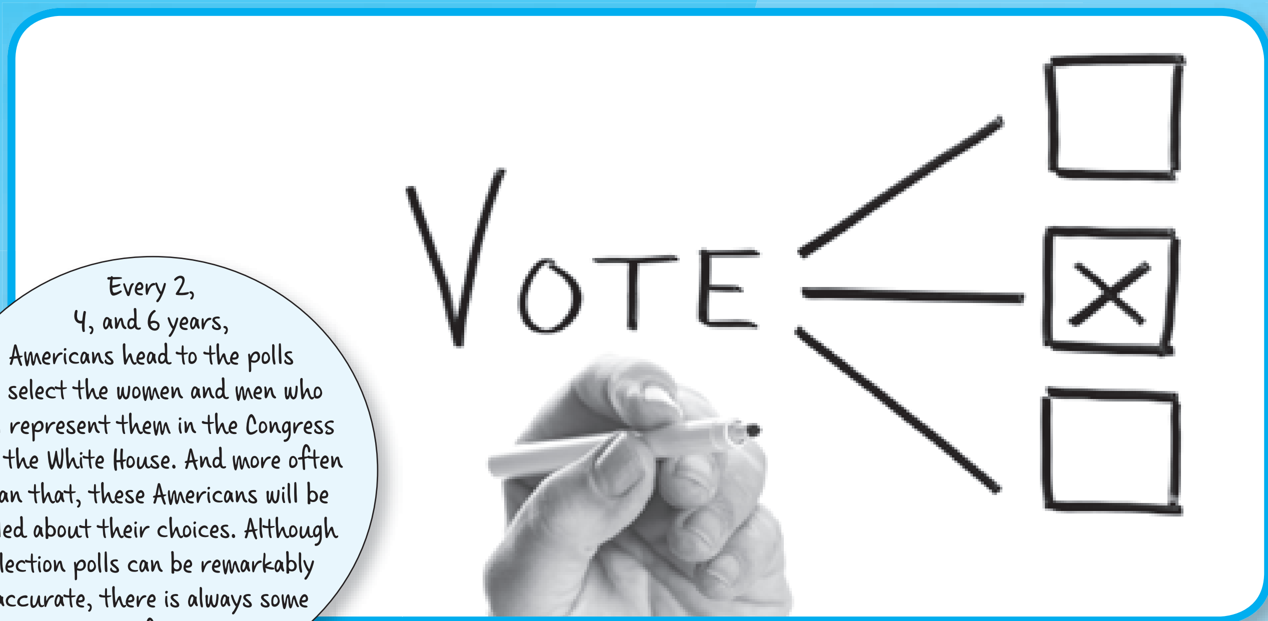
Making Inferences and Justifying Conclusions

2



Chapter 2 Overview

The first two lessons focus on methods of collecting data to analyze a question or characteristic of interest, specific sampling methods, and the significance of randomization. Then, students use data from samples to estimate population means and proportions, and determine whether results are statistically significant. In the last lesson, students have the opportunity to complete a culminating project based on concepts from the chapter.

2

Lesson		CCSS	Pacing	Highlights	Models	Worked Examples	Peer Analysis	Talk the Talk	Technology
2.1	Sample Surveys, Observational Studies, and Experiments	S.IC.1 S.IC.3	1	Students analyze the characteristics of sample surveys, observational studies, and experiments and their role in role describing populations. Random samples, biased samples, and confounding are included in this lesson. Students apply their knowledge at the end of the lesson by selecting and designing a sample survey, observational study, or experiment to gather data for a question of interest. The question of interest is revisited at the end of the first four lessons in order to prepare student for the culminating project at the end of the chapter.			x	x	
2.2	Sampling Methods and Randomization	S.IC.1 S.IC.3	2	Students analyze various sampling methods and the important role of randomization in sampling. Students apply their knowledge at the end of the lesson by selecting and describing a random sampling method for the recurring question of interest.		x	x		x
2.3	Using Confidence Intervals to Estimate Unknown Population Means	S.IC.1 S.IC.4 S.IC.6	2	Students analyze margin of error for population means and proportions. They build on their previous knowledge of normal distribution to determine margin of error for population means or population proportions using confidence intervals. At the end of the lesson, students use sample data for the recurring question of interest to determine a confidence interval for a population mean or population proportion.	x	x	x	x	

Lesson		CCSS	Pacing	Highlights	Models	Worked Examples	Peer Analysis	Talk the Talk	Technology
2.4	Using Statistical Significance to Make Inferences About Populations	S.IC.1 S.IC.2 S.IC.4 S.IC.5 S.IC.6	2	Students build on their previous knowledge of confidence intervals and margin of error to decide whether differences between population means or population proportions are statistically significant. At the end of the lesson, students use sample data for the recurring question of interest to decide whether differences between population means or population proportions are statistically significant.	x		x		
2.5	Designing a Study and Analyzing the Results	S.IC.1 S.IC.2 S.IC.3 S.IC.4 S.IC.5 S.IC.6	2	In this lesson, students have the opportunity to complete a culminating project. It involves designing, conducting, analyzing, and summarizing results of a sample survey, observational study, or experiment, for a question of interest of their choice.					

Skills Practice Correlation for Chapter 2

2

Lesson		Problem Set	Objectives
2.1	Sample Surveys, Observational Studies, and Experiments		Vocabulary
		1 – 6	Identify populations, samples, and characteristics of interest for situations
		7 – 12	Classify scenarios as sample surveys, observational studies, or experiments, and identify treatments of experiments
		13 – 18	Explain how confounding could occur in observational studies
2.2	Sampling Methods and Randomization		Vocabulary
		1 – 6	Select subjective samples that best represent the means of data sets
		7 – 12	Use a random number generator to select random samples
		13 – 18	Determine whether studies have biases
		19 – 24	Select stratified random samples from given data sets
		25 – 30	Select cluster samples from given data sets
		31 – 36	Estimate population means using data from samples
2.3	Using Confidence Intervals to Estimate Unknown Population Means		Vocabulary
		1 – 6	Determine whether scenarios represent 68%, 95%, or 99.7% confidence intervals
		7 – 12	Determine 95% confidence intervals to estimate population proportions
		13 – 18	Determine 95% confidence intervals to estimate population means
2.4	Using Statistical Significance to Make Inferences About Populations		Vocabulary
		1 – 6	Label horizontal axes of sampling distributions and determine data values that are statistically significant
		7 – 12	Use 95% confidence interval to determine whether differences between population proportion estimates are statistically significant
		13 – 18	Determine 95% confidence intervals to estimate population means
		19 – 24	Use given confidence intervals to make inferences about populations

Lesson		Problem Set	Objectives
2.5	Designing a Study and Analyzing the Results	1 – 6	Decide whether sample surveys, observational studies, or experiments are the best methods for given scenarios and describe how to obtain random samples
		7 – 12	Decide whether random sampling, stratified random sampling, or clustered sampling is the best method for given scenarios and describe how to obtain samples
		13 – 18	Identify possible biases in scenarios
		19 – 24	Create dot plots for data sets
		25 – 30	Create histograms for data sets
		31 – 36	Create stem-and-leaf plots for data sets
		37 – 42	Create box-and-whisker plots for data sets
		43 – 48	Determine means, medians, and modes of data sets and describe whether data sets are symmetric
		49 – 54	Determine standard deviations and quartiles of data sets

2

2.1

For Real?

Sample Surveys, Observational Studies, and Experiments

LEARNING GOALS

In this lesson, you will:

- Identify characteristics of sample surveys, observational studies, and experiments.
- Differentiate between sample surveys, observational studies, and experiments.
- Identify possible confounds in the design of experiments.

KEY TERMS

- characteristic of interest
- sample survey
- random sample
- biased sample
- observational study
- experiment
- treatment
- experimental unit
- confounding

ESSENTIAL IDEAS

- A characteristic of interest is a specific question that you are trying to answer or the specific information that you are trying to gather.
- A sample survey poses a question of interest to a sample of a targeted population.
- The population is the entire set of items from which data can be selected.
- A sample is a subset of data that is selected from the population.
- A random sample is a sample that is selected from the population in such a way that every member of the population has the same change of being selected.
- A biased sample is a sample that is not representative of the population.
- Confounding occurs when there are other possible reasons for the results to have occurred that were not identified prior to the study.
- An observational study gathers data about a characteristic of the population in its natural setting.
- An experiment gathers data on the effect of one or more treatments, or experimental conditions, on the characteristic of interest.
- Experimental units or members of a sample are randomly assigned to a treatment group.

COMMON CORE STATE STANDARDS FOR MATHEMATICS

S-IC Making Inferences and Justifying Conclusions

Understand and evaluate random processes underlying statistical experiments

1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

Make inferences and justify conclusions from sample surveys, experiments, and observational studies

3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.

Overview

Examples of a sample survey, an observational study, and an experiment are provided. Students answer questions related to each method and differentiate between the methods. Vocabulary terms associated with each type of data collection are defined. Students classify scenarios as a sample survey, an observational study, or an experiment and list the characteristic of interest, the population, the sample, factors that contribute to confounding, and ways to prevent bias. In the last activity, students design a data collection plan to learn how much time fellow students spend online each day.

Warm Up

Explain how each situation may lead to an invalid conclusion.

1. A student wants to determine if eating carrots is related to improved vision. She asks a local eye doctor to administer a voluntary survey to his patients. The survey contains questions about the amount of carrots they regularly eat and their vision.

   Answers will vary.

   The conclusions drawn from the survey results could be biased because many factors other than dietary factors affect vision. Good and poor vision could be related to genetics. Inadequate health care at an early age could attribute to poor eyesight.

2. A kindergarten teacher was determined to prove that there is a link between students with behavior issues and students that were raised outside of the home environment (i.e., daycare). For years, she recorded students who had behavior issues and referenced her list with the environment in which they were raised.

   Answers will vary.

   The conclusions drawn from the research could be biased because behavior issues can develop from a multitude of experiences not related to where a person was nurtured. Parent separation, death of a loved one, and a change in environment are a few examples that could be related to behavioral issues.

2

2

2.1

For Real?

Sample Surveys, Observational Studies, and Experiments

LEARNING GOALS

In this lesson, you will:

- Identify characteristics of sample surveys, observational studies, and experiments.
- Differentiate between sample surveys, observational studies, and experiments.
- Identify possible confounds in the design of experiments.

KEY TERMS

- characteristic of interest
- sample survey
- random sample
- biased sample
- observational study
- experiment
- treatment
- experimental unit
- confounding

Have you taken medicine to treat an illness? Imagine that the medicine you took was not really medicine, but just a sugar pill. In medical studies, people who have unknowingly taken a sugar pill—called a placebo—have reported that the pill has had an effect similar to medicine, even though there was no medicine in the pill at all. This is an example of what is called the placebo effect.

Researchers must always be on the lookout for placebo effects. They may be to blame for successful or unsuccessful outcomes to experiments.

2

Problem 1

A researcher wants to design a survey to determine the amount of time teenagers spend online each day. Students identify the characteristic of interest, discuss the population and write a survey question to collect data. They list strategies to consider when selecting the sample survey, and suggest what might introduce confounding when the survey is administered. The terms characteristic of interest, sample survey, random sample, biased sample, and confounding are defined in this problem.

Grouping

- Ask a student to read the definitions. Discuss as a class.
- Have students complete Questions 1 through 4 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 1 through 3

- Can there be more than one characteristic of interest in a sample survey?
- If Sandy includes all teenagers in the United States, what age group would be included in the sample survey?
- Is Augie's population and Sandy's population the same population? Why not?

PROBLEM 1 Survey Says

You can use data to help answer questions about the world. The specific question that you are trying to answer or the specific information that you are trying to gather is called a **characteristic of interest**.

For example, you can use data to help determine which drug is most effective, teenagers' favorite television program, or how often doctors wash their hands.

One way of collecting data is by using a *sample survey*. A **sample survey** poses one or more questions of interest to obtain sample data from a population. Recall, a population represents all the possible data that are of interest in a survey, and a sample is a subset of data that is selected from the population.

A researcher wants to design a sample survey to determine the amount of time that U.S. teenagers between the ages of 16 to 18 spend online each day.

1. Identify the characteristic of interest in the sample survey.
   The characteristic of interest is the amount of time spent online each day by 16- to 18-year olds.

2. Identify the population that the researcher is trying to measure by using a sample survey.
   The researcher is trying to measure the population of U.S. teenagers between 16 and 18 years old.

3. Augie and Sandy were discussing the population of the survey.

   Augie

   The population is all 16- to 18-year-olds in the United States.

   Sandy

   The population is all teenagers in the United States.

   Who is correct? Explain your reasoning.
   Augie is correct. The characteristic of interest is the amount of time spent online each day by teenagers between the ages of 16 and 18, so the population is only teenagers between those specific ages.

- Do you need to know which day of the week teenagers spend the most time online?
- Do you need to know if teenagers spend time online on Saturday or Sunday?

Guiding Questions for Share Phase, Question 4

- Is Augie's population and Sandy's population the same population? Why not?
- Do you need to know which day of the week teenagers spend the most time online?
- Do you need to know if teenagers spend time online on Saturday or Sunday?
- Do you need to know how much time teenagers spend online each week day?
- Do you think more teenagers spend time online on weekdays or weekends? Why?

Grouping

- Ask a student to read narrative and the definitions. Discuss as a class.
- Have students complete Questions 5 and 6 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Question 5

- What is the difference between a random sample and a biased sample?
- Under what circumstance, if any, would a biased sample be helpful?
- Should the representative sample include both males and females?
- Should the representative sample include minorities?
- Should the representative sample include students in college?
- Should the representative sample include students in high school?
- Should the representative sample include teenagers that do not have a computer at home?
- Should the representative sample include teenagers that do not have internet access at home?

2

4. Write a survey question or questions that the researcher could use to collect data from the participants in the survey.

Answers will vary.

On average, how many hours do you spend online a day?

On average, how many hours do you spend online each day of the week?

When sample data are collected in order to describe a characteristic of interest, it is important that such a sample be as representative of the population as possible. One way to collect a representative sample is by using a *random sample*. A **random sample** is a sample that is selected from the population in such a way that every member of the population has the same chance of being selected. A **biased sample** is a sample that is collected in a way that makes it unrepresentative of the population.

5. Joanie and Richie were discussing strategies the researcher could use to select a representative sample of 16- to 18-year-olds.

Joanie

The sample should include 16- to 18-year-olds from various states.

Richie

Include 16- to 18-year-olds who live in various types of communities, like urban, suburban, and rural.

List some additional strategies the researcher should consider when selecting the sample.

Answers will vary.

The sample should include:

- Both males and females.
- A variety of ethnicities.
- 16-, 17-, and 18-year-olds.

Guiding Questions for Share Phase, Question 6

- Which teenagers are being excluded if Cherese distributes the survey to students after school on Friday as they are leaving the building?
- Will the schools policies regarding the use of the internet influence the survey results? How?
- Why is Friday perhaps not the best day to distribute the survey?
- Should the survey be distributed at more than one location?

Problem 2

An observational study states 70% of in-house day care centers in a specific state show their children as much as 2.5 hours of television per day. Students identify the characteristic of interest, the population and the sample of the observational study. Students distinguish between an observational study and a sample survey.

Grouping

- Ask a student to read the definition and information. Discuss as a class.
- Have students complete Questions 1 and 2 with a partner. Then have students share their responses as a class.

2

6. Cherese suggested that the researcher could post the survey online and then distribute the link to the survey to students after school on Friday as they are leaving the building.

Will this method result in a biased sample? Explain your reasoning.

Yes. Putting the survey online may mean that students who answer the survey will be the ones who spend a lot of time online.

PROBLEM 2 Confound It All!

In an **observational study,** data are gathered about a characteristic of the population by simply observing and describing events in their natural settings. Recording the number of children who use the swings at a local park would be an example of a simple observational study.

The results of an observational study state that approximately 70% of in-house day care centers in one U.S. state show as much as 2.5 hours of television to the children per day. The observational study examined 132 day care centers in one state.

1. Identify the population, the sample, and the characteristic of interest in the observational study.

The population is all of the in-house day care centers in a state.

The sample is 132 day care centers in the state.

The characteristic of interest is the daily amount of TV shown to children in day care centers.

2. List some similarities and differences between an observational study and a sample survey.

Answers will vary.

Conducting sample surveys or observational studies share the same purpose. The goal of both methods is to collect meaningful data.

One difference between the two methods involves how the data are collected. For a sample survey, data are typically collected from participants using a questionnaire or survey. For an observational study, data can be collected just by observing participants in their normal, everyday environment.

Guiding Questions for Share Phase, Questions 1 and 2

- Is the population all of the in-house day care centers in a state, or 70% of the in-house day care centers in a state?
- Is 132 in-house day care centers large enough for an observational study? Why or why not?
- Is the characteristic of interest the daily amount of TV shown to children or the maximum amount of TV shown daily to children in day care centers?

Grouping

- Ask a student to read the narrative and definitions. Discuss as a class.
- Have students complete Questions 3 through 6 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 3 through 6

- Is the population all asthma patients or 200 asthma patients? What about the sample?
- Are 200 patients enough to conduct an experiment and reach a valid conclusion?
- Is the characteristic of interest the effectiveness of the new asthma drug or the effectiveness of the placebo?
- How are the two treatments in this experiment different?
- Do you know how the patients were selected for this experiment? Does that make a difference?
- Why does participating in a clinical trial involve some risk?
- What population do you think would be most willing to participate in this experiment?
- Would a monetary incentive attract participants? Why?
- Could a random number generator be used to divide the participants into 2 treatment groups? How?

An **experiment** gathers data on the effect of one or more **treatments**, or experimental conditions, on the characteristic of interest. Members of a sample, also known as **experimental units**, are randomly assigned to a treatment group.

Researchers conducted an experiment to test the effectiveness of a new asthma drug. They collected data from a sample of 200 asthma patients. One hundred of the patients received a placebo treatment along with an inhaler. The other one hundred patients received the new drug along with an inhaler. Monthly blood and breathing tests were performed on all 200 patients to determine if the new drug was effective.

3. Identify the population, the sample, and the characteristic of interest in the experiment.

The population is all asthma patients.

The sample is 200 asthma patients. One hundred receive the new drug, and the other one hundred receive the placebo.

The characteristic of interest is the effectiveness of the new asthma drug.

4. What are the treatments in the experiment?

There are 2 treatments.

The treatment for one group is an inhaler plus a placebo, and the treatment for the other group is an inhaler plus the new drug.

5. What are some ways the researchers could choose a biased sample for this experiment?

Answers will vary.

The researchers could choose only asthma patients within a very limited age range. They might also choose a sample of asthma patients who only have very mild symptoms or very severe symptoms.

Confounding occurs when there are other possible reasons, called confounds, for the results to have occurred that were not identified prior to the study.

6. Suppose one of the treatment groups was given the new drug with an inhaler and the other group was given a placebo with no inhaler. Describe how this design of the experiment introduces a confound.

If the treatment group that is given the new drug and the inhaler shows improvement, it would be impossible to tell whether that improvement is because of the new drug or because this group was also given an inhaler while the other group wasn't.

- Would a blood or breathing test help to provide a conclusive result?
- Are the first and second steps in designing an observational study or a sample survey the same as the first and second steps in designing an experiment?
- Does conducting an experiment share the same purpose as conducting an observational study or a sample survey?
- What is the common goal of administering or conducting a sample survey, an observational study and an experiment?
- What element of the design might introduce confounding?

2

Talk the Talk

Students classify different scenarios as a sample survey, an observational survey, or an experiment. They identify the population, the sample, and the characteristic of interest.

2

Grouping

Have students complete Questions 1 through 4 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 1 and 2

- What is the difference between a sample survey, an observational study, and an experiment?
- What is the difference between a sample and a population?
- What confounds could exist in each scenario?

Talk the Talk

Classify each scenario as a sample survey, an observational study, or an experiment, and explain your reasoning. Then, identify the population, the sample, and the characteristic of interest.

1. To determine whether there is a link between high-voltage power lines and illnesses in children who live in the county, researchers examined the illness rate for 100 children that live within $\frac{1}{4}$ of a mile from power lines and the illness rate for 100 children that live more than $\frac{1}{4}$ of a mile from power lines.

   This is an observational study because the researchers gathered data by observing and describing events in their natural setting.

   The population is all children in the county.

   The sample is 200 children. One hundred of the children live within $\frac{1}{4}$ of a mile from power lines, and the other 100 children live more than $\frac{1}{4}$ of a mile from power lines.

   The characteristic of interest is the illness rate of children living near power lines.

2. Seventy of the school's calculus students are randomly divided into two classes. One class uses a graphing calculator all the time, and the other class never uses graphing calculators. The math department team leader wants to determine whether there is a link between graphing calculator use and students' calculus grades.

   This is an experiment because the students are randomly assigned to classes with different treatments. The two treatments are using a graphing calculator and not using a graphing calculator.

   The population is all calculus students in the school.

   The sample is 70 students. Half of the students are randomly assigned to the class that uses a graphing calculator, and the other half are assigned to a class that does not use graphing calculators.

   The characteristic of interest is the link between using graphing calculators and calculus students' grades.

Guiding Questions for Share Phase, Questions 3 and 4

- Does the data from 15 pediatricians provide enough information to draw a valid conclusion? Explain.
- What other method could be used to gather data for the relationship between the amount of TV children watch and obesity?
- Is distributing a questionnaire during lunch a good method for collecting data? Explain.
- What other method could be used to collect data about children's daily amount of physical activity and their overall energy level? Explain.

2

3. A medical researcher wants to learn whether or not there is a link between the amount of TV children watch each day and childhood obesity in a particular school district. She gathers data from the records of 15 local pediatricians.

This is an observational study because the researcher gathered the data from observation rather than experiment.

The population is all children in a particular school district.

The sample is the records from the 15 different pediatricians.

The characteristic of interest is the link between watching TV and obesity in children.

4. In a particular school district, a researcher wants to learn whether or not there is a link between a child's daily amount of physical activity and their overall energy level. During lunch at a school, she distributed a short questionnaire to students in the cafeteria.

This is a sample survey because the students were asked to respond to questions.

The population is all children in as particular school district.

The sample is the students in the cafeteria who responded to the questionnaire.

The characteristic of interest is the link between a child's daily amount of physical activity and their overall energy level.

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Question 1 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Online Time Study

- Is it possible to have more than one population?
- What other methods are efficient for collecting data?
- Will your data be completely unbiased?

Online Time Study, Part I

To design a sample survey, observational study, or experiment, consider these steps:

- Identify the characteristic of interest.
- Identify the population.
- Identify methods to collect the sample so that the sample is not biased.
- Ensure that participants are randomly assigned to a treatment.
- Eliminate elements of the design that may introduce confounding.

1. Design a data collection plan to learn how much time students in your school spend online each day.

a. Identify the population and the characteristic of interest.

Answers will vary.

The population is all students in my school.

The characteristic of interest is how much time students spend online each day.

b. Is the most efficient method for collecting the data a sample survey, an observational study, or an experiment? Explain your reasoning.

Answers will vary.

I chose a sample survey because I can collect data from the students with a 1-question survey.

c. Explain how you can gather data from a representative, unbiased sample of students in your school.

Answers will vary.

I'll write a clear and brief survey question: "How much time, in hours, do you spend online each day?"

Then, I'll ask every teacher to randomly select 3 boys and 3 girls from their class and give them the survey questions. I'll collect the survey responses from the teachers at the end of the day.

Be prepared to share your solutions and methods.

Check for Students' Understanding

1. What are the essential questions that should be asked when designing a sample survey?
   - What is the characteristic of interest?
   - What is the population?
   - How will the sample be randomly selected so that it is representative of the population and not biased?

2

2. What are the essential questions that should be asked when designing an observational study?
   - What is the characteristic of interest?
   - What is the population?
   - How will the sample be identified so that it is representative of the population and not biased?

3. What are the essential questions that should be asked when designing an experiment?
   - What is the characteristic of interest?
   - What is the population?
   - What are the treatments in this experiment?
   - How will the experimental units be identified and how will they be randomly assigned to the treatment?
   - How will the differences in treatments be analyzed and interpreted in order to draw a valid conclusion?

2

2.2

Circle Up

Sampling Methods and Randomization

LEARNING GOALS

In this lesson, you will:

- Use a variety of sampling methods to collect data.
- Identify factors of sampling methods that could contribute to gathering biased data.
- Explore, identify, and interpret the role of randomization in sampling.
- Use data from samples to estimate population mean.

KEY TERMS

- convenience sample
- subjective sample
- volunteer sample
- simple random sample
- stratified random sample
- cluster sample
- cluster
- systematic sample
- parameter
- statistic

ESSENTIAL IDEAS

- A convenience sample is a sample whose data is based on what is convenient for the person choosing the sample.
- A subjective sample is a sample in which an individual makes a judgment about which data items to select.
- A volunteer sample is a sample whose data consists of those who volunteer to be part of a sample.
- A simple random sample is a sample in which all members of a population have an equal chance of being selected.
- A stratified random sample is a random sample obtained by dividing a population into different groups, or strata, according to a characteristic and randomly selecting data from each group.
- A cluster sample is a random sample that is obtained by creating clusters. Each cluster contains the characteristics of the population. Then, one cluster is randomly selected for the sample.
- A systemic sample is a random sample obtained by selecting every *nth* data value in a population.

COMMON CORE STATE STANDARDS FOR MATHEMATICS

S-IC Making Inferences and Justifying Conclusions

Understand and evaluate random processes underlying statistical experiments

1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

Make inferences and justify conclusions from sample surveys, experiments, and observational studies

3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.

Overview

Three sampling methods are introduced; convenience sampling, subjective sampling, and volunteer sampling. Students perform a simple random sampling by using a random digit table and a graphing calculator random number generator. Students use other types of random sampling methods such as stratified random samples, cluster samples, and systematic samples.

2

Warm Up

Explain how each situation may introduce confounding.

1. A local bookstore offered all of their customers a $10 gift certificate to complete a brief online survey in order to determine the most popular reading genre.
   Answers will vary.
   This is not a valid method to determine the most popular reading genre because only customers that wanted to do the survey and had internet access at home were able to complete the survey. They do not reasonably represent all the people that read books.

2

2. A concession stand at the community pool wanted to determine the most popular beverage. They recorded the number of water bottles and the number of soft drinks that were sold.
   Answers will vary.
   This is not a valid method to determine the most popular beverage because the only people they considered are the people that bought those two kinds of drinks. The data does not represent all people. Some people might bring their own drinks and some people may prefer drinks that are not water or soft drinks, such as juice, tea, milk, or coffee.

2

2.2

Circle Up

Sampling Methods and Randomization

2

LEARNING GOALS

In this lesson, you will:

- Use a variety of sampling methods to collect data.
- Identify factors of sampling methods that could contribute to gathering biased data.
- Explore, identify, and interpret the role of randomization in sampling.
- Use data from samples to estimate population mean.

KEY TERMS

- convenience sample
- subjective sample
- volunteer sample
- simple random sample
- stratified random sample
- cluster sample
- cluster
- systematic sample
- parameter
- statistic

What English word is missing below?

_____ _____ _____ _____ _____ _____

When you play word games like this, where you guess the letters until you figure out the word, you think about samples and populations.

For example, you know that the missing word is a sample of the population of words in the English language. Since "e" is a frequently used letter and "z" is used infrequently in words, you would probably guess "e" before you guessed "z".

It is useful in statistics, too, to assume that the characteristics of a sample match those of a population—as long as that sample is chosen wisely!

Problem 1

One hundred circles of various diameters are drawn. A table listing an identification number, the diameter, and the area for each circle is provided. These materials are used to demonstrate convenience sampling, subjective sampling, and volunteer sampling.

2

Grouping

Ask a student to read the information and definitions. Complete Question 1 and discuss as a class.

Guiding Questions for Discuss Phase, Question 1

- Is the sample size large enough to represent the entire population?
- How big should the sample size be to reasonably represent the population?
- Do you think the mean size of the convenience sample is close to the mean size of the entire circle population? Why or why not?
- Did Mauricia look at the first five circles before selecting the sample?
- Why should Maurica have looked at the first five circles before selecting the sample?

PROBLEM 1 You Gotta Mix It Up

When you use statistics, you are often measuring the values of a population by focusing on the measurements of a sample of that population. A population does not have to refer to people. It can be any complete group of data—like the areas of 100 circles.

The end of this lesson includes 100 circles and a table. The table lists an identification number, the diameter, and the area for each circle. Suppose you want to determine the mean area of all 100 circles. Calculating the areas of all of the circles would be time-consuming. Instead, you can use different samples of this population of circles to estimate the mean area of the entire population.

1. Without looking at the circles, Mauricia decided to use Circles 1–5 for her sample. Is it likely that those 5 circle areas are representative of all 100 circles? Explain your reasoning.

   Answers will vary.

   Circles 1–5 are probably not representative of the 100 circles for several reasons. One reason is that the sample size is small. Another reason is that Mauricia did not take a look at these circles before selecting the sample. The actual sizes of Circles 1–5 may not be representative of all 100 circles.

2. Analyze the circles. Select a sample of 5 circles that you think best represents the entire set of circles.

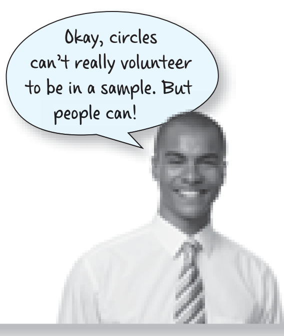
   Answers will vary.

   I chose Circles 22, 23, 25, 54, and 60.

The sample of circles Mauricia chose is called a *convenience sample*. A **convenience sample** is a sample whose data is based on what is convenient for the person choosing the sample.

The sample of circles you chose in Question 2 is called a *subjective sample*. A **subjective sample** is a sample drawn by making a judgment about which data items to select.

Another type of sample is a *volunteer sample*. A **volunteer sample** is a sample whose data consists of those who volunteer to be part of a sample.



Grouping

- Have students complete Questions 2 through 4 with a partner. Then have students share their responses as a class.
- Ask a student to read the definitions. Discuss as a class.

Guiding Questions for Share Phase, Question 2

- Was any criteria used in the selection of the subjective sample of 5 circles?
- What method was used to select a subjective sample of 5 circles?
- Why do you think your selection of 5 circles is representative of the population?
- Do you think the 5 circles you selected are close to the mean size of the circles in the population?

Grouping

Have students complete Question 4 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 3 and 4

- What is the difference between a convenience sample and a subjective sample?
- In what circumstance would a convenience sample be helpful?
- Is a convenience sample helpful in this situation?
- Do any of the three types of sampling introduce personal bias?
- How does a convenience sample introduce personal bias?

3. Olivia and Ricky discussed whether a convenience sample or a subjective sample is more likely to be representative of the population of circle areas.

> Olivia
>
> I think a subjective sample is more likely to be representative of the 100 circles than the convenience sample.

> Ricky
>
> The subjective sample and the convenience sample are equally likely to be representative of the 100 circles.

Who is correct? Explain your reasoning.

Olivia is correct.

The convenience sample may happen to have a mean that is closer to the mean of the population. However, a convenience sample may include very large or very small circles, in which case the mean will not be close to the mean of the population.

When choosing a subjective sample, a person can be careful to select circles that appear to be representative of all of the circles.

4. Olivia shared her conclusion about convenience samples, subjective samples, and volunteer samples.

> **Olivia**
>
> Even though one method may be better than another in a specific situation, collecting data using a convenience sample, subjective sample, or volunteer sample will likely result in a biased sample.

It's the sampling method that leads to the bias. It's not that an individual sample is biased or not.

Explain why Olivia's statement is correct.

Olivia's statement is correct because all three sampling methods involve personal bias, which reduces the likelihood of a representative, unbiased sample. Convenience samples are based on convenience, subjective samples are based on a person's opinion or perspective, and volunteer samples are based on person's willingness to be part of a sample.

It is possible, but less likely, for any of the three sampling methods to produce a representative, unbiased sample.

- How does a subjective sample introduce personal bias?
- How does a volunteer sample introduce personal bias?
- Which sampling method is based on a person's opinion or perspective?
- Which sampling method is based on a person's willingness to be a part of a sample?

Problem 2

The use of a random digit table enables students to perform a simple random sample of five circles. Instructions for using the graphing calculator to generate a random list of numbers are provided. Other types of random sampling methods such as stratified random samples, cluster samples, and systematic samples are defined.

Grouping

Ask a student to read the definition. Complete Question 1 and discuss as a class.

Guiding Questions for Discuss Phase, Question 1

- Which row in the table did you select?
- Did your classmates select different rows?
- What is the mean of your sample?
- What are the means of your classmates' samples?
- Do any of the means seem to be reflective of the entire population?

2

PROBLEM 2 Equal Opportunity for All

A **simple random sample** is a sample composed of data elements that were equally likely to have been chosen from the population.

1. Explain how convenience samples, subjective samples, and volunteer samples do not include data elements that were equally likely to have been chosen from the population.

   The method of choosing data for each of these types of samples is not random. In convenience sampling, data that are inconvenient to collect are ignored. In subjective sampling, a person may be biased against selecting certain types of data. And in volunteer sampling, there is no real control over what data are in the sample.

Using a random digit table is one option for selecting a simple random sample. To use the table, begin at any digit and follow the numbers in a systematic way, such as moving across a row until it ends and then moving to the beginning of the next row.

Random Digit Table

Line 1	65285	97198	12138	53010	94601	15838	16805	61004	43516	17020
Line 2	17264	57327	38224	29301	31381	38109	34976	65692	98566	29550
Line 3	95639	99754	31199	92558	68368	04985	51092	37780	40261	14479
Line 4	61555	76404	86210	11808	12841	45147	97438	60022	12645	62000
Line 5	78137	98768	04689	87130	79225	08153	84967	64539	79493	74917
Line 6	62490	99215	84987	28759	19177	14733	24550	28067	68894	38490
Line 7	24216	63444	21283	07044	92729	37284	13211	37485	10415	36457
Line 8	16975	95428	33226	55903	31605	43817	22250	03918	46999	98501
Line 9	59138	39542	71168	57609	91510	77904	74244	50940	31553	62562
Line 10	29478	59652	50414	31966	87912	87154	12944	49862	96566	48825
Line 11	96155	95009	27429	72918	08457	78134	48407	26061	58754	05326
Line 12	29621	66583	62966	12468	20245	14015	04014	35713	03980	03024
Line 13	12639	75291	71020	17265	41598	64074	64629	63293	53307	48766
Line 14	14544	37134	54714	02401	63228	26831	19386	15457	17999	18306
Line 15	83403	88827	09834	11333	68431	31706	26652	04711	34593	22561
Line 16	67642	05204	30697	44806	96989	68403	85621	45556	35434	09532
Line 17	64041	99011	14610	40273	09482	62864	01573	82274	81446	32477
Line 18	17048	94523	97444	59904	16936	39384	97551	09620	63932	03091
Line 19	93039	89416	52795	10631	09728	68202	20963	02477	55494	39563
Line 20	82244	34392	96607	17220	51984	10753	76272	50985	97593	34320

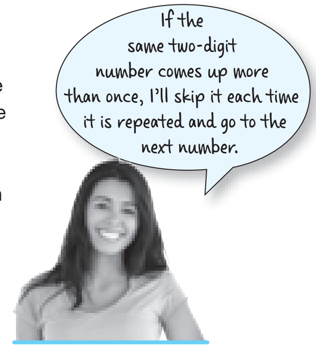
Grouping

Have students complete Questions 2 through 5 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 2 through 5

- Is it easier to use the random digit table or a graphing calculator to perform a random number sampling? Why?
- Which method appears to be more of a random selection, using the random digit table or using a graphing calculator?
- How many of your classmates' random samples contained the same circle?
- How does the mean of your random sample compare to the means of your classmates' random sample?

You can use two digits at a time to choose a sample of 5 circles.



2. Select a simple random sample of 5 circles using the random digit table. Pick any row of the table. Use the first two digits to represent the first circle of the sample, the next two digits to represent the second circle of the sample, and so on. List the identification numbers of the 5 circles.

   Answers will vary.

   Using Line 7 of the random digit table results in a simple random sample of Circles 24, 21, 66, 34, and 44.

You can also use a graphing calculator to generate a random list of numbers.

```
You can use a graphing calculator to generate
a random list of numbers and select a simple
random sample of 5 circles.

Step 1: Press MATH.
        Scroll to the PRB menu.
        Select 5:randInt(

Step 2: Enter a lower bound for the random number,
        an upper bound for the random number,
        and how many random numbers to generate.
        Use commas between values as you enter them.

Step 3: Press ENTER.
```

The lower bound is 0, the upper bound is 99, and the number of random numbers to generate is 5.

3. Use a graphing calculator to generate a random sample of 5 circles.

   Answers will vary.

   Using the random number generator on a graphing calculator, I selected Circles 41, 6, 73, 74, and 80.

4. Calculate the mean area of the circles in your simple random sample.

   Answers will vary.

5. Compare your simple random sample with your classmates' samples. What do you notice?

   Answers will vary.

   Many of my classmates had different samples. Some of the samples contained the same circle. The means of the circle areas are approximately the same.

Grouping

- Ask a student to read the narrative and Worked Example. Discuss as a class.
- Have students complete Questions 6 through 8 with a partner. Then have students share their responses as a class.

2

Guiding Questions for Share Phase, Questions 6 and 7

- How many members are in the small circle group?
- How many members are in the medium circle group?
- How many members are in the large circle group?
- Why is it necessary to maintain this 3-3-1 ratio of size related selections?
- If the lengths of the diameters in each of the three groups were differently defined, could the groups have a more even membership?
- Would dividing the circles into 4 size groups instead of 3 size groups and then choosing from each of those groups result in a more accurate representation of the population? Why or why not?

There are several other types of random samples, including *stratified random samples, cluster samples*, and *systematic samples*.

A **stratified random sample** is a random sample obtained by dividing a population into different groups, or strata, according to a characteristic and randomly selecting data from each group.

You can collect a stratified random sample of circles by first dividing the circles into groups.

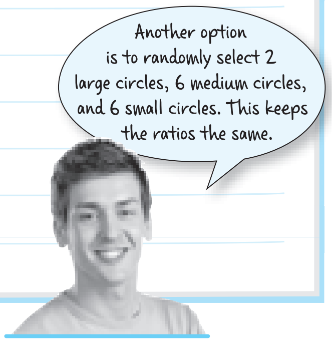
Define groups of circles based on the lengths of their diameters.

- Small circles: diameter $\leq \frac{1}{4}$ in.
- Medium circles: $\frac{1}{4}$ in. $<$ diameter $\leq 1\frac{1}{2}$ in.
- Large circles: diameter $> 1\frac{1}{2}$ in.

Small Circles (46)	Medium Circles (39)	Large Circles (15)
1, 4, 6, 13, 14, 16, 17, 19, 22, 24, 26, 28, 30, 33, 34, 37, 39, 42, 45, 46, 47, 51, 53, 56, 57, 58, 59, 62, 63, 67, 68, 72, 74, 78, 79, 82, 85, 87, 88, 89, 93, 94, 95, 97, 98, 99	0, 2, 3, 8, 9, 10, 11, 12, 21, 23, 25, 29, 31, 35, 36, 40, 41, 43, 49, 50, 52, 61, 64, 65, 66, 69, 71, 73, 75, 76, 77, 80, 81, 83, 84, 86, 90, 91, 96	5, 7, 15, 18, 20, 27, 32, 38, 44, 48, 54, 55, 60, 70, 92

There are about an equal number of small and medium circles and about a third as many small circles. To maintain this ratio in your stratified random sample, you can choose 3 small circles, 3 medium circles, and 1 small circle.

Select random circles from each group using a random digit table or a graphing calculator.



6. Collect a stratified random sample of circles. List the sample and explain your method.

Answers will vary.

Using pairs of digits in Line 13 from the random digit table, I selected Circles 1, 63, and 97 from the small circles, Circles 12, 52, and 91 from the medium circles, and Circle 32 from the large circles.

I used a graphing calculator to generate random integers between 1 and 100 until 3 small circles, 3 medium circles, and 1 large circle were selected.

7. Calculate the mean of the circle areas in your stratified random sample.

Answers will vary.

Guiding Questions for Share Phase, Question 8

- What is the difference between a stratified random sample and a cluster sample?
- What do a stratified random sample and a cluster sample have in common?
- Do there appear to be the same number of circles in each cluster?
- Could you divide the population into different clusters? How?
- If the population were divided into more than 12 clusters of equal area, would the sample result in a more accurate representation of the population? Why or why not?
- How many circles are in the cluster you selected?
- How did you determine which circles should be included in the cluster chosen?

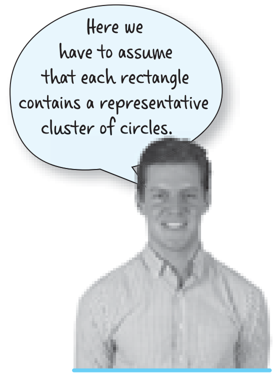
Grouping

- Ask a student to read the definition. Discuss as a class.
- Have students complete Questions 9 and 10 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 9 and 10

- What number did you randomly choose first?
- How did you choose the first number?
- What number did your classmates choose first?

A **cluster sample** is a random sample that is obtained by creating *clusters*. Then, one cluster is randomly selected for the sample. Each **cluster** contains the characteristics of a population.



8. Use the page that contains the circles at the end of this lesson to answer each question.

a. Draw 4 horizontal lines and 2 vertical lines so that the page is divided into 12 congruent rectangles. Each rectangle represents a cluster of circles. Number each cluster from 1 to 12.

See circle page.

b. Use a graphing calculator or the random digit table to randomly select one of the clusters. List the cluster sample.

I chose a cluster containing Circles 32, 43, 44, 45, 57, and 58.

c. Calculate the mean of the circle areas included in your cluster sample.

Answers will vary.

A **systematic sample** is a random sample obtained by selecting every *n*th data value in a population.

9. Select a systematic sample by choosing every 20th circle. First, randomly choose a number from 0 to 20 to start at and then choose every 20th circle after that.

One systematic sample is Circles 0, 20, 40, 60, and 80.

10. Calculate the mean of the circle areas included in your systematic sample.

Answers will vary.

- How does the average area of your selected circles compare to your classmates' average area of selected circles?
- Whose average area best represents the populations average area? How do you know?

Grouping

Have students complete Questions 11 and 12 with a partner. Then have students share their responses as a class.

2

Guiding Questions for Share Phase, Question 11

- Does using random sampling provide each member of the population an equal chance of being selected into the sample?
- Does using stratified random sampling provide each member of the population an equal chance of being selected into the sample?
- Does using cluster sampling provide each member of the population an equal chance of being selected into the sample?

11. Faheem and Calvin shared their thoughts about random sampling.

Faheem

Simple random sampling, stratified random sampling, and cluster sampling will always produce a representative, unbiased sample.

Calvin

Simple random sampling, stratified random sampling, or cluster sampling does not guarantee a representative, unbiased sample.

Who is correct? Explain your reasoning.

Calvin is correct.

Simple random sampling, stratified random sampling, and cluster sampling provide each member of the population an equal chance of being selected into the sample. These methods do not guarantee representative, unbiased samples.

Guiding Questions for Share Phase, Question 12

- If each member of the population has an equal chance of being selected into the sample, will this always result in an unbiased sample? Why not?
- Is it more likely that the best estimate of the population comes from the largest sample?
- Is it always the case that best estimate of the population comes from the largest sample?
- Should the sample mean approach the estimate of the population mean as the sample size increases or decreases? Why?

The mean of a sample, $\bar{x}$, can be used to estimate the population mean, μ. The population mean is an example of a **parameter**, because it is a value that refers to a population. The sample mean is an example of a **statistic**, because it is a value that refers to a sample.

The population mean for the 100 circles is $\mu = 0.58\pi$ square inches, or approximately 1.82 square inches.

12. Carla collected three simple random samples from the population of 100 circles and calculated the mean of each sample.

> *Carla*
>
> *I didn't expect the sample of 5 circles to have a mean closest to the mean of the population. I must have done something wrong when collecting the samples.*
>
> *Mean of 5 circles $\approx 0.55\pi$ square inches*
>
> *Mean of 15 circles $\approx 0.49\pi$ square inches*
>
> *Mean of 30 circles $\approx 0.65\pi$ square inches*

Is Carla's statement correct? Explain your reasoning.

Carla's statement is incorrect.

It is more likely for the best estimate of the population to come from the largest sample, but it is possible that the closest estimate comes from the smallest sample. In general, as sample size increases, the sample mean will be a better estimate of the population mean.

2

2

Online Time Study, Part II

Students designed a data collection plan to learn how much time students in their school spend online each day in the previous lesson. Now that they have been given several sampling methods they are asked to select the most appropriate sampling method to fit the plan.

Grouping

Have students complete the problem with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Online Time Study, Part II

- Which sampling method is best to use in this situation? Why?
- Is a different sampling method equally effective?
- Which other sampling methods could be used?
- Should students be randomly selected in each grade level or class?
- How many students should be randomly selected in each grade level or class?
- Should an equal number of males and females be selected?
- Should the classes chosen represent all subjects taught?
- What is the characteristic of interest in this situation?

Online Time Study, Part II

In the first lesson of this chapter, you designed a plan to learn about the amount of time students in your school are online each day.

1. Which sampling method would be best to select the data? Explain your reasoning.

Answers will vary.

My plan is to ask teachers from a variety of classes to randomly select 3 boys and 3 girls in their class and give them the survey questions. I'll collect the survey responses from the teachers at the end of the day.

It is important to choose a variety of classes that represents all subjects and all levels, which breaks down to 15 different classes.

	Regular	Honors	Advanced Placement
Math	✓	✓	✓
English	✓	✓	✓
Science	✓	✓	✓
Social Science	✓	✓	✓
Electives	✓	✓	✓

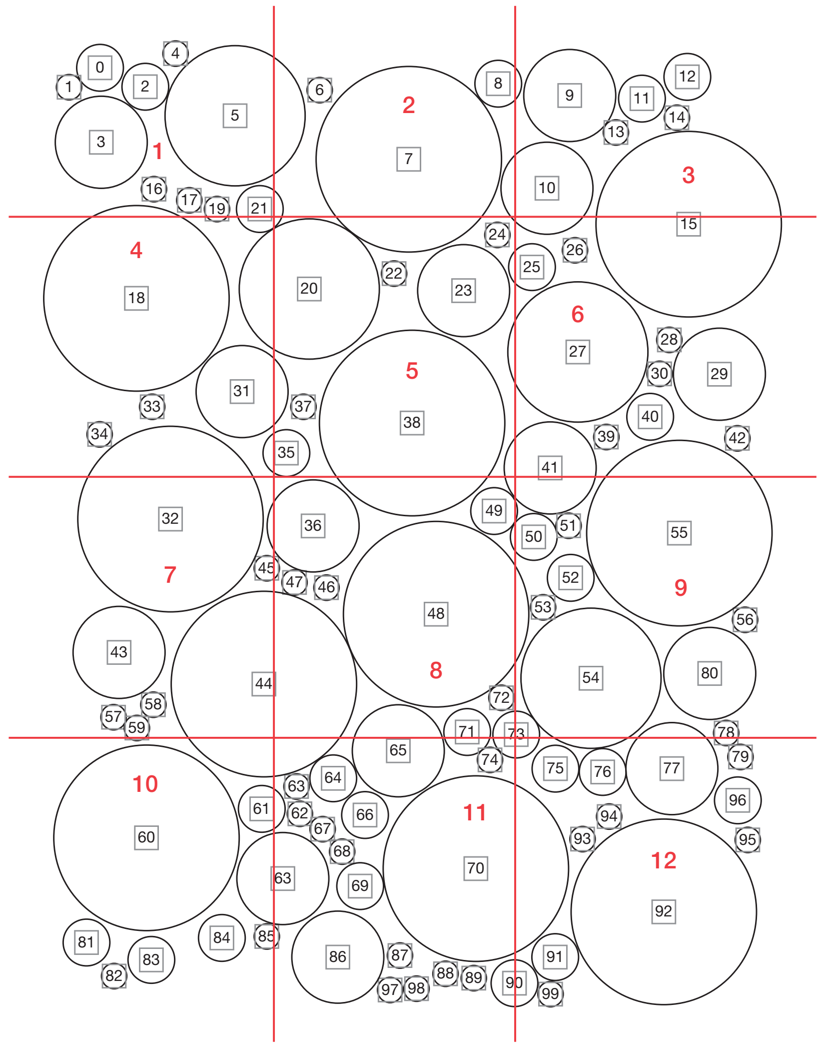
I plan to deliver the surveys to the teachers at the beginning of the day, ask them to randomly select 3 boys and 3 girls in their class, and give them the survey questions.

At the end of the day, I'll collect the surveys from the teachers.

Survey Question: "How much time, in hours, do you spend online each day?"

Be prepared to share your results and methods.

- How can you maximize on the randomness of this design plan?


4
0
1 2 5 6 2 8 9 11 12
3 1 7 13 14
16 17 19 21 10 3 15
24 26
4 22 25
18 20 23
6
27
28
5
30 29
31
33 37 38 40
34 39 42
35 41
49
32 36 55
50 51
45 47 46 52
7 9
53
48 56
43 8
80
44 54
72
58
57 59 71 73 78
65 74
10 64 75 76 77 79
63
61 62 66 11 96
67 94
60 93 95
12
68
63 69 70
92
85
84 81 87 91
86
83
88 89
82
97 98 90 99

2

2

Circle Number	Diameter (in.)	Area (in.^{2})	Circle Number	Diameter (in.)	Area (in.^{2})	Circle Number	Diameter (in.)	Area (in.^{2})
0	$\frac{1}{2}$	$\frac{1}{16}\pi$	**18**	2	π	**36**	1	$\frac{1}{4}\pi$
1	$\frac{1}{4}$	$\frac{1}{64}\pi$	**19**	$\frac{1}{4}$	$\frac{1}{64}\pi$	**37**	$\frac{1}{4}$	$\frac{1}{64}\pi$
2	$\frac{1}{2}$	$\frac{1}{16}\pi$	**20**	$1\frac{1}{2}$	$\frac{9}{16}\pi$	**38**	2	π
3	1	$\frac{1}{4}\pi$	**21**	$\frac{1}{2}$	$\frac{1}{16}\pi$	**39**	$\frac{1}{4}$	$\frac{1}{64}\pi$
4	$\frac{1}{4}$	$\frac{1}{64}\pi$	**22**	$\frac{1}{4}$	$\frac{1}{64}\pi$	**40**	$\frac{1}{2}$	$\frac{1}{16}\pi$
5	$1\frac{1}{2}$	$\frac{9}{16}\pi$	**23**	1	$\frac{1}{4}\pi$	**41**	1	$\frac{1}{4}\pi$
6	$\frac{1}{4}$	$\frac{1}{64}\pi$	**24**	$\frac{1}{4}$	$\frac{1}{64}\pi$	**42**	$\frac{1}{4}$	$\frac{1}{64}\pi$
7	2	π	**25**	$\frac{1}{2}$	$\frac{1}{4}\pi$	**43**	1	$\frac{1}{4}\pi$
8	$\frac{1}{2}$	$\frac{1}{16}\pi$	**26**	$\frac{1}{4}$	$\frac{1}{64}\pi$	**44**	2	π
9	1	$\frac{1}{4}\pi$	**27**	$1\frac{1}{2}$	$\frac{9}{16}\pi$	**45**	$\frac{1}{4}$	$\frac{1}{64}\pi$
10	1	$\frac{1}{4}\pi$	**28**	$\frac{1}{4}$	$\frac{1}{64}\pi$	**46**	$\frac{1}{4}$	$\frac{1}{64}\pi$
11	$\frac{1}{2}$	$\frac{1}{16}\pi$	**29**	1	$\frac{1}{4}\pi$	**47**	$\frac{1}{4}$	$\frac{1}{64}\pi$
12	$\frac{1}{2}$	$\frac{1}{16}\pi$	**30**	$\frac{1}{4}$	$\frac{1}{64}\pi$	**48**	2	π
13	$\frac{1}{4}$	$\frac{1}{64}\pi$	**31**	1	$\frac{1}{4}\pi$	**49**	$\frac{1}{2}$	$\frac{1}{16}\pi$
14	$\frac{1}{4}$	$\frac{1}{64}\pi$	**32**	2	π	**50**	$\frac{1}{2}$	$\frac{1}{16}\pi$
15	2	π	**33**	$\frac{1}{4}$	$\frac{1}{64}\pi$	**51**	$\frac{1}{4}$	$\frac{1}{64}\pi$
16	$\frac{1}{4}$	$\frac{1}{64}\pi$	**34**	$\frac{1}{4}$	$\frac{1}{64}\pi$	**52**	$\frac{1}{2}$	$\frac{1}{64}\pi$
17	$\frac{1}{4}$	$\frac{1}{64}\pi$	**35**	$\frac{1}{2}$	$\frac{1}{16}\pi$	**53**	$\frac{1}{4}$	$\frac{1}{64}\pi$

Circle Number	Diameter (in.)	Area (in.^{2})	Circle Number	Diameter (in.)	Area (in.^{2})	Circle Number	Diameter (in.)	Area (in.^{2})
54	$1\frac{1}{2}$	$\frac{9}{16}\pi$	72	$\frac{1}{4}$	$\frac{1}{64}\pi$	90	$\frac{1}{2}$	$\frac{1}{16}\pi$
55	2	π	73	$\frac{1}{2}$	$\frac{1}{16}\pi$	91	$\frac{1}{2}$	$\frac{1}{16}\pi$
56	$\frac{1}{4}$	$\frac{1}{64}\pi$	74	$\frac{1}{4}$	$\frac{1}{64}\pi$	92	2	π
57	$\frac{1}{4}$	$\frac{1}{64}\pi$	75	$\frac{1}{2}$	$\frac{1}{16}\pi$	93	$\frac{1}{4}$	$\frac{1}{64}\pi$
58	$\frac{1}{4}$	$\frac{1}{64}\pi$	76	$\frac{1}{2}$	$\frac{1}{16}\pi$	94	$\frac{1}{4}$	$\frac{1}{64}\pi$
59	$\frac{1}{4}$	$\frac{1}{64}\pi$	77	1	$\frac{1}{4}\pi$	95	$\frac{1}{4}$	$\frac{1}{64}\pi$
60	2	π	78	$\frac{1}{4}$	$\frac{1}{64}\pi$	96	$\frac{1}{2}$	$\frac{1}{16}\pi$
61	$\frac{1}{2}$	$\frac{1}{16}\pi$	79	$\frac{1}{4}$	$\frac{1}{64}\pi$	97	$\frac{1}{4}$	$\frac{1}{64}\pi$
62	$\frac{1}{4}$	$\frac{1}{64}\pi$	80	1	$\frac{1}{4}\pi$	98	$\frac{1}{4}$	$\frac{1}{64}\pi$
63	$\frac{1}{4}$	$\frac{1}{64}\pi$	81	$\frac{1}{2}$	$\frac{1}{16}\pi$	99	$\frac{1}{4}$	$\frac{1}{64}\pi$
64	$\frac{1}{2}$	$\frac{1}{16}\pi$	82	$\frac{1}{4}$	$\frac{1}{64}\pi$			
65	1	$\frac{1}{4}\pi$	83	$\frac{1}{2}$	$\frac{1}{16}\pi$			
66	$\frac{1}{2}$	$\frac{1}{16}\pi$	84	$\frac{1}{2}$	$\frac{1}{16}\pi$			
67	$\frac{1}{4}$	$\frac{1}{64}\pi$	85	$\frac{1}{4}$	$\frac{1}{64}\pi$			
68	$\frac{1}{4}$	$\frac{1}{64}\pi$	86	1	$\frac{1}{4}\pi$			
69	$\frac{1}{2}$	$\frac{1}{16}\pi$	87	$\frac{1}{4}$	$\frac{1}{64}\pi$			
70	2	π	88	$\frac{1}{4}$	$\frac{1}{64}\pi$			
71	$\frac{1}{2}$	$\frac{1}{16}\pi$	89	$\frac{1}{4}$	$\frac{1}{64}\pi$			

Check for Students' Understanding

2

A high school principal wants to determine the students' favorite menu item in the school cafeteria. Help her by describing how each sampling method could be used to gather data.

1. Subjective sample:
   Answers will vary.
   To determine the students' favorite menu item in the school cafeteria, the principal could interview 40 students that she knows.

2. Simple random sample:
   Answers will vary.
   To determine the students' favorite menu item in the school cafeteria, the principal could interview 40 students whose names were randomly selected by a computer in the school's attendance office.

3. Convenience random sample:
   Answers will vary.
   To determine the students' favorite menu item in the school cafeteria, the principal could interview the first 40 students who entered the main office on a given day.

4. Volunteer random sample:
   Answers will vary.
   To determine the students' favorite menu item in the school cafeteria, the principal could interview the first 40 students who agreed to speak to her.

5. Stratified random sample:
   Answers will vary.
   To determine the students' favorite menu item in the school cafeteria, the principal could interview 40 randomly selected students that were in the first of four lunch periods.

6. Cluster sample:
   Answers will vary.
   To determine the students' favorite menu item in the school cafeteria, the principal could interview 10 students from each of the four grades on a given day.

7. Systemic sample:
   Answers will vary.
   To determine the students' favorite menu item in the school cafeteria, the principal could interview every 4th student that visits her office.

2.3

Sleep Tight

Using Confidence Intervals to Estimate Unknown Population Means

LEARNING GOALS

In this lesson, you will:

- Interpret the margin of error for estimating a population proportion.
- Interpret the margin of error for estimating a population mean.
- Recognize the difference between a sample and a sampling distribution.
- Recognize that data from samples are used to estimate population proportions and population means.
- Use confidence intervals to determine the margin of error of a population proportion estimate.
- Use confidence intervals to determine the margin of error of a population mean estimate.

KEY TERMS

- population proportion
- sample proportion
- sampling distribution
- confidence interval

ESSENTIAL IDEAS

- A sampling distribution is the set of sample proportions for all possible equal-sized samples.
- The formula for calculating the standard deviation of a sampling distribution to estimate a population proportion is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where $\hat{p}$(p-hat) is the sample proportion and n is the sample size.
- As the sample size increases, the standard deviation of a sampling distribution decreases.
- The confidence interval gives an estimated range of values that will likely include a population proportion or population mean.
- Confidence intervals for a population proportion or a population mean are calculated using the sample proportion or mean of a sample and the standard deviation of its sampling distribution:
  - The lower bound of a 68% confidence interval ranges from one standard deviation of the sampling distribution below the sample proportion to one standard deviation of the sampling distribution above the sample proportion.
  - The lower bound of a 95% confidence interval ranges from two standard deviations of the sampling distribution below the sample proportion to two standard deviations of the sampling distribution above the sample proportion.

- The lower bound of a 99.7% confidence interval ranges from three standard deviations of the sampling distribution below the sample proportion to three standard deviations of the sampling distribution above the sample proportion.
- The formula for calculating the standard deviation of a sampling distribution for continuous data is $\frac{S}{\sqrt{n}}$, where S is the standard deviation of the original sample and n is the sample size.

COMMON CORE STATE STANDARDS FOR MATHEMATICS

2

S-IC Making Inferences and Justifying Conclusions

Understand and evaluate random processes underlying statistical experiments

1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

Make inferences and justify conclusions from sample surveys, experiments, and observational studies

4. Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.

6. Evaluate reports based on data.

Overview

A poll measuring support for the re-election of a mayor is a context of the first scenario. Students conduct a simulation using a random number generator to increase the sample size and compare the results to the original poll. The terms population proportion, sample proportion, sampling distribution and confidence interval are defined and the formula for calculating the standard deviation of the sampling distribution is provided. The formula is applied and the confidence interval is calculated for different problem situations. A survey involving a sleep study is context of a second scenario. A formula is given to calculate the standard deviation for the population mean. Students use this formula in different scenarios to determine a range of values for population means associated with a confidence interval of 95%.

Warm Up

Determine the type of sampling method being used in each scenario. Choose from simple random sampling, cluster sampling, stratified sampling, systematic sampling, subjective sampling, convenience sampling, or volunteer sampling.

1. A math teacher chooses every third student to present their solution to the class.
   systematic sampling

2. The activity director used a lottery system to assign parking permits to 150 seniors.
   simple random sampling

3. A cafeteria worker asks the first 100 students that walk into the cafeteria to complete a satisfaction survey.
   convenience sampling

4. A counselor visits every 9th grade homeroom and gives interested students a student government survey to complete.
   volunteer sampling

5. A math teacher chooses her favorite student work to post on the wall.
   subjective sampling

2

2

2.3

Sleep Tight

Using Confidence Intervals to Estimate Unknown Population Means

2

LEARNING GOALS

In this lesson, you will:

- Interpret the margin of error for estimating a population proportion.
- Interpret the margin of error for estimating a population mean.
- Recognize the difference between a sample and a sampling distribution.
- Recognize that data from samples are used to estimate population proportions and population means.
- Use confidence intervals to determine the margin of error of a population proportion estimate.
- Use confidence intervals to determine the margin of error of a population mean estimate.

KEY TERMS

- population proportion
- sample proportion
- sampling distribution
- confidence interval

Why do we have dreams? Scientists still don't really have the answer to that question, but there have been many theories.

Some suggest that dreaming is the brain's way of discarding memories you have gathered during the day but no longer need, and studies have shown that dreaming increases as a result of learning. Another theory suggests that your brain is simply constantly churning out thoughts and images and that this doesn't stop when the rest of your body is asleep.

Some scientists are looking to evolution to provide some clues about why we dream—especially since humans don't seem to be the only animals that dream.

Why do you think some animals dream?

2

Problem 1

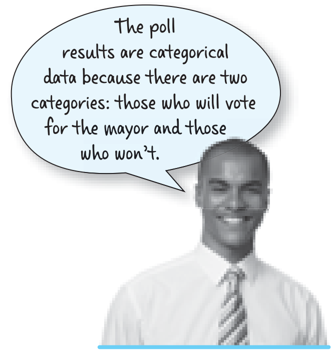
A poll measuring the support for the re-election of a mayor is conducted and a measure of error for the poll is given. Students answer questions related to the situation, and conduct a simulation using a random number generator. They combine their results with classmates to enlarge the sample size and compare the results of the simulation to the results of the original poll. The terms population proportion, sample proportion and sampling distribution are defined and the formula for calculating the standard deviation of the sampling distribution is given. Students apply the formula to the problem situation and conclude that as the sample size increases, the standard deviation of a sampling distribution decreases. Confidence interval is defined and students calculate confidence intervals.

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Questions 1 through 3 with a partner. Then have students share their responses as a class.

PROBLEM 1 Every Vote Counts: Exploring Categorical Data

In a poll of 1100 registered voters before an upcoming mayoral election, 594 people, or 54%, said they would vote to re-elect the current mayor, while the remaining voters said they would not vote for the mayor. The margin of error for the poll was ± 3 percent, which means that the poll predicts that somewhere between 51% (54% − 3%) and 57% (54% + 3%) of people will actually vote to re-elect the mayor.



1. Does the poll represent a sample survey, an observational study, or an experiment?

   The poll represents a sample survey.

2. Based on the poll, can you conclude that the current mayor will be re-elected? Explain your reasoning.

   Answers will vary.

   The results of the poll suggest that the current mayor will be re-elected because the range of votes, 51% to 57%, is more than the 50% needed to win. However, the actual result of the election could be different for a number of reasons. People who responded to the poll may not vote, might change their minds, or may not have answered the poll question honestly.

3. Is it possible for fewer than 50% of respondents in a new sample to respond that they will vote for the mayor in the election? Is it likely? Explain your reasoning.

   Answers will vary.

   It is possible for fewer than 50% of respondents in a new sample to support the mayor for re-election. I would say that 50% is somewhat likely because the number is just below the 51% to 57% range.

Guiding Questions for Share Phase, Questions 1 through 3

- What is the difference between a sample survey, an observational study, and an experiment?
- What is the margin of error?
- What does the margin of error represent with respect to the problem situation?
- What does the margin of error suggest with regard to the lower bound?
- What does the margin of error suggest with regard to the upper bound?

continued on the next page

- How is 54% calculated?
- What percent is needed to win the mayoral election?
- Do all people that respond to a poll about the election definitely vote in the election?
- Do all people honestly respond in polls?
- How does the margin of error support the mayor's chances of re-election?
- Why is it somewhat likely the mayor will receive 50% of the vote?

Grouping

Have students complete Question 4 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Question 4

- How many students are in this class?
- How do we determine the size of each student's sample?
- What is the sample size for which each student is responsible?
- In your sample, how many votes will be cast to re-elect the mayor?
- In your sample, how many votes will be cast to not re-elect the mayor?
- When your sample is combined with your classmate's sample, how many votes will be cast to re-elect the mayor?

2

4. With your classmates, conduct a simulation to represent polling a new sample of 1100 voters.

a. Divide 1100 by the number of students in your class to determine the size of each student's sample.

Answers will vary.

There are 20 students in my class and 1100 divided by 20 equals 55. So, each student can conduct a simulation that has a sample size of 55.

b. Generate an amount of random numbers equal to the sample size in part (a) to represent responses to the polling question. Generate random numbers between 1 and 100, with numbers from 1 to 54 representing support for re-electing the mayor and the numbers 55 to 100 representing support for not re-electing the mayor. Tally the results of your simulation, and then list the total number of tallies for each category.

Number of People Who Respond that They Will Vote to Re-elect the Mayor	Number of People Who Respond that They Will Not Vote to Re-elect the Mayor
𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 |	𝍸 𝍸 𝍸 ||||

Answers will vary.

In my simulation, 36 out of 55 people state that they will vote to re-elect the mayor and 19 out of 55 state that they will vote to not re-elect the mayor.

- When your sample is combined with your classmate's sample, how many votes will be cast to not re-elect the mayor?
- How was the percent of people supporting the re-election determined?
- How was the percent of people not supporting the re-election determined?
- What were the results of the original poll?
- Are the results of the simulation higher or lower than the results of the original poll?
- Do you think the results of the simulation should be close to 54%?

2

c. Calculate the percent of people who state that they will vote to re-elect the mayor and the percent of people who state that they will vote to not re-elect the mayor based on your simulation.

Answers will vary.

Based on my simulation, approximately 65.45% of people state that they will vote to re-elect the mayor, and approximately 34.55% of people state that they will vote to not re-elect the mayor.

d. Complete the simulation for the 1100 voters by combining the data from your classmates. List the percent of votes for each category.

Percent of People Who Respond that They Will Vote to Re-elect the Mayor	Percent of People Who Respond that They Will Vote to Not Re-elect the Mayor
56%	44%

Answers will vary.

e. Are the results of the simulation different from the results of the original poll? Explain.

Answers will vary.

In the original poll, 54% of the people supported re-electing the mayor. In my simulation, 56% of people supported re-electing the mayor.

f. If you conducted the simulation over and over, would you expect to get the same results or different results each time? Explain your reasoning.

It is possible to get some results that are the same, but it is more likely to get a different result for each simulation. The results for each simulation will probably be close to 54% in favor of the mayor's re-election.

Grouping

Ask a student to read the information and Worked Example. Discuss as a class.

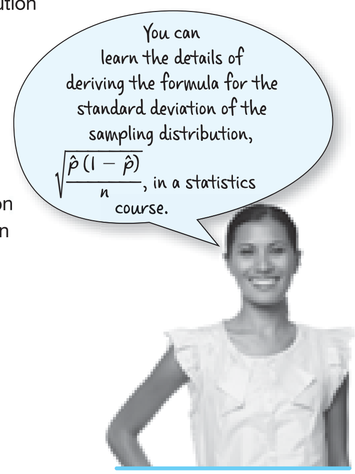
Guiding Questions for Discuss Phase

- What is the difference between a sample proportion and a population proportion?
- In Question 4, did you and your classmates create a sampling distribution? Explain.
- How many sample proportions are used to create a sampling distribution?
- How can you calculate the standard deviation of a sampling distribution?

The percent of voters who actually vote for the mayor in the election is the **population proportion**. The percent of voters in the sample who respond that they will vote for the mayor is the **sample proportion**. The population proportion and sample proportion are measures used for discrete, or categorical, data. For continuous data, these are called the population mean and sample mean.

When you and your classmates generated random numbers to simulate multiple samples of the 1100 voters, you came up with different sample proportions. The set of all of your classmates' sample proportions is part of a *sampling distribution*.

A **sampling distribution** is the set of sample proportions for all possible equal-sized samples. A sampling distribution will be close to a normal distribution, and the center of a sampling distribution is a good estimate of a population proportion—in this case, the percent of people who will actually vote to re-elect the mayor.



But rather than collecting a very large number of samples, a more practical method for estimating a population proportion is to use the sample proportion of a single sample to estimate the standard deviation of the sampling distribution. The standard deviation of a sampling distribution can give you a range in which the population proportion is likely to fall, relative to the sample proportion.

For example, to estimate the standard deviation of the sampling distribution for the sample of 1100 voters, you can use the formula $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where ($\hat{p}$) is the sample proportion and n is the sample size.

The sample proportion from the original poll is 54%, or 0.54. This is the percent of the 1100 people in the poll who said they would vote to re-elect the mayor.

The standard deviation of the sampling distribution for this poll is

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.54(1-0.54)}{1100}}$$
$$\approx 0.0150$$

This means that 1 standard deviation below the sample proportion of 54% is 54% − 1.5%, or 52.5%. And 1 standard deviation above the sample proportion of 54% is 54% + 1.5%, or 55.5%.

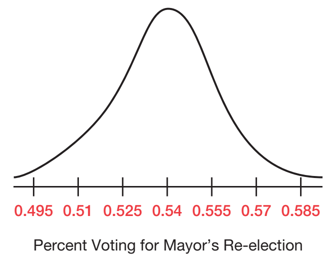
Grouping

Have students complete Questions 5 and 6 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 5 and 6

- What is the difference between a population proportion and a sample proportion?
- Which term is associated with the voters that actually vote?
- Which term is associated with the voters that respond on the poll that they will vote?
- Why are sampling distributions approximately normal?
- How is the center of a sampling distribution related to the population proportion?
- What does the standard deviation for a sampling distribution mean with respect to the problem situation?
- Are the smaller samples generally more spread out from the center? Why?
- Are the larger samples generally more clustered around the center? Why?
- Can the formula for the standard deviation of a sampling distribution of sample proportions be used to support Bobbie's claim?

5. Use the sample proportion and standard deviation of the sampling distribution to label the horizontal axis of the normal curve.



6. Bobbie made an observation about the standard deviation of a sampling distribution.

> *Bobbie*
>
> *The standard deviation of a sampling distribution gets smaller and smaller as the size of the sample gets larger and larger.*

Is Bobbie's statement correct? Explain why or why not.

Bobbie's statement is correct because smaller samples are generally more spread out from the center and larger samples are more clustered around the center.

I can also use the formula for the standard deviation of a sampling distribution of sample proportions, $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, to justify Bobbie's statement.

If the sample size was increased to 10,000 and the sample proportion remained the same, then the standard deviation of the sampling distribution of sample proportions would be less than 0.0150. This occurs because the standard deviation formula involves division by the sample size.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.54(1-0.54)}{10{,}000}}$$

$$= \sqrt{\frac{0.54(0.46)}{10{,}000}}$$

$$\approx 0.0050$$

- In the formula, if the value in the numerator stays the same as the value in the denominator increases, what happens to the value of the radical?
- What is the value of the mean on the normal curve?
- What is the value of one standard deviation on the normal curve?
- What is the value of two standard deviations on the normal curve?

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Questions 7 through 9 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Question 7

- How is the confidence interval associated with 68% calculated?
- How is the confidence interval associated with 95% calculated?
- How is the confidence interval associated with 99.7% calculated?

An estimated range of values that will likely include the population proportion or population mean is called a **confidence interval**. When stating the margin of error, a 95% confidence interval is typically used. However, other confidence intervals may also be used.

For example, the standard deviation of the sampling distribution for the election sample is 0.015, or 1.5%. Two standard deviations is 3%, so the margin of error is reported as $\pm 3\%$.

Confidence intervals for a population proportion are calculated using the sample proportion of a sample and the standard deviation of the sampling distribution.

- The lower bound of a 68% confidence interval ranges from 1 standard deviation below the sample proportion to 1 standard deviation above the sample proportion.
- The lower bound of a 95% confidence interval ranges from 2 standard deviations below the sample proportion to 2 standard deviations above the sample proportion.
- The lower bound of a 99.7% confidence interval ranges from 3 standard deviations below the sample proportion to 3 standard deviations above the sample proportion.

7. Determine each confidence interval for the election poll.

a. 68%

The interval between 52.5% and 55.5% represents a 68% confidence interval.

b. 95%

The interval between 51.0% and 57.0% represents a 95% confidence interval.

c. 99.7%

The interval between 49.5% and 58.5% represents a 99.7% confidence interval.

2

Guiding Questions for Share Phase, Questions 8 and 9

- When stating a margin of error, why do you suppose a 95% confidence interval is typically used?
- How are confidence intervals for a population proportion calculated?
- Are the lower bounds of the confidence intervals associated with the Empirical Rule of Normal Distribution?
- How are the lower bounds of the confidence intervals associated with the Empirical Rule of Normal Distribution?
- What value is associated with one standard deviation of the sampling distribution of sample proportions?
- What value is associated with two standard deviations of the sampling distribution of sample proportions?
- How is the value associated with two standard deviations of the sampling distribution of sample proportions related to the margin of error?

2

8. Explain the similarities and differences between each confidence interval for the election poll.

The confidence intervals are based on the sample proportion and the standard deviation of the sampling distribution of sample proportions.

A 68% confidence interval is based on the data values of the sampling distribution that are between 1 standard deviation below the sample proportion and 1 standard deviation above the sample proportion.

A 95% confidence interval is based on the data values of the sampling distribution that are between 2 standard deviations below the sample proportion and 2 standard deviations above the sample proportion.

A 99.7% confidence interval is based on the data values of the sampling distribution that are between 3 standard deviations below the sample proportion and 3 standard deviations above the sample proportion.

9. The result of the original poll was 54% with 3% margin of error. What confidence interval does 3% represent? Explain your reasoning.

The 3% margin of error in the original poll represents a 95% confidence interval.

I know this because a 95% confidence interval is based on the data values of the sampling distribution of sample proportions that are between 2 standard deviations below the sample proportion and 2 standard deviations above the sample proportion. For this election poll, one standard deviation of the sampling distribution of sample proportions is equal to 0.015 and two standard deviations of the sampling distribution of sample proportions is 0.03, or 3%.

Grouping

Have students complete Question 10 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Question 10

- What is the value of *p*-hat in this situation?
- What is the value of *n* in this situation?
- What is the value of one standard deviation of the sampling distribution in this situation?
- What is the value of two standard deviations of the sampling distribution in this situation?
- How can you determine the lower bound related to the range of values for the population proportion?
- How can you determine the upper bound related to the range of values for the population proportion?

10. Use a 95% confidence interval to determine a margin of error and a range of values for each population proportion.

a. A survey of 1500 teenagers shows that 83% do not like waking up early in the morning.

The interval from 81% to 85% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.83(1-0.83)}{1500}}$$
$$= \sqrt{\frac{0.83(0.17)}{1500}}$$
$$\approx 0.010$$

b. A survey of 200 licensed high school students shows that 16% own their own car.

The interval from 10.8% to 21.2% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.16(1-0.16)}{200}}$$
$$= \sqrt{\frac{0.16(0.84)}{200}}$$
$$\approx 0.026$$

c. A survey of 500 high school students shows that 90% say math is their favorite class.

The interval from 87.4% to 92.6% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.90(1-0.90)}{500}}$$
$$= \sqrt{\frac{0.90(0.10)}{500}}$$
$$\approx 0.013$$

2

Problem 2

The sample mean and sample standard deviation of a sleep survey is the context of the problem situation. Students use a formula to calculate the standard deviation for the population mean and use the standard deviation to determine a 95% confidence interval for the population mean. They also use a 95% confidence interval to determine a range of values for various population means.

Grouping

- Ask a student to read the information. Complete Question 1 and discuss as a class.
- Have students complete Questions 2 through 5 with a partner. Then have students share their responses as a class.

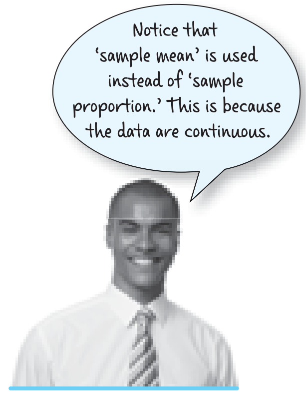
Guiding Questions for Discuss Phase, Question 1

- Do you suppose the means for each sample will be close to 7.7? Why?
- What mean values would be considered close to 7.7 hours?

PROBLEM 2 Sweet Dreams: Exploring Continuous Data

A sample of 50 students at High Marks High School responded to a survey about their amount of sleep during an average night. The sample mean was 7.7 hours and the sample standard deviation was 0.8 hour.

Let's determine an estimate for the population mean sleep time for all High Marks High School students.



1. If you gathered data from many new samples, would you expect the samples to have equal means or different means? Explain your reasoning.

   It is possible that some of the sample means would be equal, but it is more likely that the sample means would be different. The means for each sample will probably be close to 7.7.

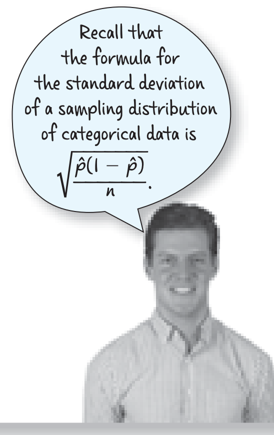
Collecting additional samples of 50 students and plotting the sample mean of each sample will result in a sampling distribution. The sampling distribution will be approximately normal, and the mean of the sampling distribution is a good estimate of the population mean.

Just like with the categorical data, a more practical method for estimating the population mean amount of sleep for High Marks High School students is to use the sample mean to calculate an estimate for the standard deviation of the sampling distribution. The formula for the standard deviation of a sampling distribution for continuous data is $\frac{s}{\sqrt{n}}$, where s is the standard deviation of the original sample and n is the sample size.

2. Use the standard deviation from the original sample to determine the standard deviation for the sampling distribution. Explain your work.

   The standard deviation for the sampling distribution is approximately 0.11 hour.

   $\frac{s}{\sqrt{n}} = \frac{0.8}{\sqrt{50}} \approx 0.11$



Guiding Questions for Share Phase, Question 2

- What is the standard deviation from the original sample?
- What information is needed to calculate the standard deviation for the population mean?

Guiding Questions for Share Phase, Questions 3 through 5

- How is the standard deviation used to determine the upper bound of a 95% confidence interval for the population mean?
- How is the standard deviation used to determine the lower bound of a 95% confidence interval for the population mean?
- Is the margin of error associated with one standard deviation or two standard deviations above and below the mean?
- What information is needed to determine the margin of error?
- How is the margin of error calculated?
- Which formula is used to solve this problem?
- What information is needed to use the formula?
- What is the value of n in this situation?
- What is the value of S in this situation?
- Once the standard deviation of the mean population is determined, how is it used to determine the range of values?

3. Use the standard deviation of the sampling distribution to determine a 95% confidence interval for the population mean. Explain your work.

A 95% confidence interval for the population mean is 7.48 hours to 7.92 hours.

I used the value of 2 standard deviations from the mean to calculate the confidence interval.

$7.7 - 2(0.11) = 7.48$

$7.7 + 2(0.11) = 7.92$

4. Write the 95% confidence interval in terms of the population mean plus or minus a margin of error.

In terms of the mean of the sampling distribution of sample means plus or minus the margin of error, the 95% confidence interval is 7.7 ± 0.22.

Two standard deviations is $2(0.11) = 0.22$.

5. Use a 95% confidence interval to determine a range of values for each population mean.

a. A sample of 75 students responded to a survey about the amount of time spent online each day. The sample mean was 3.2 hours, and the standard deviation of the sampling distribution was 0.9 hour.

The interval from 3.0 to 3.4 represents a 95% confidence interval for the population mean.

$\frac{s}{\sqrt{n}} = \frac{0.9}{\sqrt{75}} \approx 0.10$

Two standard deviations is $2(0.10) = 0.2$.

b. A sample of 1000 teachers responded to a survey about the amount of time they spend preparing for class outside of school hours. The sample mean was 2.5 hours, and the standard deviation of the sampling distribution was 0.5 hour.

The interval from 2.46 to 2.56 represents a 95% confidence interval for the population mean.

$\frac{s}{\sqrt{n}} = \frac{0.5}{\sqrt{1000}} \approx 0.02$

Two standard deviations is $2(0.02) = 0.04$.

c. A sample of 400 adults responded to a survey about the distance from their home to work. The sample mean was 7.8 miles, and the standard deviation of the sampling distribution was 1.6 miles.

The interval from 7.64 to 7.96 represents a 95% confidence interval for the population mean.

$\frac{s}{\sqrt{n}} = \frac{1.6}{\sqrt{400}} = 0.08$

Two standard deviations is $2(0.08) = 0.16$.

Talk the Talk

Students distinguish between the terms sample and sampling distribution, sample proportion and sample mean, and standard deviation and margin of error.

2

Grouping

Have students complete Questions 1 and 2 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 1 and 2

- Which term is associated with collecting all of the possible samples of equal sizes from a population?
- Which term is associated with a set of data that is collected from a population?
- Which term(s) are associated with measures of center for a sample?
- Which term is associated with only categorical data?
- Which term is associated with only continuous data?
- Which term is associated with the measure of spread used to create a confidence interval?
- Which term is associated with a range of numbers estimating a population mean, or population proportion?
- The margin of error is usually associated with which percent confidence interval?

Talk the Talk

1. What is the difference between a sample and a sampling distribution?

   A sample is set of data that is collected from a population. A sampling distribution is the set of all sampling proportions that result from collecting all of the possible samples from a population. Sampling distributions are approximately normal, and the mean of a sampling distribution is a good estimate of the population mean.

2. What is the difference between a sample proportion and a sample mean?

   Sample proportions and sample means are measures of center for a sample.

   The term sample proportion is used for categorical data. The election scenario in Problem 1 is an example of a situation using a sample proportion. Fifty-four percent of the sample respondents were in favor of re-electing the mayor.

   The term sample mean is used for continuous data. The average amount of sleep scenario is Problem 2 is an example of a sample mean. The amount of sleep in hours is continuous data.

Online Time Study, Part III

Students recall the scenario from the previous lessons about the online survey, calculate a 95% confidence interval, and label the normal curve of the sampling distribution.

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Questions 1 through 3 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 1 and 2

- Which formula is used to solve this problem?
- What information is needed to use the formula?
- What is the value of *n* in this situation?
- What is the value of *S* in this situation?
- Once the standard deviation of the mean population is determined, how is it used to determine the range of values?

Online Time Study, Part III

To summarize data from a sample survey, observational study, or experiment:

- Calculate measures of center.
- Calculate measures of spread.
- Select the most appropriate method(s) to display the data (dot plot, histogram, stem-and-leaf plot, box-and-whisker plot, normal curve).
- Describe the characteristics of the graphical display.

To analyze data from a sample survey, observational study, or experiment:

- Use confidence intervals to determine a range of values for the population mean(s) or proportion(s).

Recall the study described in previous lessons about the amount of time students in your school are online each day.

1. Will your study involve estimating a population mean or a population proportion? Explain your reasoning.

My study will involve recording amounts of time spent online, which are continuous data. So, I will be estimating the population mean.

2. Use a 95% confidence interval to determine a range of values for the population mean, given a random sample of 60 students with a sample mean of 3.5 hours and a standard deviation of 1.1.

The interval from 3.22 to 3.78 represents a 95% confidence interval for the population mean.

$\frac{s}{\sqrt{n}} = \frac{1.1}{\sqrt{60}} \approx 0.14$

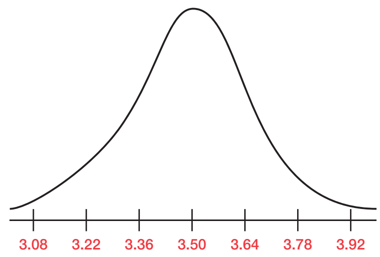
$2(0.14) = 0.28$

Guiding Questions for Share Phase, Question 3

- Describe how to label the horizontal axis.

2

3. Use the sample mean and standard deviation of the sampling distribution to label the horizontal axis of the normal curve.



Be prepared to share your results and methods.

Check for Students' Understanding

1. A survey of 500 students reports that 82% will attend the winter carnival. Determine a range of values for the population proportion. Use a 95% confidence interval.

   The interval from 78.6% to 85.4% represents a 95% confidence interval for the population proportion.

   $$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.82(1-0.82)}{500}}$$
   $$= \sqrt{\frac{0.82(0.18)}{500}}$$
   $$\approx 0.017$$

   $82 - 2(1.7) = 78.6$

   $82 + 2(1.7) = 85.4$

2. A sample of 300 dog owners responded to a survey about the amount of money they spend on dog chew toys each year. The sample mean was \$150 and the sample standard deviation was \$12. Determine a range of values for the population mean. Use a 95% confidence interval.

   The interval from \$148.62 to \$151.38 represents a 95% confidence interval for the population mean.

   $$\frac{S}{\sqrt{n}} = \frac{12}{\sqrt{300}}$$
   $$\approx 0.6928$$

   $150 - 2(0.69) = 148.62$

   $150 + 2(0.69) = 151.38$

2

2

2.4

How Much Different?

Using Statistical Significance to Make Inferences About Populations

LEARNING GOALS

In this lesson, you will:

- Use sample proportions to determine whether differences in population proportions are statistically significant.
- Use sample means to determine whether differences in population means are statistically significant.

KEY TERM

- statistically significant

ESSENTIAL IDEAS

- Statistically significant are unlikely to have occurred by chance.
- Sample proportions or means more than 2 standard deviations from the proportion or mean of its sampling distribution are considered statistically significant.

COMMON CORE STATE STANDARDS FOR MATHEMATICS

S-IC Making Inferences and Justifying Conclusions

Understand and evaluate random processes underlying statistical experiments

1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

2. Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation.

Make inferences and justify conclusions from sample surveys, experiments, and observational studies

4. Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.

5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

6. Evaluate reports based on data.

Overview

Within the context of a situation, students use a sample proportion or mean, and a 95% confidence interval to determine the margin of error. The sample proportion or mean and standard deviation of its sampling distribution are used to label the axis of a normal curve and students determine whether or not results are statistically significant. A random number generator is used to create an additional sample for the purpose of comparison.

2

Warm Up

1. A satisfaction survey was given to 220 gym members. Ninety-one percent of the respondents in the sample reported that they work out regularly and intend to renew their contract with the gym for an additional year. Determine a range of values for the population proportion. Use a 95% confidence interval.

   The interval from 87.2% to 94.8% represents a 95% confidence interval for the population proportion.

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

   $$\sqrt{\frac{0.91(1 - 0.91)}{220}} = \sqrt{\frac{0.91(0.09)}{220}} \approx 0.019$$

   $91 - 2(1.9) = 87.2$

   $91 + 2(1.9) = 94.8$

2. A sample of 630 grocery store shoppers responded to a survey about the average cost of their groceries each week. The sample mean was \$125 and the sample standard deviation was \$20. Determine a range of values for the population mean. Use a 95% confidence interval.

   The interval from \$123.40 to \$126.60 represents a 95% confidence interval for the population mean.

   $$\frac{s}{\sqrt{n}}$$

   $$\frac{20}{\sqrt{630}} \approx 0.7968$$

   $125 - 2(0.80) = 123.40$

   $125 + 2(0.80) = 126.60$

2

2

2

How Much Different?

2.4

Using Statistical Significance to Make Inferences About Populations

LEARNING GOALS

In this lesson, you will:

- Use sample proportions to determine whether differences in population proportions are statistically significant.
- Use sample means to determine whether differences in population means are statistically significant.

KEY TERM

- statistically significant

A person's blood pressure is typically measured using two numbers. One number represents the pressure in the arteries when the heart beats. This is the systolic pressure. The other number represents the pressure in the arteries between heartbeats. This is the diastolic pressure. For example, $\frac{118}{74}$ represents a systolic pressure of 118 and a diastolic pressure of 74.

Problem 1

An experiment compares two different types of water. Students discuss the results of the experiment and determine a sample proportion. The sample proportion, a confidence interval of 95% for the population proportion and the formula are used to determine the margin of error and the range of values for the population proportion. Students use the sample proportion and standard deviation of the sampling distribution to label the axis of a normal curve. They conclude the difference between the two different types of water is not statistically significant. Another survey is conducted to determine consumer preference. Students calculate the margin of error and the range of values for the population proportion. Using the sample proportion and standard deviation of the sampling distribution, students label the axis on a second normal curve. This time, students conclude the difference between the two different types of water is statistically significant. A random number generator is used to create new samples and the results are checked for statistical significance.

2

PROBLEM 1 Whatta Water: Exploring Categorical Data

Commercials on a local TV station claim that Whatta Water tastes better than tap water, but a local news anchor does not believe the claim. She sets up an experiment at a local grocery store to test the claim. A representative, unbiased sample of 120 shoppers participate in the tasting survey using unmarked cups. Out of the 120 people, 64 said Whatta Water tastes better than tap water.

1. If shoppers had to choose one or the other and there was no difference in the tastes of the two waters, what proportion of shoppers would you expect to say that Whatta Water tastes better? Explain your reasoning.

   If there was no difference in the tastes of the two waters and shoppers had to choose one or the other, then I would expect that 50% of shoppers would have responded that Whatta Water tastes better, and 50% would have said that tap water tastes better.

2. What is the sample proportion of shoppers who stated that Whatta Water tastes better?

   The sample proportion of shoppers who stated that Whatta Water tastes better is approximately 53%.

   $\frac{64}{120} = 0.5\overline{3}$

3. Based on your answers to Questions 1 and 2, what reason(s) can you give to doubt Whatta Water's claim? Explain your reasoning.

   Answers will vary.

   The sample proportion of people who preferred Whatta Water is only a little higher than 50%. If Whatta Water's claim was true, I would expect a higher proportion of people to prefer Whatta Water's taste.

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Questions 1 through 3 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 1 through 3

- How many shoppers are half of all of the shoppers?
- Half of the shoppers are represented by what percent?
- What information is needed to determine the sample proportion of shoppers who preferred Whatta Water?
- How is the sample proportion of shoppers who preferred Whatta Water determined?
- How does the sample proportion of people who preferred Whatta Water compare to 50%?
- If Whatta Water's claim was true, would you expect a higher proportion of people preferring the taste of Whatta Water?

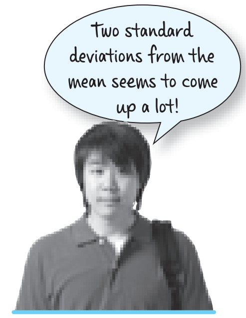
Grouping

- Ask a student to read the definition. Discuss as a class.
- Have students complete Questions 4 and 6 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 4 through 6

- What information is needed to use the formula to determine the margin of error?

The term **statistically significant** is used to indicate that a result is very unlikely to have occurred by chance. Typically, a result that is more than 2 standard deviations from the mean, or outside a 95% confidence interval, is considered statistically significant.



4. Use a 95% confidence interval to determine a range of values for the population proportion of people who prefer the taste of Whatta Water. Explain your work.

The interval from 43.8% to 62.2% represents a 95% confidence interval for the population proportion.

The margin of error is approximately ±9.2%.

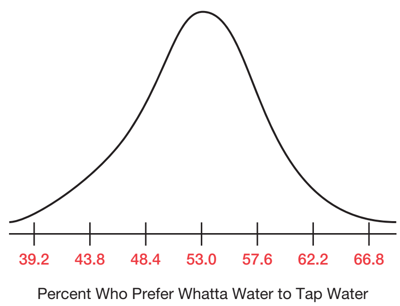
$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.53(1-0.47)}{120}}$$

$$= \sqrt{\frac{0.53(0.47)}{120}}$$

$$\approx 0.046$$

$2(0.046) = 0.092$

5. Use the sample proportion and standard deviation of the sampling distribution to label the horizontal axis of the normal curve.



6. Based on the range of values of the 95% confidence interval, what conclusion can you make about Whatta Water's claim that their water tastes better than tap water?

Whatta Water's claim is probably not true, because 50% is in the 95% confidence interval. The sample proportion of 53% is not statistically significant.

- What formula is used to determine the margin of error?
- How is the margin of error used to determine a range of values for the population proportion?
- What is the mean of the normal distribution curve?
- What is the value of one standard deviation?
- Is 50% in the 95% confidence interval?
- Is the difference between 53% and 50% statistically significant? Why not?
- What determines statistical significance?

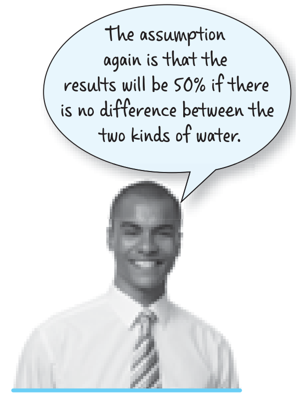
Grouping

Have students complete Questions 7 through 10 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Question 7

- What information is needed to use the formula to determine the margin of error?
- What formula is used to determine the margin of error?
- How is the margin of error used to determine a range of values for the population proportion?
- How does the margin of error in this sample compare to the margin of error in the previous sampling?
- What is the mean of the normal distribution curve?
- What is the value of one standard deviation?
- Is 50% in the 95% confidence interval?
- Is the difference between 34% and 50% statistically significant? Why?
- What determines statistical significance?

7. The local water company also conducted a survey of 120 people which they said showed that people prefer tap water over Whatta Water. Forty-one of the respondents said Whatta Water tastes better.



a. Use a 95% confidence interval to determine a range of values for the population proportion of people who prefer Whatta Water. Explain your work.

The interval from 25.4% to 42.6% represents a 95% confidence interval for the population proportion.

The margin of error is approximately ±8.6%.

$$\frac{41}{120} \approx 0.34$$

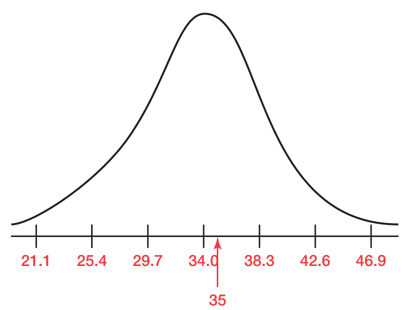
$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.34(1-0.34)}{120}}$$

$$= \sqrt{\frac{0.34(0.66)}{120}}$$

$$\approx 0.043$$

$$2(0.043) = .086$$

b. Use the sample proportion and standard deviation of the sampling distribution to label the horizontal axis of the normal curve.



c. Based on the range of values of the 95% confidence interval, what conclusion can you draw about the local water company's claim that tap water tastes better than Whatta Water?

The local water company's claim is probably true because 50% is not in the 95% confidence interval. The difference between 34% and 50% is statistically significant.

Guiding Questions for Share Phase, Questions 8 through 10

- Are the sample results created using the random number generator different than the previous sample? How?
- Is the location of the sample proportion close to the mean?
- Where is the location of the sample proportion in terms of standard deviations?
- Did the results of your simulation result in a sample proportion that is within the 95% confidence interval? What does this imply about statistical significance?

8. Use a random number generator to conduct a simulation of the local water company's survey, for a new sample of 120 people. Generate a random number between 1 and 100, with numbers from 1 to 34 representing that Whatta Water tastes better and numbers from 35 to 100 representing that tap water tastes better. List the results in the table.

Percent of People in Simulation Who Said Whatta Water Tastes Better	Percent of People in Simulation Who Said Tap Water Tastes Better
35%	65%

Answers will vary.

9. On the normal curve in Question 7 part (b), locate and mark the sample proportion of your simulation. Describe the location of the sample proportion on the normal curve.

Answers will vary.

The location of the sample proportion is close to the mean, less than 1 standard deviation above the mean.

10. Compare the results of your simulation with the water company's study and with Whatta Water's study. Are your results significantly different? Explain your reasoning.

Answers will vary.

The results of my simulation are not significantly different from the water company's study because the sample proportion of my simulation, 35%, is within the 95% confidence interval. The results of my simulation are significantly different from Whatta Water's survey because the sample proportion of 50% still more than 2 standard deviations away from my simulation results.

2

Problem 2

The sample mean and sample standard deviation of a homework survey is the context of the problem situation. Students calculate the standard deviation for a population mean estimate and use it to determine a 95% confidence interval.

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Questions 1 through 4 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 1 through 3

- What information is needed to use the formula to determine the margin of error?
- What formula is used to determine the margin of error?
- How is the margin of error used to determine a range of values for the population mean?
- What is the mean of the normal distribution curve?
- What is the value of one standard deviation?

PROBLEM 2 Nonstop Homework: Exploring Continuous Data

A sample of 40 students at High Marks High School responded to a survey about the average amount of time spent on homework each day. The sample mean was 2.9 hours and the sample standard deviation was 0.8 hour.

1. Use a 95% confidence interval to determine a range of values for the population mean. Explain your work.

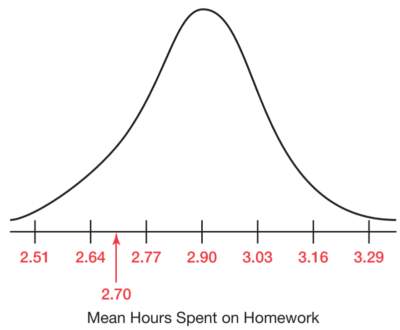
   The interval from 2.64 to 3.16 hours represents a 95% confidence interval for the population mean.

   The margin of error is approximately ±0.26.

   $\frac{s}{\sqrt{n}} = \frac{0.8}{\sqrt{40}} \approx 0.13$

   $2(0.13) = 0.26$

2. Label the horizontal axis of the normal curve that represents the sampling distribution.



3. A new sample of 40 students was taken and the resulting sample mean was 2.70 hours.

   a. On the normal curve in Question 2, locate and mark the sample mean of the new sample. Describe the location of the sample mean on the normal curve.

   Answers will vary.

   The location of the sample mean is almost 2 standard deviations below the mean of the sampling distribution.

Guiding Questions for Share Phase, Question 4

- If the sample mean value of the new sampling is statistically significant, where will it be located on the normal distribution curve with respect to the mean of the sampling distribution of sample means?
- If the sample mean value of the new sampling is not statistically significant, where will it be located on the normal distribution curve with respect to the mean of the sampling distribution of sample means?
- Which values are outside of the 95% confidence interval?
- Which values are within the 95% confidence interval?

Grouping

Have students complete Question 5 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 5

- What does statistical significance compare?
- Is statistical significance used to compare an individual data value to a population mean?
- Is statistical significance used to compare sample means to population means?

b. Are the results of the new sample statistically significant? Explain your reasoning.

The results of the new simulation are not statistically significant because the sample mean, 2.70, lies within the 95% confidence interval.

4. What sample mean values are statistically significant? Explain your reasoning.

Sample mean values less than 2.64 hours and greater than 3.16 hours are statistically significant because those values are outside of the 95% confidence interval.

5. Mary shared a comment about the time she spends on homework.

Mary

I spend an average of 3.5 hours on homework every night. Compared to the sample mean, the average amount of time I spend on homework every night is statistically significant.

Is Mary's reasoning valid? Explain why or why not.

Mary's reasoning is not valid because statistical significance is used to compare sample means to population means, not an individual data value and a population mean.

- Is Mary comparing an individual data value to a population mean or sample means to population means?

2

Problem 3

Two local newspapers conducting a poll with regard to a tax increase is the context of the problem situation. The number of residents polled and the number of residents in favor of the tax increase are given. Students determine the sample proportion for each poll, and use the results to estimate a range of values for a 95% confidence interval. If the two confidence intervals overlap, then the difference is not statistically significant whereas if they do not overlap, the difference is statistically significant. A third local newspaper poll is part of the scenario. Students calculate the 95% confidence interval for the third newspaper poll and determine whether its results are statistically significant in comparison with the other two newspaper polls.

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Questions 1 through 4 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 1 and 2

- How is the sample proportion for each newspaper poll determined?

2

PROBLEM 3 Read Between the Lines: Comparing Categorical Data

Two hometown newspapers conducted a poll about whether residents are for or against a tax to provide funding for school renovations in the district. Today's News polled 75 residents and 53 stated that they are in favor of the tax increase. Local Time polled 100 residents and 54 stated they are in favor of the tax increase.

1. Calculate the sample proportion for each poll.

   The sample proportion for the Today's News poll is $\frac{53}{75}$, or about 71%.

   The sample proportion for the Local Time poll is $\frac{54}{100}$, or 54%.

2. Use the results from each poll to estimate a range of values for the population proportion using a 95% confidence interval. Explain your work.

   For the Today's News poll, the margin of error is approximately ±0.104. So, the interval from 60.6% to 81.4% represents a 95% confidence interval for the population proportion.

   For the Local Time poll, the margin of error is approximately ±0.10. So, the interval from 44% to 64% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.71(1-0.71)}{75}}$$
$$= \sqrt{\frac{0.71(0.29)}{75}}$$
$$\approx 0.052$$
$$2(0.052) = 0.104$$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.54(1-0.54)}{100}}$$
$$= \sqrt{\frac{0.54(0.46)}{100}}$$
$$\approx 0.050$$
$$2(0.050) = 0.100$$

- What information is needed to use the formula to determine the margin of error for each newspaper poll?
- What formula is used to determine the margin of error for each newspaper poll?
- How is the margin of error used to determine a range of values for the population proportion for each newspaper poll?

Guiding Questions for Share Phase, Questions 3 and 4

- When the population proportion intervals overlap, what does this imply with regard to statistical significance of the estimates?
- When the population proportion intervals do not overlap, what does this imply with regard to statistical significance of the estimates?
- How does the margin of error associated with the third newspaper compare to the margin of error associated with the first two newspapers?
- How does the range of values of the population proportion associated with the third newspaper compare to the population proportion associated with the first two newspapers?
- Do the population proportion intervals overlap with regard to the Reporter and Local Times? What does this imply in terms of statistical significance?
- Do the population proportion intervals overlap with regard to the Reporter and Today's News? What does this imply in terms of statistical significance?

2

3. The Reporter newspaper published a survey of 90 residents and 38 stated that they are in favor of the tax increase. Use a 95% confidence interval to determine a range of values for the population proportion. Explain your work.

For the Reporter poll, the margin of error is approximately ±0.104. So, the interval from 31.6% to 52.4% represents a 95% confidence interval for the population proportion.

$$\frac{38}{90} \approx 0.42$$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.42(1-0.42)}{90}}$$

$$= \sqrt{\frac{0.42(0.58)}{90}}$$

$$\approx 0.052$$

$$2(0.052) = 0.104$$

If two confidence intervals overlap, then the difference between the population proportions or population means is not statistically significant. If the intervals do not overlap then the difference between the population proportions or population means is statistically significant.

4. Compare the population proportion estimates and determine whether their differences are statistically significant. Explain your reasoning.

a. The Reporter and Local Times

The difference between the population proportion estimates of the Reporter and Local Time polls is not statistically significant because the two population proportion intervals overlap.

The 95% confidence interval for the population proportion of the Reporter poll ranges from 31.6% to 52.4%.

The 95% confidence interval for the population proportion of the Local Time poll ranges from 44% to 64%.

b. The Reporter and Today's News

The difference between the population proportion estimates of the Reporter Today's News polls is statistically significant because the two population proportion intervals do not overlap.

The 95% confidence interval for the population proportion of the Reporter poll ranges from 31.6% to 52.4%.

The 95% confidence interval for the population proportion of the Today's News poll ranges from 60.6% to 81.4%.

2

Problem 4

A medical blood pressure supplement experiment is the context. Half of the participants in the study are given a supplement and the other half is given a placebo. The mean difference in blood pressure and standard deviation associated with each group are given. Students interpret the mean difference for each treatment, use a 95% confidence interval to determine a range of values for the population mean of each treatment, and use the results to describe the correlation between taking the supplement and lowering high blood pressure.

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Questions 1 through 5 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 1 and 2

- Would you expect the treatment group taking the placebo to have any change in blood pressure? Why not?
- If you expect little to no change in blood pressure, would you expect any difference between the two sample means?

PROBLEM 4 Pressure Situation: Comparing Continuous Data

A researcher conducted a randomized experiment to see whether there was a link between a new supplement and blood pressure. She collected data from a representative, unbiased sample of 200 people who had high blood pressure. One hundred of the people were randomly selected to take the supplement and the other 100 people were given a placebo. Recall that a placebo is a treatment that is assumed to have no real effect on the characteristic of interest.

The participants' blood pressures were recorded at the beginning and at the end of the 12-week experiment, and the difference (end − beginning) was calculated.

1. For the 100-person treatment that took the placebo, what value would you expect for the difference of sample means at the beginning of the experiment and at the end of the experiment. Explain your reasoning.

   Answers will vary.

   For the 100-person treatment that took the placebo, I would expect that the difference of the two sample means would be close to zero. Taking the placebo would probably not change the sample mean blood pressure very much.

2. For the 100-person treatment that took the supplement, what value would you expect for the difference of sample means at the beginning of the experiment and at the end of the experiment. Explain your reasoning.

   Answers will vary.

   For the 100-person treatment that took the supplement, I would expect that the difference of the two sample means will be a negative number. Taking the supplement will probably lower the sample mean blood pressure.

- If the supplement is effective, would you expect that taking it would result in a lower blood pressure?
- If the supplement is not effective, would you expect that taking it would result in no change in the blood pressure?
- Would a drop in blood pressure be signified by a negative number or a positive number?
- If the blood pressure drops, will the difference of the two sample mean be a negative number?

Guiding Questions for Share Phase, Questions 3 and 4

- Is the mean difference in blood pressure related to blood pressure increasing or decreasing?
- How do you know if the mean blood pressure in the treatment group increased or decreased?
- Does -15 indicate a decrease by an average of 15 points or does it indicate an increase by an average of 15 points?
- Does 1.7 indicate a decrease by an average of 1.7 points or does it indicate an increase by an average of 1.7 points?
- Why would the mean difference associated with the placebo group have any value greater than 0?
- What information is needed to use the formula to determine the margin of error for each treatment group?
- What formula is used to determine the margin of error for each treatment group?
- How is the margin of error used to determine a range of values for the population mean?

Suppose that the mean difference in blood pressure of the group who took the supplement was −15 with a standard deviation of 3.2, and the mean difference in blood pressure of the group who took the placebo was 1.7 with a standard deviation of 0.3.

3. Interpret and explain the meaning of a negative mean difference for the treatment that took the supplement and a positive mean difference for the treatment that took the placebo.

For the group that took the supplement, a negative mean difference means that the blood pressure of people in that group decreased by an average of 15 points.

For the group that took the placebo, a positive mean difference means that the blood pressure of people in that group increased by an average of 1.7 points.

4. Use a 95% confidence interval to determine a range of values for the population mean of each treatment. Explain your work.

For the treatment that took the supplement, the interval from −16.24 to −13.76 represents a 95% confidence interval for the population mean difference in blood pressure.

The margin of error is ±1.24.

$\frac{s}{\sqrt{n}} = \frac{6.2}{\sqrt{100}} = 0.62$

$2(0.62) = 1.24$

For the treatment that took the placebo, the interval from 1.54 to 1.86 represents a 95% confidence interval for the population mean difference in blood pressure.

The margin of error is ±0.16.

$\frac{s}{\sqrt{n}} = \frac{0.8}{\sqrt{100}} = 0.08$

$2(0.08) = 0.16$

2

Guiding Questions for Share Phase, Question 5

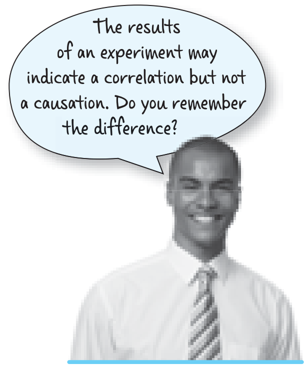
- Do the range of values for the population mean of each difference overlap? What does this imply about the statistical significance?
- Is there a link between taking the supplement and lowering systolic blood pressure?

2

5. What conclusion can you make about whether or not the supplement effectively lowers high blood pressure? Explain your reasoning.

It is likely that there is a correlation between taking the supplement and lowering high blood pressure. I drew this conclusion based on the statistical significance of the results described in Question 3. The range of values for the population mean of each treatment group do not overlap.

I cannot conclude that taking the supplement will result in lower blood pressure. However, I can say that there is statistical evidence of a link between taking the supplement and lowering blood pressure.



Problem 5

Quality control issues in a manufacturing company and the GPA of a random sample of high school students who have a part-time job are the scenarios. Students use the information provided in each situation to determine whether the results are statistically significant. When a given proportion is within the 95% confidence interval, the result is not statistically significant in the first scenario. When confidence intervals to not overlap, the results are statistically significant.

Grouping

Have students complete Questions 1 and 2 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Question 1

- What information is needed to use the formula to determine the margin of error?

PROBLEM 5 Decisions, Decisions . . .

1. A manufacturing company has a policy that states that if significantly more than 2% of computer parts are defective during an 8-hour shift, then the parts from that shift will not be shipped. During an 8 hour shift, 1020 parts were produced and 22 were defective. Should the parts be shipped? Explain your reasoning.

The parts should be shipped.

The interval from 1.2% to 2.8% represents a 95% confidence interval for the population proportion.

The margin of error is approximately ± 0.008

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.02(1-0.02)}{1020}}$$
$$= \sqrt{\frac{0.02(0.98)}{1020}}$$
$$\approx 0.004$$

$2(0.004) = 0.008$

Twenty-two defective parts out of 1020 total parts means that approximately 2.2% of the parts are defective.

$\frac{22}{1020} = 0.022$

Because 2.2% is within the 95% confidence interval, the result is not statistically significant and the parts should be shipped.

- What formula is used to determine the margin of error?
- How is the margin of error used to determine a range of values for the population proportion?
- Twenty-two defective parts out of a total of 1020 parts is what percent of defective parts?
- Is 2.2% within the 95% confidence interval? What does this imply with respect to statistical significance?
- Does the result suggest the parts should be shipped?

Guiding Questions for Share Phase, Question 2

- What information is needed to use the formula to determine the margin of error for each sample of students?
- What formula is used to determine the margin of error for each sample of students?
- How is the margin of error used to determine a range of values for the population mean?
- Do the confidence intervals overlap? What does this imply with respect to statistical significance?
- Does the data suggest a possible link between a student's part-time job status and GPA?

2. The mean grade point average (GPA) of a random sample of 50 High Mark High School students who had a part-time job during the previous grading period is 3.15 with a standard deviation of 0.44. The mean GPA of a random sample of 50 High Mark High School students who did not have a part-time job during the previous grading period is 2.77 with a standard deviation of 0.35. Does that data suggest a possible link between High Mark High School students' part-time job status and their GPA?

The data does suggest a possible link between a student's part-time job status and GPA because population mean confidence intervals for each sample do not overlap.

For the sample of students who have a part-time job, the interval from 3.03 to 3.27 represents a 95% confidence interval for their GPA population mean.

The margin of error is ± 0.12.

$$\frac{s}{\sqrt{n}} = \frac{0.44}{\sqrt{50}} \approx 0.06$$

$2(0.06) = 0.12$

For the sample of students who do not have a part-time job, the interval from 2.67 to 2.87 represents a 95% confidence interval for their GPA population mean.

The margin of error is ± 0.10.

$$\frac{s}{\sqrt{n}} = \frac{0.35}{\sqrt{50}} \approx 0.05$$

$2(0.05) = 0.10$

The population mean interval for the sample of students who have a part-time job, 3.03 to 3.27, does not overlap with population mean interval for the sample of students who do not have a part-time job, 2.67 to 2.87. Therefore, the results of the two samples are statistically significant and suggest that there is a link between part-time job status and GPA.

Problem 6

Student revisit a GPA scenario from the previous lesson to address how statistical significance is related to correlation and causation. Students also consider whether two populations can be considered statistically significant.

2

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Questions 1 through 3 with a partner. Then have students share their responses as a class.

Guiding Questions for Share Phase, Questions 1 and 2

- What is the difference between correlation and causation?
- Is it possible to conclude causation? Explain.
- Can you use data from two different populations to determine statistical significance? Explain.

PROBLEM 6 The End of the Line

Recall the problem from the previous lesson about part-time job status and grade point average (GPA).

The population mean interval for the GPA of High Mark High School students who have a part-time job, 3.03 to 3.27, does not overlap with population mean interval for the GPA of High Mark High School students who do not have a part-time job, 2.67 to 2.87.

1. Carmen shared a conclusion about part-time job status and GPA.

> *Carmen*
>
> *Because the results of the statistical analysis are statistically significant, I can conclude that holding a part-time job will result in a higher GPA.*

Is Carmen's statement correct? Explain why or why not.

Carmen's statement is incorrect. Statistically significant results provide evidence of correlation, not causation.

The interval for the estimate of the population mean for the GPA of neighboring Great Beginnings High School students who do not have a part-time job is 3.18 to 3.39.

2. Is the GPA of students who do not have a part-time job statistically different at High Mark High and Great Beginnings High School? Explain your reasoning.

Depending on the population, the results could be statistically significant.

If the population is High Mark High School students then the results are not statistically significant because the samples are not from the same population. One sample is students from High Mark High School and the other sample is students from Great Beginnings High School.

If the population is neighboring high school students, then the results could be statistically significant as long as the samples are representative and unbiased.

Guiding Questions for Share Phase, Question 3

- Explain how different samples from the same school could represent the same population?
- Explain how different samples from the same school could represent different populations.

3. The estimate for the population mean for the math GPA of Great Beginnings High School students using a sample of the math club is 3.27 to 3.54. The estimate for the population mean for the math GPA of Great Beginnings High School students using a sample of the government club is 3.11 to 3.40.

> Max
>
> *The results of the statistical analysis are not statistically significant because the population mean intervals for math GPA overlap.*

Is Max's statement correct? Explain why or why not.

Max's statement is incorrect. In order to have statistically significant results the samples must be randomly selected from the same population. It is likely that this scenario contains sampling bias. Even though both samples are from the same school, one sample consists of students exclusively from the math club and the other sample consists of students exclusively from the government club. A likely source of sampling bias is that student in the math club will probably have a higher math GPA than students in the government club.

2

Online Time Study, Part IV

The student online survey continues to be the topic. Two samples of data were collected and the sample mean and standard deviation associated with each sample are given. Students use a 95% confidence interval to determine whether the estimates of the population means are statistically significant.

2

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students complete Question 1 with a partner. Then have students share their responses as a class.

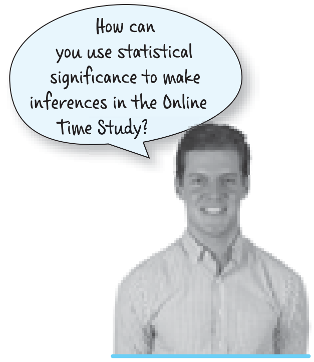
Guiding Questions for Share Phase, Question 1

- Which term is associated with collecting all of the possible samples of equal sizes from a population?
- Which term is associated with a data set that is collected from a population?
- Which term(s) are associated with measures of center for a sample?
- Which term is associated with only categorical data?
- Which term is associated with only continuous data?
- Which term is associated with the measure of spread used to create a confidence interval?

Online Time Study, Part IV

To analyze data from a sample survey, observational study, or experiment, you can use statistical significance to make inferences about populations.

Recall the study you have been planning about the amount of time students in your school are online each day.



Suppose two samples of data were collected. One sample of 40 students in your school has a sample mean of 2.3 hours and a standard deviation of 0.7 hour. Another sample of 40 students in your school has a sample mean of 3.7 hours and a standard deviation of 1.1 hours.

1. Use a 95% confidence interval to determine whether the estimate of the population means using each sample is statistically significant. Explain your work.

The results are statistically significant because there is no overlap between the two 95% confidence intervals of the population means.

The interval from 2.08 to 2.52 hours represents a 95% confidence interval for the population mean of the sample with a mean of 2.3 hours.

The margin of error is approximately ±0.22.

$$\frac{s}{\sqrt{n}} = \frac{0.7}{\sqrt{40}} \approx 0.11$$

$2(0.11) = 0.22$

The interval from 3.36 to 4.04 hours represents a 95% confidence interval for the population mean of the sample with a mean of 3.7 hours.

The margin of error is approximately ±0.34.

$$\frac{s}{\sqrt{n}} = \frac{1.1}{\sqrt{40}} \approx 0.17$$

$2(0.17) = 0.34$

Be prepared to share your results and methods.

- Which term is associated with a range of numbers estimating a population mean or population proportion?
- The margin of error is usually associated with which percent confidence interval?

Check for Students' Understanding

1. A sample of test scores from Ms. Baker's first period shows that 32 out of 36 students passed and a sample of test scores from Ms. Baker's second period shows that 25 out of 30 students passed. Determine whether the estimate of the population proportion using the sample space is statistically significant using a 95% confidence interval.

   The difference between the population proportion estimates of the passing test scores for both classes is not statistically significant because the two confidence intervals overlap.

   The 95% confidence interval for the population proportion of passing test scores for the first period ranges from 79% to 99.0%. The margin of error is approximately ±0.10.

   The 95% confidence interval for the population proportion of passing test scores for the second period ranges from 68.48% to 97.52%. The margin of error is approximately ±0.1452.

$$\sqrt{\frac{0.89(1-0.89)}{36}}$$

$$\sqrt{\frac{0.89(1-0.89)}{36}} = \sqrt{\frac{0.89(0.11)}{36}} \approx 0.05$$

$89 - 2(5) = 79$

$89 + 2(5) = 99$

$$\sqrt{\frac{0.83(1-0.83)}{30}}$$

$$\sqrt{\frac{0.83(1-0.83)}{30}} = \sqrt{\frac{0.83(0.17)}{30}} \approx 0.0726$$

$83 - 2(7.26) = 68.48$

$83 + 2(7.26) = 97.52$

2

2. A sample of 30 dog owners responded to a survey about the number of hours they use doggie daycare per week. The sample mean was 32 hours and a standard deviation of 4 hours. Another sample of 30 dog owners that responded to the same survey has a sample mean of 40 hours with a standard deviation of 8 hours. Use a 95% confidence interval to determine whether the estimate of the population means using each sample is statistically significant.

The results are statistically significant because there is no overlap between the two 95% confidence intervals of the population means.

The interval from 30.54 to 33.46 hours represents a 95% confidence interval for the population mean of the sample with a mean of 4 hours.
The margin of error is approximately ±1.46

$$\frac{s}{\sqrt{n}}$$

$$\frac{4}{\sqrt{30}} \approx 0.73$$

$$2(0.73) = 1.46$$

The interval from 37.08 to 42.92 hours represents a 95% confidence interval for the population mean of the sample with a mean of 8 hours.
The margin of error is approximately ±2.92.

$$\frac{s}{\sqrt{n}}$$

$$\frac{8}{\sqrt{30}} \approx 1.46$$

$$2(1.46) = 2.92$$

2

2.5

DIY

Designing a Study and Analyzing the Results

LEARNING GOALS

In this lesson, you will:

- Analyze the validity of conclusions based on statistical analysis of data.
- Design a sample survey, observational study, or experiment to answer a question.
- Conduct a sample survey, observational study, or experiment to collect data.
- Summarize the data of your sample survey, observational study, or experiment.
- Analyze the data of your sample survey, observational study, or experiment.
- Summarize the results and justify conclusions of your sample survey, observational study, or experiment.

ESSENTIAL IDEAS

- A real world situation is used to design and conduct a sample survey, observational study, or experiment.
- Statistical analysis of sample data can be used to interpret the data and draw conclusions.

COMMON CORE STATE STANDARDS FOR MATHEMATICS

S-IC Making Inferences and Justifying Conclusions

Understand and evaluate random processes underlying statistical experiments

1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
2. Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation.

Make inferences and justify conclusions from sample surveys, experiments, and observational studies

3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
4. Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.
5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.
6. Evaluate reports based on data.

Overview

In this final lesson, students are given the guidelines developed throughout the chapter for designing and conducting a sample survey, observational study, or an experiment for a characteristic of interest of their choice.

Warm Up

Which method best answers each question, a sample survey, an observational study, or an experiment? Explain your reasoning.

1. Are senior citizens more likely to purchase dark color cars?

   An observational study or sample survey could determine whether senior citizens more likely to purchase dark color cars.

2. Are pet owners healthier than people that do not own a pet?

   An experiment could determine whether pet owners are healthier than people who do not own a pet.

3. Do more people drive SUVs or sedans?

   An observational study could determine whether more people drive SUVs or sedans.

4. Does a particular herb increase memory?

   An experiment could determine whether a particular herb increases memory.

5. Which population uses more electronic reading devices, adults or teenagers?

   A sample survey could determine whether adults or teenagers use more electronic reading devices.

2

2.5

DIY

Designing a Study and Analyzing the Results

2

LEARNING GOALS

In this lesson, you will:

- Analyze the validity of conclusions based on statistical analysis of data.
- Design a sample survey, observational study, or experiment to answer a question.
- Conduct a sample survey, observational study, or experiment to collect data.
- Summarize the data of your sample survey, observational study, or experiment.
- Analyze the data of your sample survey, observational study, or experiment.
- Summarize the results and justify conclusions of your sample survey, observational study, or experiment.

DIY stands for "do it yourself." So, why not? Try to write an interesting opener yourself for this lesson. Use these hints to help you get started:

- Make your opener related to something about the lesson or the whole chapter.
- Write about something you think other students would be interested in reading.
- Be creative!

Share your opener with your classmates. Which one did you like best?

Problem 1

Students use the guidelines provided to design and conduct a sample survey, observational study, or experiment, summarize and analyze the data, and draw conclusions.

2

Grouping

- Ask a student to read the information. Discuss as a class.
- Have students choose a characteristic of interest and follow the guidelines to complete the project.

PROBLEM 1 Do It Yourself!

Use the following guidelines to design and conduct a sample survey, observational study, or experiment, summarize and analyze the data, and draw conclusions. You can use this page as a checklist while planning and conducting your study.

I. Design a sample survey, observational study, or experiment.	
• Select a characteristic of interest to learn about from a sample survey, observational study, or experiment.	
• Select a question that can be answered by collecting quantitative data.	
• Identify the population.	
• Identify the characteristic being studied.	
• Describe the method for choosing a random sample.	
• Address potential sources of bias.	
II. Conduct the sample survey, observational study, or experiment.	
• Use the sampling method to collect data for your sample survey, observational study, or experiment.	
III. Summarize the data of the sample survey, observational study, or experiment.	
• Calculate measures of center.	
• Calculate measures of spread.	
• Select the most appropriate method(s) to display the data (dot plot, histogram, stem-and-leaf plot, box-and-whisker plot, normal curve).	
• Describe the characteristics of the graphical display.	
IV. Analyze the data of the sample survey, observational study, or experiment.	
• Use confidence intervals to determine a range of values for the population mean(s) or proportion(s).	
• Using statistical significance to make inferences about populations.	
V. Draw conclusions based on the results of the sample survey, observational study, or experiment.	
• Write a conclusion that answers the question of interest of your sample survey, observational study, or experiment. Use the data and data analysis to justify your conclusion.	

Be prepared to share your results and methods.

Check for Students' Understanding

1. Is it easier to design and conduct a sample survey, observation study or experiment? Why?

Answers will vary.

2

2. When designing a sample survey, observation study or experiment, what is the difference between selecting a characteristic of interest and selecting a question that can be answered by collecting data?

Answers will vary.

3. What should be considered when identifying the population for a sample survey, observation study or experiment?

Answers will vary.

4. What are important elements associated with designing a sample survey, observation study or experiment?

- Select a characteristic of interest
- Select a question that can be answered by collecting quantitative data
- Identify the population
- Identify the characteristic being studied
- Describe the method for choosing a random sample
- Address potential sources of bias

5. What are important elements associated with summarizing the data of a sample survey, observation study or experiment?

- Calculate the measure of center
- Calculate the measures of spread
- Select the most appropriate method(s) to display the data
- Describe the characteristics of the graphical display

2

Chapter 2 Summary

KEY TERMS

- characteristic of interest (2.1)
- sample survey (2.1)
- random sample (2.1)
- biased sample (2.1)
- observational study (2.1)
- experiment (2.1)
- treatment (2.1)
- experimental unit (2.1)
- confounding (2.1)
- convenience sample (2.2)
- subjective sample (2.2)
- volunteer sample (2.2)
- simple random sample (2.2)
- stratified random sample (2.2)
- cluster sample (2.2)
- cluster (2.2)
- systematic sample (2.2)
- parameter (2.2)
- statistic (2.2)
- population proportion (2.3)
- sample proportion (2.3)
- sampling distribution (2.3)
- confidence interval (2.3)
- statistically significant (2.4)

2.1 Identifying Characteristics of Sample Surveys, Observational Studies, and Experiments

The characteristic of interest is the specific question to be answered or the specific information to be gathered for sample surveys, observational studies, and experiments. The entire set of items from which data can be selected is the population. A subset of the population that is selected is a sample.

Example

Fifty-five deer are randomly selected from a park in the township. They are anesthetized, weighed, and then released back into the park.

The population is all of the deer in the park. The sample is the 55 deer selected. The characteristic of interest is the mean weight of the deer.

2

2.1 Differentiating Between Sample Surveys, Observational Studies, and Experiments

A sample survey poses a question of interest to a sample of the targeted population. An observational study gathers data about a characteristic of the population without trying to influence the data. An experiment gathers data on the effect of one or more treatments on the characteristic of interest.

Example

A study states that approximately 78% of planes arrived on time during a 3 hour period at an airport.

This is an observational study since the study only gathered data about the number of planes that arrived on time and did not try to influence the data.

2.2 Using a Variety of Sampling Methods to Collect Data

Sampling methods could include convenience sampling, volunteer sampling, simple random sampling, stratified random sampling, cluster sampling, and systematic sampling.

Example

The data set below shows the number of late student arrivals at four elementary schools each week for five weeks.

Number of Late Arrivals				
Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**
49	37	45	44	43
47	41	45	46	48
39	43	38	44	42
43	47	39	39	42
52	55	50	54	55

You can create a stratified random sample with 5 data values to describe the number of late arrivals by randomly choosing one school from each of the 5 weeks and recording the number of late arrivals: {39, 37, 50, 46, 42}.

2.2

Identifying Factors of Sampling Methods that could Contribute to Gathering Biased Data

Some sampling methods introduces bias, which reduces the likelihood of a representative, unbiased sample.

Example

A cereal company conducts taste tests for a new cereal on a random sample of its employees.

There is bias in this study because the taste test is only conducted on the company's employees. It is possible that the employees will prefer the cereal of the company that employs them for other reasons than taste.

2.2

Exploring, Identifying, and Interpreting the Role of Randomization in Sampling

You can use random sampling by using a random digit table or a graphing calculator to create unbiased samples.

Example

For the data set, you can use a calculator to generate four random numbers between 1 and 10. Then you can use the numbers generated to create a random sample of four from the data set.

The 25-meter freestyle times, in seconds, of ten young swimmers are shown.

Swimmer	1	2	3	4	5	6	7	8	9	10
Time	21.2	19.3	18.7	20.6	20.5	18.4	22.9	23.5	18.2	17.9

Possible random numbers: 19.3, 18.7, 22.9, 17.9.

2

2.3 Recognizing that Data from Samples are Used to Estimate Population Proportions and Population Means

Data from samples are used to calculate confidence intervals that estimate population proportions and population means.

Example

A sample of 250 women responded to a survey about the amount of money they spend on cosmetics each month. The sample mean was $45.50 and the sample standard deviation was $10.75.

The interval from $44.14 to $46.86 represents a 95% confidence interval for the population mean.

$$\frac{s}{\sqrt{n}} = \frac{10.75}{\sqrt{250}} \approx 0.68$$

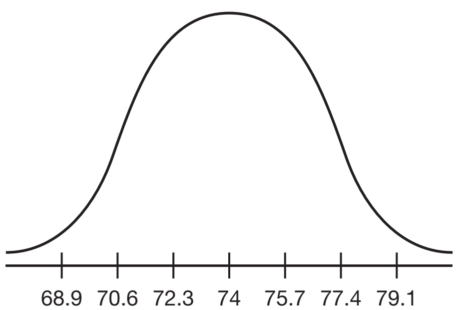
2.4 Using Sample Proportions to Determine Whether Differences in Population Proportions are Statistically Significant

To determine the sample proportions that would be statistically significant, use the normal curve and label it based on the standard deviation from the sample.

Example

Use the sample proportion and standard deviation of the sampling distribution to label the horizontal axis of the normal curve. Then, determine what sample proportions would be statistically significant.

A sample proportion of families that own dogs is 74%, and the standard deviation is 0.017.



Sample proportion values less than 70.6% and greater than 77.4% are statistically significant because those values are outside of the 95% confidence interval.

2.4 Using Sample Means to Determine Whether Differences in Population Means are Statistically Significant

Use the sample mean and standard deviation to determine the margin of error for the confidence interval. The margin of error is 2 times the standard deviation.

Example

Use a 95% confidence interval to determine a range of values for the population mean. Explain your work.

A sample of 80 doctors took a stress test. The sample mean was 44.5 and the sample standard deviation was 14.8.

The interval from 41.2 to 44.5 represents a 95% confidence interval for the population mean.

The margin of error is approximately ± 3.30.

$$\frac{s}{\sqrt{n}} = \frac{14.8}{\sqrt{80}} \approx 1.65$$

$2(1.65) = 3.30$

2.5 Conducting a Sample Survey, Observational Study, or Experiment to Answer a Question

You can determine what type of sample technique would be most appropriate to answer a question for a sample survey, observational study, or experiment.

Example

Suppose you want to estimate the number of senior citizens in a town that are on public assistance.

You can assign all the senior citizens in the town an ID number and use a computer to randomly generate a sample of senior citizens. This technique provides a random sample of the population of the senior citizens in the town, and random sampling is typically representative of a population.

2