And the Survey Says . . .

How can television newscasters predict the next president of the United States long before the votes are counted? In this module, you explore some of the basics of sampling and surveys.

Staci Auck • Kyle Boyce • Tom Teegarden

© 1996-2019 by Montana Council of Teachers of Mathematics. Available under the terms and conditions of the Creative Commons Attribution NonCommerical-ShareAlike (CC BY-NC-SA) 4.0 License (https://creativecommons.org/licenses/by-nc-sa/4.0/)

And the Survey Says . . .

Introduction

Opinion polls have become a familiar feature of American life. You read about them in magazines and newspapers, and hear them discussed on radio and television. Such surveys are administered by a number of different polling organizations. Some—like Gallup, Roper, and Harris—are independent companies. Others, like the New York Times and CBS, are members of the news media themselves.

The pollsters question you about everything from religion to politics, from your views on health care to your preferred brand of toothpaste. The information they obtain has many applications. A candidate for the U.S. Senate might use polls to plan a campaign strategy. A Hollywood producer might use audience surveys to edit an upcoming movie. A cookie manufacturer might use consumer taste preferences to target a profitable new market.

How do polling organizations discover the likes and dislikes of a nation of more than 250 million? Fortunately for the pollsters, reliable results can be obtained by surveying only about 1500 people—a small percentage of the entire population. Why is this sample big enough? This module will help you answer that question.

Activity 1

Whether you watch television, listen to the radio, or read newspapers, what you get from the news often comes in the form of statistics. Statistics arise from the collection of data. In the case of opinion polls, data is collected by surveying small samples of the entire population. In this activity, you investigate some methods pollsters use to determine whom to survey and examine some of the factors that might prevent a sample from representing the entire population.

Mathematics Note

A **population** is a group of all objects, individuals, or observations about which information is to be gathered. A quantity that describes a population is a **parameter**.

A **sample** is a portion of the population. A quantity that describes a sample is a **statistic**.

A **bias** is a factor that prevents a sample from representing the entire population.

A **census** is the collection of data about an entire population.

For example, a pollster's study of voter preferences for an upcoming presidential election might describe the responses from all registered voters as a *population*. The percentage of voters who intend to vote for a particular candidate is a *parameter*.

A *sample* of the population might include 1500 registered voters who answered their home telephones on a Monday morning. The percentage of those surveyed who intend to vote for a particular candidate is a *statistic*.

The fact that most working people are not home on weekday mornings would be a *bias* of this sample, since these people would not have a chance of being selected. To perform a *census* of the population, the polling organization would have to question every registered voter in the country—a nearly impossible task.

Exploration

In this exploration, you describe the characteristics of a population of beans based on the data obtained from samples. **Note:** Do not examine the contents of your container before beginning the exploration.

a.
1. Shake the container containing the population of beans.
2. Take a small handful of beans from the container. This is one sample of the population.
3. Record the percentage of each kind of bean in the sample.
4. Return the sample to the container and mix it thoroughly with the other beans.

b. Repeat Part **a** four more times.

c. Based on the data collected in your five samples, predict the percentage of each kind of bean in the entire population.

d. Take a census of the beans. Determine the percentage of each kind of bean in the population. Compare the actual percentages to your predictions in Part **c**.

Discussion

a.
1. Which values in the exploration are parameters?
2. Which values are statistics?

b. How did the results of the census compare with your predictions based on five samples?

c. Describe any biases in the selection process that might have prevented samples from accurately representing the entire population.

d. How could you eliminate bias in this sampling method?

e. What kinds of bias could affect a real-life survey?

f. Consider the following survey question, "Would you prefer the powerful and graceful eagle or the clumsy bison as the school mascot?" How could you reword this question to minimize possible bias?

Assignment

1.1 Some television programs offer their audiences a chance to express opinions on everything from favorite rock videos to favorite presidential candidates. Viewers typically dial a phone number to register their votes. Is such a survey biased? If so, describe the biases that might exist. If not, explain why not.

1.2 To help determine consumer preferences, an electronics company conducts a poll in a shopping mall. The pollsters stand outside a department store and ask all the people who pass by if they own a videocassette recorder (VCR). Describe any biases this sample might contain.

1.3 Washington High School has 2380 students: 500 seniors, 597 juniors, 626 sophomores, and 657 freshmen. Elections for student government are coming up next month. In order to predict the next student body president, the school newspaper has asked you to take a preliminary poll.

a. What is the population for this poll?

b. Describe how to conduct a census of the population.

c. Describe one method for sampling the population.

d. What advantages or disadvantages are there to the sampling method you described in Part **c**?

1.4 In 1936, incumbent Franklin Delano Roosevelt was campaigning against Alf Landon for the presidency of the United States. Shortly before the election, pollsters for the *Literary Digest* conducted a survey. They sent postcard ballots to people selected at random from a list of automobile registrations and from a telephone directory. Their results indicated that Landon would win the election. That prediction was incorrect. In a landslide victory, Franklin Roosevelt won all but two states—Maine and Vermont.

a. What biases in the poll may have caused the incorrect prediction?

b. How might the pollsters have obtained a more accurate picture of the nation's preferences?

1.5 According to pollster George Gallup, "Nothing is so difficult, nor so important, as the selection and wording of questions." In fact, poor word choice and misleading questions have biased many surveys. Write a survey question that introduces bias due to its own wording and explain the cause of that bias.

* * * * *

1.6 Jordan is conducting a survey of music preferences at his school using the following questionnaire:

Which of the following music types do you prefer?

(1) rock and roll	(2) alternative
(3) country western	(4) classical
(5) heavy metal	(6) other

To obtain a sample of 50 people, he polls 20 students in his mathematics class, 20 friends and acquaintances sitting nearby at lunch, and 10 fellow band members at an after-school practice session.

a. Is Jordan's sample representative of the entire school population?

b. Describe any biases that might exist in his survey question or sampling techniques.

c. Suggest how Jordan could change the sampling method to minimize bias.

1.7 For an article in the school paper, Susanne wants to survey students on a proposal to open the campus for lunch. This has been a hot topic for weeks. If the school board approves the proposal, the lunch period will be lengthened from 30 minutes to 45 minutes, which in turn will add 15 minutes to the end of the day. She plans to ask the following question:

"Do you want to have an open campus for the lunch period in order to go anywhere you want to eat instead of being confined to the school cafeteria?"

a. What do you think is Susanne's opinion on this proposal?

b. Write a question that shows a bias for the opposing view.

c. Write a question on this proposal that minimizes bias.

* * * * * * * * * *

Research Project

Select and complete one of the projects described below.

a. Find a report of an interesting survey in a magazine or newspaper. Write a summary of its results, including a description of the population, the sample size, and the sampling method. Identify potential sources of bias in the survey and suggest how these biases might have been avoided.

b. Identify a poll or survey (other than the one mentioned in Problem **1.4**) that produced an incorrect prediction. Describe some of the biases that may have caused the mistake. Present your findings to the class.

Activity 2

In order to design reliable surveys, many polling organizations work hard to remove biases from both their samples and their survey questions. At the same time, however, they must also try to control the costs of administering the poll. Selecting an appropriate sampling method can help achieve all of these goals. In this activity, you examine several sampling methods that can reduce polling costs, simplify administration, and minimize bias.

Exploration

The map in Figure **1** shows a small town located at the intersection of two highways. The town is divided into 40 different districts, each identified by a number. On this map, the area of each district indicates its population (1 square unit represents 100 people).

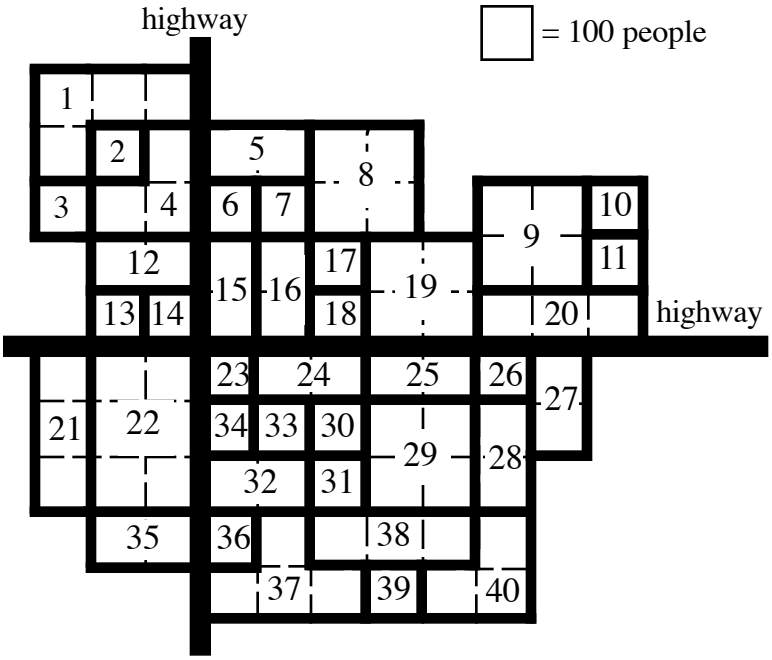


Figure 1: Population of a town, by district

a. Select a sample of 4 of the 40 districts from the map in Figure **1**.

1. Determine the population of each district in your sample.

2. Using only your sample, predict the mean number of people per district for the entire town.

3. Using only your sample, estimate the population of the entire town.

b. Determine the population of the entire town. Compare the actual value with your prediction in Part **a**.

Mathematics Note

A **simple random sample** is selected so that each member of the population has the same chance of being included in the sample.

For example, consider a population of students in a classroom. To obtain a simple random sample of this population, each student's name could be written on an identical slip of paper. The slips of paper could then be placed into a bowl, mixed thoroughly, and drawn one at a time.

c. **1.** Randomly generate four different integers in the interval [1, 40]. Use these integers to identify four districts on the map in Figure **1**.

2. Use this sample to predict the mean number of people per district.

3. Use this sample to estimate the population of the entire town.

Mathematics Note

Stratified sampling requires that a population be divided into parts. Each part is a **stratum** (plural, **strata**). To produce a stratified sample, simple random samples are taken from each stratum. These samples do not necessarily have to be the same size.

For example, differences in age could determine the strata of a voting population. To select a stratified sample of this population, a polling organization might survey groups of people with ages 18–28 years, 29–45 years, 46–60 years, and 61 years or older.

Systematic sampling is accomplished by collecting data from every *n*th unit of a population after randomly choosing a starting point.

For example, a systematic sample of the customers in a shopping mall could be obtained by selecting every fifth person, beginning with the first person who enters after 10:00 A.M.

d. **1.** Devise a stratified sampling method to obtain a sample of four districts from the map in Figure **1**.

2. Use your stratified sample to predict the mean number of people per district and to estimate the population of the entire town.

e. **1.** Devise a systematic sampling method to obtain a sample of four districts from the map in Figure **1**.

2. Use your systematic sample to predict the mean number of people per district and to estimate the population of the entire town.

Discussion

a. Compare your results in the exploration with those of others in your class.

Is it reasonable to expect population estimates made using the same sampling technique to be similar? Why or why not?

b. You used four different sampling methods in the exploration: sampling based on intuition, simple random sampling, stratified sampling, and systematic sampling. Where might bias occur in each of these methods?

c. Which method has the most potential for bias? Which has the least potential for bias? Explain your responses.

Assignment

2.1 Parts **a–f** each describe a sampling method. Identify those that produce a simple random sample. If a method does not produce a simple random sample, explain why this occurs.

a. Select the first 20 students who enter the cafeteria for lunch.

b. Assign a number to each student in the school, then use a random number generator to select 20 students.

c. Assign a number to each student in the school, then use a random number generator to select 10 boys and 10 girls.

d. Write all student identification numbers on identical slips of paper, place the slips in a bin, mix them up, and select 20 from the bin.

e. Select the first person who walks into a pep assembly and every 10th person thereafter.

f. Select only those students whose school identification numbers end with a 0.

2.2 Which of the sampling methods described in Problem **2.1** represent stratified sampling? Which represent systematic sampling?

2.3 **a.** The editors of a school paper want to survey student preferences for an upcoming election. The school has 400 students. Suggest one way to obtain a sample of 20 students from this population using each of the following methods:

1. simple random sampling

2. stratified sampling

3. systematic sampling

b. Identify potential sources of bias in each sampling technique you suggested in Part **a**.

2.4 Identify an issue of concern at your school. Devise a method of surveying student opinion on this issue.

a. Describe your poll in a paragraph, including a list of survey questions and an explanation of your sampling technique.

b. Describe both some advantages and some disadvantages of the method you devised.

2.5 Write a paragraph describing a situation in which a biased sample could be used to influence people. Identify the source of bias and suggest why this influence might be considered misleading.

* * * * *

2.6 Consider a district that contains 1400 registered voters, 748 of which are women. Describe how you could obtain a sample of 100 registered voters from this population using each of the following methods:

a. systematic sampling

b. simple random sampling

c. stratified sampling

2.7 A car manufacturer is conducting a marketing study. The proposed sampling method consists of interviewing customers who request test drives at a dealership.

a. Will this sampling method produce a simple random sample of all potential car buyers?

b. Describe the advantages and disadvantages of the proposed method.

* * * * * * * * *

Activity 3

Amelia and Bernadette are both running for president of Washington High School. Two weeks before election time, the school newspaper surveyed the entire school population of 2380 students. According to the results of this census, 40% of the students favored Amelia, 30% preferred Bernadette, 20% were undecided, and 10% did not plan to vote.

With the election approaching quickly, Amelia felt that she had a comfortable lead. A week after the census, however, she took a strong stand in favor of an unpopular dress code. Concerned that Amelia might have lost some support, her campaign staff surveyed a random sample of 25 students. Of those polled, only 8 students planned to vote for Amelia.

Now Amelia is worried. Do the results of this most recent poll mean her support has slipped to 8/25 or 32%? Her campaign manager points out that these are the findings of only one sample. Another sample might show completely different results. "But if that's true," says Amelia, "what's the use of taking any samples at all?" Before drawing any conclusions, Amelia and her friends decide to investigate the sampling process.

Exploration

If 40% of Washington High students plan to vote for Amelia, how likely is it for a random sample of 25 students to show only 32% support? In this exploration, you use simulations to model the results of such samples.

a. The percentage of a population with a given characteristic is a **population proportion**. Create a population of beans in which 40% have a distinguishing characteristic other than size or shape.

In this model, the marked beans represents students who planned to vote for Amelia. The remaining 60% represents students who did not indicate that they planned to vote for Amelia.

b. In a simple random sample of 25 beans from this population, about how many would you expect to be marked?

c. The percentage of a sample with a given characteristic is a **sample proportion**. Take a random sample of 25 beans from the container. Record the percentage of marked beans in the sample, then return the sample to the container and mix the beans thoroughly.

d. Repeat Part **c** nine more times, for a total of 10 samples.

e.

1. Combine the results of your 10 samples with the results of the rest of your class and sort the data from least to greatest.

2. Determine the mean of the sample proportions.

3. Display the class data in a histogram, using intervals with a width of 4%. Represent the percentage of marked beans on the x-axis and the frequency on the y-axis.

4. Find the percentage of sample proportions that are less than or equal to 32%.

f.

1. Use technology to simulate Parts **a**–**d** of the exploration. Generate as many samples as the technology allows, not to exceed the number of samples in the class data from Part **e**.

2. Determine the mean of the sample proportions.

3. Create a histogram of the data using the same intervals as the histogram in Part **e**.

Discussion

a. What patterns did you observe in the two histograms created in the exploration?

b. How does the mean of the sample proportions compare with the population proportion?

c. **1.** If the population proportion had been 52% instead of 40%, how would the histograms have been affected?

2. What would the histograms have looked like if twice as many samples were taken?

d. **1.** In the class data, what percentage of the sample proportions were less than or equal to 32%?

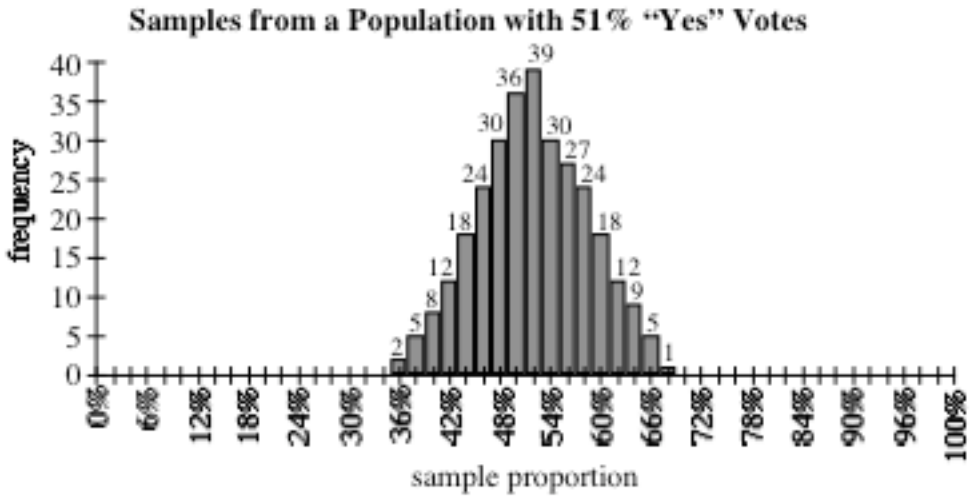
2. Given this percentage, should Amelia be concerned about a loss of student support? Explain your response.

e. How might you change the sampling method described in Part **c** of the exploration to obtain a more accurate characterization of the population?

Assignment

3.1 In order for your community to build a new swimming pool, 51% of the voting population must approve its construction. Before hiring an architect, the local council decided to assess community support by surveying a random sample of 25 voters. Of those surveyed, 9 people indicated that they would vote for the new pool.

To determine what conclusions they could draw from this sample, the council used a computer simulation to collect numerous samples of size 25 from a population in which 51% favored a proposal. The following histogram shows the results of this simulation.



a. What percentage of the people surveyed indicated that they would vote for the pool?

b. What is the total number of samples in the simulation?

c. What is the mean of the sample proportions in the simulation?

d. Should the town council hire an architect? Write a recommendation to the council that supports your position.

3.2 In the election for president of Washington High School, 10% of the students do not intend to vote.

a. What percentage of the school population would have to vote for Amelia in order for her to win? Explain your response.

b. **1.** Use a simulation to obtain 90 samples of size 25 from a population that would narrowly elect Amelia.

2. Create a histogram of this data.

c. When Amelia's campaign crew polled a sample of 25 students, they found that only 8 planned to vote for Amelia. What is the estimated probability that, in a sample of 25 students, 8 or fewer would support Amelia?

d. Do you think Amelia has a good chance of winning the election? Justify your response.

* * * * *

3.3 A group of friends is playing a board game. In this game, players use a six-sided die to determine the number of spaces they may move around the board. Because 4 of the first 5 players have each rolled a six, the group suspects that the die may not be fair.

a. Create a simulation to model the roll of a fair die. Use the simulation to obtain 300 samples of size 5 and determine the number of sixes in each sample.

b. Create a histogram to display the results of the simulation.

c. Use the histogram to estimate the probability of obtaining a six in 4 out of 5 rolls with a fair die.

d. Do you think that the die in the board game is fair? Explain your response.

* * * * * * * * * *

Activity 4

Amelia was relieved to know that her popularity might not have slipped. With 20% of voters still undecided and the election just a few days off, she felt that she still had a chance to win.

Her campaign staff, however, was beginning to have some doubts. In their previous survey of 25 students, only 32% indicated that they would vote for Amelia. What if they had surveyed a sample of 100 students? What would a 32% rating have meant then?

As the staff discussed these questions, they realized that selecting a random sample of size 25 from the Washington High population could result in many different combinations of students. In fact, for a population of 2380, there are about 10^{60} different samples of size 25. Some of these samples would reflect an accurate picture of Amelia's support, while others might give misleading information. How could they possibly be confident in their results?

The campaign staff paid a visit to their mathematics teacher. She told them that even though they could never survey all the possible samples, they could still obtain a reasonable estimate of Amelia's chances in the election using the mean and standard deviation of all possible sample proportions.

Mathematics Note

The mean of all possible sample proportions for samples of size n equals the population proportion p.

Standard deviation is a measure of the spread in a data set. For a population consisting of all possible sample proportions for a given sample size, the standard deviation can be calculated using the following formula:

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

where p is the population proportion and n is the sample size.

For example, if 40% of Washington High students support Amelia, the population proportion p is 0.4. For samples of size 20, the standard deviation of all possible sample proportions is:

$$\sigma = \sqrt{\frac{0.4(1-0.4)}{20}} \approx 0.11 = 11\%$$

Exploration 1

In this exploration, you investigate the effects of sample size on the reliability of the resulting data.

a. Assume that 40% of the Washington High student population still supports Amelia.

1. Use a simulation to obtain 90 samples of size 25 from this population. Record the proportion of favorable votes in each sample.

2. Sort the data from least to greatest.

3. Create a histogram of the sample proportions.

b. Using the formula given in the mathematics note, calculate the standard deviation of all possible sample proportions for the simulation in Part **a**.

1. Determine the percentage of sample proportions from Part **a** that lie within 1 standard deviation of the population proportion.

2. Determine the percentage of sample proportions that lie within 2 standard deviations of the population proportion.

c. Repeat Parts **a** and **b** using a sample size of 50.

d. Repeat Parts **a** and **b** using a sample size of 100. **Note:** Save this data for use in Exploration **2**.

Discussion 1

a. Compare your histogram from Part **a** of Exploration **1** with the class histogram from Activity **3**.

b. Describe how changing the sample size affected each of the following:

1. the shape of the histograms

2. the standard deviation of all possible sample proportions

3. the interval that represents values within 1 standard deviation of the population proportion.

c. Statisticians have found that, for reasonably large sample sizes, approximately 68% of sample proportions fall within 1 standard deviation of the population proportion, while approximately 95% fall within 2 standard deviations of the population proportion.

1. How do these values compare with the percentages you found in Exploration **1**?

2. What do these percentages indicate about the results of a single sample?

d. If the campaign staff found that 32% of a sample of 100 students planned to vote for Amelia, what should they conclude? Explain your response.

Exploration 2

As election day approaches, the number of undecided voters grows smaller and smaller. Some choose Amelia, while others throw their support to Bernadette. Amelia's campaign staff decides to take one last poll, this time with a larger sample size.

One thing worries them, however. In the two weeks that have passed since the school-wide census, the proportion of students who favor Amelia almost certainly has changed. The campaign staff knows that no matter what the level of her support, the sample proportion is likely to be within 2 standard deviations of the population proportion. The hard part involves calculating this standard deviation, since its formula depends on knowing the population proportion.

Without knowing the actual population proportion, how could they draw a reliable conclusion from their sample? In this exploration, you determine how confident you can be in the results of a single sample.

a. The formula for the standard deviation of all possible sample proportions, where p is the population proportion and n is the sample size, is shown below:

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

1. Graph this equation for a sample size of 100 ($n = 100$).

2. Use the graph to determine the largest possible standard deviation for a sample size of 100.

3. What population proportion corresponds with this maximum standard deviation?

b. Repeat Part **a** for some other sample sizes. Use your results to make a general statement about the maximum standard deviation for any sample size.

c. In Part **d** of Exploration **1**, you simulated 90 samples of size 100 from a population with a known population proportion of 40%.

Add and subtract 1 maximum standard deviation (MSD) from each of these sample proportions to obtain 90 intervals. For example, given a sample proportion of 34%, the corresponding interval would be:

[34 – MSD, 34 + MSD]

d. Determine the number of intervals from Part **c** that contain the population proportion (40%).

e. Combine your results from Part **d** with those of the rest of the class. Determine the percentage of intervals that contain the population proportion.

f. Repeat Parts **c**–**e** using twice the maximum standard deviation (2 • MSD).

Discussion 2

a. Describe the domain and range of the following function, when $n = 100$:

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

b. Would you be more confident in predicting that an unknown population proportion is within 1 MSD of a sample proportion, or within 2 MSDs of a sample proportion? Explain your response.

c. In Part **c** of Exploration **2**, why do you think it was important to use the maximum standard deviation to determine each interval?

d. How do you think that changing the sample size would affect these intervals?

e. Would you be more confident in predicting an unknown population proportion using a sample size of 100 or a sample size of 400? Explain your response.

Mathematics Note

The **maximum standard deviation** of all possible sample proportions occurs when the population proportion $p = 0.5$. The value of 2 maximum standard deviations for all possible sample proportions equals:

$$2 \bullet \sigma = 2 \bullet \sqrt{\frac{0.5(1-0.5)}{n}} = \frac{2 \bullet 0.5}{\sqrt{n}} = \frac{1}{\sqrt{n}}$$

A **confidence statement** declares that a population parameter lies within a specific range of values. Given the value of 2 maximum standard deviations calculated above, the following confidence statement can be made, where n is the sample size: "It is highly likely that the population proportion is within $1/\sqrt{n}$ of the sample proportion."

For example, consider a random sample of 1500 registered voters. Of those surveyed, 700 plan to vote for candidate A. In this case, the sample proportion is $700/1500$ or 47%. The value of 2 maximum standard deviations is $1/\sqrt{1500}$ or 2.6%. Therefore, the following confidence statement can be made: "It is highly likely that the proportion of the population who plan to vote for candidate A is within the interval [44.4%, 49.6%].

The interval within which a parameter is likely to fall is often reported using a **margin of error**. Identifying a margin of error that is twice the maximum standard deviation provides an interval that is highly likely to contain the population proportion.

For example, the results of the survey described above might be reported as follows: "At the time of our poll, it was estimated that 47% of registered voters preferred candidate A, with a margin of error of 2.6%."

f. **1.** When using the maximum standard deviation to identify a margin of error, how is the margin of error affected by the size of the sample?

2. How is it affected by the size of the population?

g. Why do you think a pollster would be willing to predict the outcome of a national election based on a random sample of only 1500 registered voters?

Assignment

4.1 On the day before the election, Amelia's campaign staff surveyed a random sample of 400 students. Of those polled, 49% favored Amelia.

a. Write a confidence statement describing the true proportion of Amelia's supporters.

b. If you were Amelia's campaign manager, what would you say about her chances of winning the election? Explain your response.

4.2 A television commercial made the following claim: "Our survey shows that 4 out of 5 dentists prefer sugarless gum for their patients who chew gum." Can you make a confidence statement about this proportion? If so, make one. If not, explain why not.

4.3 A poll taken the day before a congressional election reported the following results: "Senator Rodriguez currently has 54% of the vote. The poll's margin of error is ± 3%." Is Senator Rodriguez assured of re-election? Explain your response.

4.4 For its annual election survey, one polling organization selects a random sample of 1000 registered voters. Using the maximum standard deviation of all possible sample proportions, will the survey's margin of error be more or less than three percentage points? Explain your response.

4.5 The editors of the Washington High newspaper surveyed a random sample of 100 students on the morning of the election. Of those surveyed, 56 planned to vote for Amelia. What predictions can you make about the election? Justify your response.

* * * * *

4.6 To assess customer satisfaction with its new checkout system, a supermarket surveyed a random sample of 40 shoppers. Of those surveyed, 83% were happy with the new system. In what interval is the actual proportion of satisfied customers likely to fall? Explain your response.

4.7 As part of its quality control process, a car dealership surveyed a random sample of 25 customers. Of those surveyed, 73% expressed satisfaction with the service they received. Using the maximum standard deviation of all possible sample proportions, what is the margin of error for this survey? Explain your response.

* * * * * * * * * *

Summary Assessment

Imagine that you are a quality control specialist for a light bulb manufacturer. Your job involves testing bulbs to make sure that the production line is working properly. When the equipment is functioning as designed, the proportion of defective bulbs is less than 5%. Since the factory produces many thousands of bulbs each day, you must use a sampling technique to estimate the actual proportion.

1. Design a practical method of sampling light bulbs so that the sample proportions provide a reliable estimate of the population proportion.

2. Use technology to create a simulation of the sampling process. Display the results in a histogram.

3. Determine the margin of error for a sample proportion obtained using your method.

4. Describe how you would respond to a sample proportion of 6.2% defective bulbs obtained using your method.

Module Summary

- A **population** is a group of all objects, individuals, or observations about which information is to be gathered. A quantity that describes a population is a **parameter**.
- A **sample** is a portion of the population. A quantity that describes a sample is a **statistic**.
- A **bias** is a factor that prevents a sample from representing the entire population.
- A **census** is the collection of data about an entire population.
- A **simple random sample** is selected so that each member of the population has the same chance of being included in the sample.
- **Stratified sampling** requires that a population be divided into parts. Each part is a **stratum** (plural, **strata**). To produce a stratified sample, simple random samples are taken from each stratum. These samples do not necessarily have to be the same size.
- **Systematic sampling** is accomplished by collecting data from every *n*th unit of a population after randomly choosing a starting point.
- The percentage of a population with a given characteristic is a **population proportion**.
- The percentage of a sample with a given characteristic is a **sample proportion**.
- The mean of all possible sample proportions for samples of size n equals the population proportion p.
- **Standard deviation** is a measure of the spread in a data set. For a population consisting of all possible sample proportions for a given sample size, the standard deviation can be calculated using the following formula:

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

  where p is the population proportion and n is the sample size.
- The **maximum standard deviation** of all possible sample proportions occurs when the population proportion $p = 0.5$. The value of 2 maximum standard deviations for all possible sample proportions equals:

$$2 \bullet \sigma = 2 \bullet \sqrt{\frac{0.5(1-0.5)}{n}} = \frac{2 \bullet 0.5}{\sqrt{n}} = \frac{1}{\sqrt{n}}$$

- A **confidence statement** declares that a population parameter lies within a specific range of values.
- The interval within which a parameter is likely to fall is often reported using a **margin of error**.

Selected References

Gallup, G. *The Sophisticated Poll Watcher's Guide*. Ephrata, PA: Princeton Opinion Press, 1972.

Landwehr, J. M., J. Swift, and A. E. Watkins. *Exploring Surveys and Information from Samples*. Palo Alto, CA: Dale Seymour Publications, 1988.

Mizrahi, A., and M. Sullivan. *Finite Mathematics with Applications for Business and Social Sciences*. New York: John Wiley & Sons, 1973.

Moore, D. S. *Statistics: Concepts and Controversies*. New York: W. H. Freeman and Co., 1991.

Sincich, T. *Statistics by Example*. San Francisco, CA: Dellen Publishing Company, 1982.