREE METHODS FOR PRODUCING DATA—SURVEYS, OBSERVATIONAL STUDIES, AND experiments—were discussed in the Introduction. In this section, we examine surveys in more detail. A survey is a type of study in which individuals are asked one or more questions. The survey questions are worded so that the resulting responses will provide data that help answer questions about some population of interest.



If every individual in the population provides responses to the survey questions, the study is called a **census**. A census is the usual method of collecting data only if the population of interest is very small—the students in your math class, for example. However, if the population is large, it is more common for only a subset of the population to provide responses to a survey. In this case, the group of individuals who respond to the survey is referred to as a **sample**.

When only a sample participates in a survey, the way in which the individuals in the sample are selected is critical. As with observational studies, if we want to generalize the results of a survey to the entire population, we need to select the sample in a way that is likely to result in a representative sample.

A popular classic movie called “Magic Town” (1936) featured an actor named Jimmy Stewart playing a very successful pollster. He was able to accurately determine the opinions of the entire United States simply by surveying all the residents of a small town called Magic Town. Because this town was a flawless mirror of the entire country, its residents constituted the perfect sample. Unfortunately for those planning surveys, Magic Town is fictional and much more care needs to go into sample selection!

Just as with observational studies, sample selection can be random or nonrandom. To be reasonably confident that the selected sample will be representative of the population, some type of random selection is required. It is sometimes tempting to select the sample in a nonrandom way just because it is convenient to do so. For example, it might be easy to use the students in your math class as a sample of the students at your high school, but there are many reasons why this sample may not be representative of the entire school—the class may consist of mostly seniors, for example. Because there is no way to tell by just looking at a sample if it is representative of the population, our only assurance comes from the method that was used to choose the sample and from the role that random selection played in the choice.

In addition to being thoughtful about how the sample will be selected, it is also important to think carefully about how the actual survey questions will be worded. Each question should be evaluated to determine if it uses appropriate vocabulary and simple sentence structure and to make sure that the question is clear. This will help to ensure that the survey responses, in addition to being representative of the population, are unambiguous and can be generalized in a straightforward manner.

There is one last thing to think about when planning a survey—how large should your sample be? You want the sample to be large enough so that it can reasonably represent the population of interest. On the other hand, it can be both costly and time-consuming to carry out a survey with a large sample size. Because larger samples tend to provide more information than smaller samples, you will need to consider both the desire for a large sample and the available resources for carrying out the survey to arrive at a reasonable sample size.

Planning and carrying out a good survey is a complex task. This overview and the following investigations just provide the basics. You can learn more about surveys in a course in statistics and data analysis. In the investigations that follow, you will explore aspects of planning surveys and analyzing the data that result from them.

In Investigation #9, you will have the opportunity to design and carry out a survey. Collecting survey data involves asking people to share personal opinions or ideas. Not everyone feels comfortable doing that. Any individual has the right to refuse to participate in a survey. When you are in the role of researcher, you must respect that right. It is also your responsibility to preserve the anonymity and confidentiality of responses.





Investigation #7: Welcome to Oostburg!



Oostburg is a small town in Wisconsin. The 306 residents of this town are very data-driven! They are willing and anxious to respond to surveys and give their opinions about various issues. A recent survey was conducted in Oostburg and every person who lives there responded. (Although baby Edna, the youngest citizen of Oostburg at only 8 months old, was not able to answer any of these questions, her parents were willing to respond for her.) This particular survey included questions about age, sex, voting behavior, and participation in various activities during the last month. Data from the survey are summarized in the following two tables.

Age	17 and Younger		18 to 40 Years Old		41 to 60 Years Old		61 and Older	
Sex	Male	Female	Male	Female	Male	Female	Male	Female
Number of Responses	36	41	32	35	38	46	32	46

Age	17 and Younger		18 to 40 Years Old		41 to 60 Years Old		61 and Older	
Sex	Male	Female	Male	Female	Male	Female	Male	Female
Voted in last town election	0	0	10	12	28	40	29	43
Attended a movie during the last month	4	5	16	22	12	22	4	7
Ate fast food at least once during the last month	6	7	28	33	8	5	3	4
Shopped for clothes online during the last month	5	6	26	28	4	7	0	1
Watched "The Simpsons" during the last month	23	27	29	31	12	8	1	2

What do the data tell us about Oostburg residents? Given that the entire population of Oostburg was surveyed, the above data is a census of the town. Use the given data to answer the following questions.

1. If a resident of Oostburg is to be selected at random, what is the probability that the person selected:

(a) attended a movie during the last month?

(b) attended a movie and is 18 to 40 years old?

(c) is male?

(d) is between 18 and 60 years old?

2. Estimate the probability that a person selected at random is between 10 and 30 years old. Why is this probability more difficult to compute than those of question 1?

3. What is the probability that a person selected at random is male and did not watch “The Simpsons” in the past month?

4. Pose two other probability questions that could be answered using the survey data and then answer those questions by computing the relevant probabilities.

5. Below are seven headlines from the *Oostburg Herald*, a local newspaper. Evaluate the accuracy of each headline based on the survey data. Write a sentence or two giving your assessment of the headline, using the survey data to support your evaluation.

(a) “70% of Eligible Voters Turned Out for Election” (Assume the eligible age of voting in Oostburg is 18.)

(b) “Over 60 Crowd Not Responding to Online Shopping”

(c) “Movies Are Reaching Across ALL the Generations”

(d) “Fast Food Eating a Big Thing with the Younger Crowd”

(e) “The Simpsons’ Not Popular with Older TV Viewers”

(f) “40% of People Over 60 Voted in the Election!”

(g) “Oostburgians Eating Preferences Dependent on Age!”

6. The section overview describes “Magic Town,” a town that is a flawless mirror of the entire country. Do you think Oostburg could be such a magic town? Explain your reasoning.

Hugo VanHorn, a senior at Oostburg High School, did not have access to the data from the survey described here. For a school project, Hugo decided to investigate the popularity of online shopping in Oostburg. After band practice, he quickly asked 40 band members if they had shopped for clothes online in the past month. The results from his survey are summarized below:

Have you shopped for clothes online during the past month?	
Yes	No
28	12

Hugo was quite impressed with his results so he wrote a report about the popularity of online shopping in Oostburg. His report indicated that 70% of the residents of Oostburg had shopped for clothing online in the past month.

7. Is the statement that 70% shopped for clothing online in the past month an accurate summary of Hugo’s sample? Explain your answer.

8. Is Hugo’s statement that 70% of all Oostburg residents shopped for clothing online during the past month an accurate statement? Justify your answer.

9. Hugo's sample was a convenience sample; he did not randomly select his survey participants from the residents of Oostburg. As a consequence, Hugo's sample was not representative of the Oostburg population. In fact, residents in one of the age groups were over-represented in his sample. Based on the census survey data, which age group do you think was over-represented in Hugo's sample? Explain your reasoning.



10. Why would it have been better for Hugo to have used random selection in choosing the 40 people who would participate in his survey?

11. Assuming that Hugo would like to be able to use survey data to generalize to the Oostburg population, write a brief set of instructions that Hugo could use to select 40 participants for a new survey.



Investigation #8: Student Participation in Sports



The short article below is from the student newspaper at Rufus King High School. Use the information from the article to answer the following questions.

Student Survey Finds Females More Involved in Sports

By Kayla Johnson

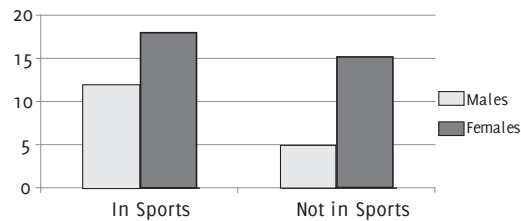
A Rufus King mathematics class conducted a survey to investigate student participation in extra-curricular activities. Fifty randomly selected students participated in the survey.

One question on the survey asked about participation in school sports programs. The accompanying graph shows participation by sex.

Shauna Rafferty, a junior at Rufus King and captain of the girls' soccer team, noted that there were more females in the sample that participated in sports. She said, "I think the girls in our school are more active in sports than the males. I am sure the success of our soccer team played a large role in this increased interest."

Bryon Jones, a junior on last year's state championship boys' basketball team did not agree. He responded, "All of the guys in my classes belong to one of the sports programs in the school."

Participation in Sports at Rufus King High School



Mr. Samuelson, the athletic director of the school, indicated that the number of sports programs and the number of students participating in the programs has posed a real problem in scheduling practice sessions. "Hopefully we will not have to eliminate some of the programs available to our students because students are not able to get adequate practice session time," Mr. Samuelson indicated.

Mr. Samuelson further stated that it was difficult to balance the demand for time in the weight room, the gym, and the outdoor fields.

1. How many of the 50 students surveyed were females involved in sports?
2. How many of the 50 students surveyed were females not involved in sports?

- 3.** How many of the 50 students surveyed were males involved sports?
- 4.** How many of the 50 students surveyed were males not involved in sports?
- 5.** The article says that the 50 students surveyed were randomly selected. Describe one way in which this random selection might have been accomplished.
- 6.** For each of the sample selection methods listed below, give two reasons why random selection of survey participants would be preferable.
- (a) Give the survey to the first 50 students who arrive on campus on a Friday morning.
- (b) Give the survey to all of the students enrolled in the school's two sections of pre-calculus.
- (c) Give the survey to all students who use the weight room on a particular day.

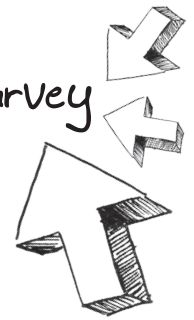
7. What percent of the 50 students in the survey sample were female?
8. What percent of the 50 students in the survey sample were male?
9. What are two reasons that the percent of females in the survey sample might not be 50%?
10. Based on the results of the survey, do you think that the number of girls attending the school is about the same as the number of boys attending the school? Explain your reasoning.
11. Is Shauna correct in her statement that there were more girls in the sample who participated in sports than boys who participated in sports?
12. Explain why Shauna may not be correct in her statement that the survey results imply that girls at the school are more interested in sports than boys.

13. The headline in the school newspaper states more females participate in sports. Explain how this statement could be considered accurate and explain how this statement is at the same time misleading.

14. Write a replacement headline that is not misleading, and then write a few sentences that you think accurately summarize the survey results.



Investigation #9: Planning and Conducting a Survey



In this investigation, you will develop a sampling plan and carry out a survey to investigate tooth-brushing behavior of students at your school. Consider the following two recommendations.

From *www.animated-teeth.com*:

As you might guess, many humans simply aren't self-disciplined enough to brush properly when they use a manual toothbrush. As a general rule, most people should brush their teeth at least twice a day with each **brushing period encompassing at least two to three minutes**. The fact of the matter is that most of us fail to routinely meet these guidelines.

From a *Los Angeles Daily News* (December 15, 2007) article titled "Water District Asks Users to Cut Back by 10 Percent. Drought Depleted Supplies Spur Voluntary, Mandatory Measures to Limit Consumption"

The Las Virgenes Water District is asking residents to reduce water use by 10 percent and is ordering farmers to cut back by a third. In seeking voluntary and mandatory cutbacks, the district follows the lead of Long Beach and other cities responding to an ongoing drought. "With no relief to the drought in sight, we must take steps now to ensure we have adequate supplies for the coming year," said John Mundy, the district's general manager. "We are dealing with water cutbacks throughout the state." ... Since nearly 70 percent of water is used outdoors, the district is asking residents to reduce use, water every other day and to sweep, rather than hose off, driveways. They also called upon residents to fix leaks, take shorter showers, and **shut off faucets while shaving or brushing teeth**.

1. Write a set of survey questions that would allow you to get responses regarding the following three characteristics of selected students at your school:

Sex of the survey participant

Whether or not the survey participant leaves the water on or turns the water off while brushing his or her teeth

How long, in seconds, the survey respondent thinks that he or she spends when brushing his or her teeth

You can also include other questions you think might be of interest.

2. As a class, discuss the proposed survey questions and come to an agreement on the wording of the questions to be included in the survey. Record the final version of the survey questions below.

3. As a class, discuss whether you think it would be easy or difficult to obtain a random sample of 50 students at your school and to obtain the desired survey information from all the students selected for the sample. Write a few sentences summarizing the class discussion in the space below.

4. As a class, decide how you will go about selecting a sample of 50 students that reasonably could be considered representative of the population of students from your school. Write a brief description of the sampling plan, and point out the aspects of the plan that make it reasonable to argue that it will be representative.

5. Carry out the survey and record the responses in the table below. If you included additional questions in your survey, you can modify the data sheet as needed. As a reminder: *Collecting survey data involves asking people to share personal opinions or ideas. Not everyone feels comfortable doing that. Any individual has the right to refuse to participate in a survey. When you are in the role of researcher, you must respect that right. It is also your responsibility to preserve the anonymity and confidentiality of students' responses.*

Survey Data							
Respondent	Sex (M or F)	Water Off (Y or N)	Time Spent Brushing	Respondent	Sex (M or F)	Water Off (Y or N)	Time Spent Brushing
1				26			
2				27			
3				28			
4				29			
5				30			
6				31			
7				32			
8				33			
9				34			
10				35			
11				36			
12				37			
13				38			
14				39			
15				40			
16				41			
17				42			
18				43			
19				44			
20				45			
21				46			
22				47			
23				48			
24				49			
25				50			

Now use the survey data to answer the following questions.

6. Construct a bar chart of the “water off” data. What proportion of survey respondents reported that they turn the water off while brushing their teeth?

7. Think for a minute about how the students in the sample were chosen. Do you think the proportion of students at your school who report that they turn the water off while brushing is likely to be much smaller than, much larger than, or somewhere near the value of the proportion computed in question 6? What aspect of the survey design supports your answer?

8. Sometimes there is a difference between what people say they do and what they *actually* do. Do you think this might be the case for the “water off” question? Explain your reasoning.

9. What proportion of the girls in the survey sample report that they turn the water off while brushing? How does this compare to the proportion of boys that say they turn off the water?

10. Use the reported brushing time data to construct a dotplot. Write a few sentences describing what the dotplot tells you about the distribution of brushing times.

11. Now construct a dotplot that uses color to distinguish between the reported brushing times of females and the reported brushing times of males (use one color for dots that correspond to responses that came from females and a different color for the dots that represent responses from males). Does this plot suggest that females tend to report longer brushing times? Explain.

12. Find the median of the data set consisting of the 50 reported brushing times. Divide the survey responses into two groups—those whose reported brushing times were less than the median brushing time and those whose reported brushing times were equal to or greater than the median brushing time. Use the table below to organize the information needed to compute the proportion that report turning off the water while brushing for each of these two groups. Do these proportions suggest that people who brush longer may be more likely to turn off the water while brushing? Explain.

	Below the Median Brushing Time	Equal to or Greater than the Median Brushing Time
Number in the Sample		
Number Who Report They Turn Off Water		
Proportion Who Report They Turn Off Water		

13. The web site referenced earlier (www.animated-teeth.com) also included the following:

Actually, the statement that most people aren't self-disciplined enough to brush properly when they use a manual toothbrush is probably a little bit harsh. Research has found that there can be a major discrepancy between the amount of time that a person actually does brush, as compared to the amount of time that they perceive they have brushed.

One study (*Journal of Clinical Dentistry*, 1998, 9(2):49-51) found that their test subjects, on average, brushed their teeth for 78 seconds (a little longer than a minute) when they actually thought they were brushing for 141 seconds (over two minutes, an adequate amount of time). So, the intention of these people was appropriate but in reality their actions (actual brushing time) were lacking.

Compute the mean of the 50 reported brushing times in the survey data set. How does your sample mean compare to the value of 78 seconds in the quote above?

14. As a class, discuss how you might design a study that would help you determine if there is a discrepancy between reported brushing times and actual brushing times for students at your school. Write a few sentences summarizing the class discussion.