Teacher Notes for Section II: Surveys

In Section II of the module, students consider a special type of observational study—the survey. As you saw in the previous section, random selection plays an important role here in that it is likely to produce a "representative" sample of some larger population of interest. As with all observational studies, when random selection is used to obtain the sample that will participate in a survey, it is possible to generalize results from the sample to the larger population with confidence. Without random selection, our ability to generalize is limited.

There are three investigations in this section.

Investigation #7: Welcome to Oostburg!

Students use data from a census to answer questions about the corresponding population and then begin to explore the idea of sampling from a population by considering data from a subset of the population.

Investigation #8: Student Participation in Sports

Students draw conclusions from survey data and consider how the way in which a sample is selected affects the conclusions that can be drawn from the resulting survey data.

Investigation #9: Planning and Conducting a Survey—A Class Project

In this culminating investigation, students design, implement, analyze data from, and draw conclusions from a survey to investigate the tooth-brushing behavior of students at their school.

Prerequisites

Students should be able to:

Use census data to compute simple probabilities

Apply proportional reasoning

Compute simple percentages

Construct and interpret a bar graph

Construct and interpret a dotplot

Describe the center, shape, and spread of the distribution of a numerical variable

Compute numerical summary measures, such as the mean and median

Learning Objectives

As a result of completing this section, students should be able to:

Distinguish between a census and a sample

Use survey data to evaluate claims about a population

Explain the importance of random selection when choosing a sample

Explain when it is reasonable to generalize from a sample to the population from which it was selected

- Extract information from a comparative bar graph
- Explain why it is sometimes best to make group comparisons based on percentages, rather than counts or frequencies
- Write simple, clear, and unambiguous survey questions
- Develop a reasonable method for obtaining a representative sample of students from their school
- Develop and implement a sampling plan
- Summarize survey data graphically using bar charts and dotplots
- Summarize survey data numerically using measures such as counts and percents for categorical data, or the mean and median for numerical data

Teaching Tips

Throughout the investigations of this section, be sure to emphasize the dangers associated with convenience sampling.

Stress the role of random selection in good sample selection plans.

Encourage students to articulate why they think a particular sampling plan that is being proposed would be likely to result in a sample that could be considered representative of the population.

Obtain administrative permission before allowing your students to conduct the survey in Investigation #9.

Remind students about the ethics of data collection using surveys:

- Every individual has the right to refuse to participate in a survey
- Preserve the anonymity and confidentiality of individual responses

Possible Extensions

You might want to introduce students to other methods of sampling that involve random selection, such as stratified sampling, cluster sampling, systematic sampling, and multi-stage sampling.

You might also want to conduct a class discussion on the dangers associated with surveys that rely on volunteers or those who self-select to participate in the survey. For example, you might have the class think about the possible problems in generalizing the results of a survey about Internet use that is conducted online or a survey to assess opinion on a controversial topic where people must make a telephone call to participate in the survey.

Section II: Surveys

Corresponds to pp. 44-45
in Student Module

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.

Teacher Notes for Investigation #7: Welcome to Oostburg!

This investigation introduces students to survey data resulting from a census of a small town. Students use the data to compute various probabilities and to evaluate the accuracy of claims that have been made about the town. In the second part of the activity, students are given survey data resulting from a sample from the population and are asked to compare results based on this sample to population results. Finally, students are led through a discussion of why convenience samples are not a good way to collect data if the goal is to generalize to the larger population.

Prerequisites

Students should be able to:

- Use census data to compute simple probabilities
- Apply proportional reasoning

Learning Objectives

As a result of completing this investigation, students should be able to:

- Distinguish between a census and a sample
- Use survey data to evaluate claims about a population
- Explain the importance of random selection when choosing a sample
- Explain when it is reasonable to generalize from a sample to the population from which it was selected

Teaching Tips

When discussing question 2, you may need to help students see how they can use proportional reasoning to compute the desired estimate. You can give an example like the following: There are 38 + 46 = 84 people in the 41 to 60 age group. If we were interested in how many were between 41 and 45 years old, a reasonable estimate would be ¼(84) = 21, because 41 to 45 is about ¼ of the 41 to 60 age range.

Question 3 is a good place to remind students that they can use what they know about sets (in particular about the complement of a set) to help them compute the desired probability.

Question 3 is also a good place to remind students that there is a difference between the probability that a randomly selected person is male and did not watch "The Simpsons" and the probability that a randomly selected male did not watch "The Simpsons." The latter probability is a conditional probability (the probability of not watching "The Simpsons" *given* that the selected person is male), and the two probabilities are not the same.

Encourage group discussion in the evaluation of the headlines in question 5. You might begin by asking how many students think a particular headline is accurate and how many think it is not accurate. Then choose one student on each side to argue their case. A consensus should then develop about each headline.

Be sure to have a class discussion about the dangers associated with convenience sampling. Ask students to think of a good way and a poor way of selecting a sample of students at your school if they plan to use the resulting sample to estimate how much time students spend in the school library.

Suggested Answers to Questions

1. (a) The total number who attended a movie is 4 + 5 + 16 + 22 + 12 + 22 + 4 + 7 = 92. Because the data represent a census of the 306 people in the town,

$$P(\text{selected person attended a movie}) = 92/306 = .301$$

(b) The number who attended a movie and who are also between 18 and 40 years old is 16 + 22 = 38, so

$$P(\text{selected person attended a movie and is between 18 and 40 years old}) = 38/306 = .124$$

(c) The total number of males is 36 + 32 + 38 + 32 = 138, so

$$P(\text{male}) = 138/306 = .451$$

(d) The total number who are between 18 and 60 years old is 32 + 35 + 38 + 46 = 151, so

$$P(\text{selected person is between 18 and 60}) = 151/306 = .493$$

2. You might first want to address why this estimate is more difficult. The age range 10 to 30 does not match the age groups used to summarize the survey data. As a result, we do not know the exact number of people in this range and can only estimate. A student may indicate that any guess is possible, but a more acceptable answer (one used in applications involving census data) is to determine a reasonable estimate of the number of people in the age range based on the totals for the reported categories. For this question, the 10- to 17-year-old age group is very close to half the ages represented in the category "17 and younger," so an estimate of one-half of the 77 people in the 0–17 age range would be a reasonable estimate (or, approximately 35 to 40 people). Similarly, 18 to 30 years is approximately half the number of ages in the range of 18 to 40. (Some students might be more precise with this proportion.) Therefore, an approximation based on half of the 67 people in that age range would be reasonable (or, 30 to 35 people). If we put these estimates together, we get an estimate of 65 to 75 people in that age range. This would indicate a probability of approximately .19 to .23. As all answers are estimates, evaluate students' thinking based on their use of proportional reasoning.

3. There are a total of 138 males in the town. Of these, 23 + 29 + 12 + 1 = 65 reported watching "The Simpsons." So, 138 – 65 = 73 males did not watch "The Simpsons."

$$P(\text{selected person is male and did not watch "The Simpsons"}) = 73/306 = .239$$

4. Answers will vary depending on questions posed.

5. (a) This is an accurate headline as 162 of the eligible voters turned out for the last election. Since there are 229 eligible voters, 162/229, or approximately .707 or 71% of the eligible voters, indicated they voted.

(b) This is also an accurate headline, as one out of 78 people in the over 60 age category indicated that they shopped for clothes online (or approximately 1%), compared to 54 out of 67 people (or approximately 81%) in the "18–40 years old" category.

(c) Most interpretations would indicate that this is not accurate based on a comparison of the proportion who attended a movie for each age group, or

0–17 years old: 9/77 or ≈ 11.7%

18–40 years old: 38/67 or ≈ 58%

41–60 years old: 34/84 or ≈ 41%

Over 60 years old: 11/78 or ≈ 14.1%

The variation in percents is rather large, indicating that the generations represented by these age categories are not attending movies in the same way.

However, if the headline is interpreted as simply stating that people from each generation attend movies, then yes, the headline is accurate. Ask students how they would interpret this headline.

(d) Most interpretations would indicate that this headline is accurate. Comparing the "18 to 40 years old" group to the "41 to 60 years old" group would be a good indicator to support the headline, or

61/67 (≈ 90%) ate fast food in the "18 to 40 years old" group

13/84 (≈ 16%) ate fast food in the "41 to 60 years old" group

(e) Most interpretations would indicate that this is accurate. Comparing the "over 60 years old" group to the "18–40 years old" group would support this answer:

3/78 (≈ 4%) watched "The Simpsons" in the "over 60 years old" group, compared to 60/67 (≈ 90%) in the "18–40 years old" group.

(f) This is clearly not an accurate headline. 72/78 (≈ 92%) of the people over 60 voted in the last election. This is not close to 40%.

(g) Most interpretations of this question would indicate that it is accurate. This would be based on the variation in the percent of people in each age category who indicated that they eat out:

0–17 years old: 13 out of 77 people (≈ 17%)

18–40 years old: 61 out of 67 people (≈ 90%)

41–60 years old: 13 out of 84 people (≈ 15%)

Over 60 years old: 7 out of 78 people (≈ 9%)

6. Probably not. Oostburg has gender and age distributions that are not similar to the country as a whole.

7. Yes, Hugo was accurate, as 28 people responded yes to the question. Therefore, 28/40 = 70%.

8. No. Only 77/306 (or approximately 25.2%) of the total town indicated that they shopped online. The difference in the sample Hugo collected and the census data indicates Hugo is not accurately summarizing the town. This is very likely due to the method Hugo used to select his sample.

9. As the percent of people who shopped online from the sample is similar to the percent of the age category of 18 to 40 years old in the census data, it is likely the band is made up primarily of people who are 18 to 40 years old. This would make sense if the band in which Hugo is a member is a high-school band (as opposed to a community band, which might have members from all age categories).

10. With random selection, we can be confident that the resulting sample will be representative of the population.

11. Answers will vary. Good answers will employ some form of random selection and will provide a convincing argument that the proposed method will result in a representative sample.

Investigation #7: Welcome to Oostburg!

Corresponds to pp. 46-50 in Student Module

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.

Teacher Notes for Investigation #8: Student Participation in Sports

In this investigation, students use survey data from a random sample to investigate the plausibility of statements about a population. Students are also asked to think about why methods of choosing a sample that do not use random selection may result in an unrepresentative sample. Finally, the idea of sampling variability is introduced, setting the stage for an introduction to inference in Section IV.

Prerequisites

Students should be able to:

- Interpret a bar graph
- Compute simple percentages

Learning Objectives

As a result of completing this investigation, students should be able to:

- Extract information from a comparative bar graph
- Use survey data to evaluate claims about a population
- Explain the importance of random selection when choosing a sample
- Explain why it is sometimes best to make group comparisons based on percentages, rather than counts or frequencies

Teaching Tips

Focus the discussion of question 6 answers on why the proposed methods, which do not involve random selection, may yield a sample that is not representative of the population.

Question 9 is a good place to introduce the idea of sampling variability. Even if 50% of the students at the school are female, we would not expect the percentage of females in a sample to be exactly 50%.

Question 10 begins to hint at the ideas that will be formalized in Section IV. Answering this question requires an informal assessment of whether a sample percentage of 66% could have resulted due to just sampling variability when the population percentage is 50%. Here, appeal to the students' intuition. With random selection and a sample size of 50, a sample percentage of 66% is unlikely to have occurred by chance. Point out that more formal methods for making such assessments will come in Section IV.

In the discussion of questions 11 to 13, make the case for using percentages to compare groups, rather than making the comparison based on raw frequencies.

Suggested Answers to Questions

1. 18

2. 15

3. 12

4. 5

5. Answers will vary.

6. (a) Answers will vary, but should address possible reasons why such a sample might not be representative of the population. For example, maybe some sports teams practice before classes start in the morning, so athletes may be over-represented in a sample that consists of the first 50 students to arrive on campus.

(b) Answers will vary, but should address possible reasons why such a sample might not be representative of the population. For example, maybe the two sections of pre-calculus are offered in the afternoon only, when athletes are typically required to travel to athletic events.

(c) Answers will vary, but should address possible reasons why such a sample might not be representative of the population. For example, it is more likely that students in the weight room would be athletes, making this sample not representative of the general student population.

7. 33/50 = .66 = 66%

8. 17/50 = .34 = 34%

9. (1) There may be more females than males at the school, and (2) the percent of females in a sample will not necessarily be exactly equal to the actual percent of females in the population (sampling variability).

10. No. There were twice as many girls in the sample as there were boys. Because the sample was randomly selected and the sample size was 50, it is not likely that we would see twice as many girls in the sample if the number of boys and the number of girls at the school were about the same.

11. Yes. In the sample, there were 18 girls and only 12 boys who reported participating in sports.

12. The greater number of girls in the sample who participated in sports could just be due to there being more girls in the sample. The proportion of girls in the sample who participated in sports is 18/33 = .55. For boys, this proportion is 12/17 = .71.

13. There were more girls in the sample who reported participating in sports (18 for girls versus 12 for boys). However, the greater number of girls in the sample who participated in sports could just be due to there being more girls in the sample. The proportion of girls in the sample who participated in sports is 18/33 = .55. For boys, this proportion is 12/17 = .71.

14. Answers will vary.

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.

Corresponds to pp. 51-54 in Student Module

Student Survey Finds Females More Involved in Sports

By Kayla Johnson

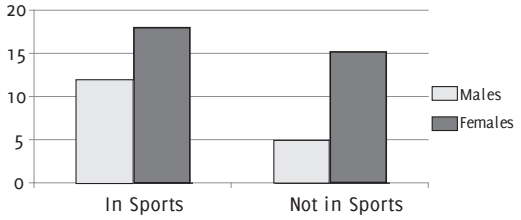
A Rufus King mathematics class conducted a survey to investigate student participation in extra-curricular activities. Fifty randomly selected students participated in the survey.

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.

Teacher Notes for Investigation #9: Planning and Conducting a Survey

In this culminating investigation for Section II, students develop and implement a sampling plan to investigate two aspects of tooth-brushing behavior—the reported length of time that students spend brushing and whether students leave water running while brushing.

Prerequisites

Students should be able to:

- Construct and interpret a bar graph
- Construct and interpret a dotplot
- Describe the center, shape, and spread of the distribution of a numerical variable
- Compute numerical summary measures, such as counts and proportions for categorical data, or the mean and median for numerical data

Learning Objectives

As a result of completing this investigation, students should be able to:

- Write simple, clear, and unambiguous survey questions
- Develop a reasonable method for obtaining a representative sample of students from their school
- Develop and implement a sampling plan
- Summarize survey data graphically using bar charts and dotplots
- Summarize survey data numerically using measures such as the mean and median

Teaching Tips

Obtain permission from your administration, if needed, before having your students carry out the survey.

Remind students about the ethics of survey data collection, including an individual's right to "opt out" and the importance of preserving respondents' anonymity and confidentiality.

In the discussion of the survey questions, focus on whether the questions are clear and unambiguous and whether they will elicit the desired information. For example, the question about brushing time should specify that the response should be in seconds.

Sometimes it is difficult to obtain a true random sample from a population. In this investigation, it may be impractical to select at random from a list of students at the school and then somehow find the selected students to participate in the survey. So, other methods may need to be considered. In a similar situation, for example, a student suggested the following plan:

> Post students at each entrance to campus and then (1) use a random number generator or table of random numbers to obtain a random number between 1 and 19, (2) use the random number generated in step 1 to determine the next student to be approached

(for example, if the random number is 8, the students would let the next 7 students go by and then approach the next student—the 8th student—to participate in the survey), (3) repeat steps 1 and 2, each time generating a new random number.

While not an "official" random sample, this method uses a form of random selection and would be a reasonable alternative.

The class discussion of question 13 is a good place to revisit the idea of sampling variability by talking about how the sample mean will vary somewhat from sample to sample.

Suggested Answers to Questions

1. Answers will vary.

2. Answers will vary.

3. Answers will vary.

4. Answers will vary. It may not be practical to select a true random sample, but some sort of random selection should be employed. For an example, see the Teaching Tips above.

5. Answers will vary.

6. Answers will vary.

7. Answers will vary. Answers should appeal to the way random selection was employed and should argue why the student believes the sampling plan is likely to produce a representative sample.

8. Answers will vary. Even if students leave the water on, they may be tempted to say they turn if off because they know this is the right thing to do from a conservation standpoint.

9. Answers will vary. Answers should comment on how the two proportions (for females and for males) compare.

10. Answers will vary. Answers should comment on center, shape, and spread.

11. Answers will vary. Answers should be based on the dotplot and whether the responses for females tend to fall to the right end of the dotplot or whether they are intermingled with the responses from males.

12. Answers will vary.

13. Answers will vary.

14. Answers will vary. Students will need to think about how to measure actual brushing time. This may be difficult to do, as having an observer present or asking someone to time his or her brushing time may actually influence behavior.

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.

Corresponds to pp. 55-61
in Student Module

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.