Statistics

6

Unit Overview

In this unit, you will begin your study of statistics. You will learn how to select a random sample from a population and how to use data from the random sample to learn about the population. You will also use sample data to compare two populations.

Key Terms

As you study this unit, add these and other terms to your math notebook. Include in your notes your prior knowledge of each word, as well as your experiences in using the word in different mathematical examples. If needed, ask for help in pronouncing new words and add information on pronunciation to your math notebook. It is important that you learn new terms and use them correctly in your class discussions and in your problem solutions.

Academic Vocabulary

- population
- sample
- census

Math Terms

- sampling
- sampling variability
- sample mean
- sample statistic
- random sample
- population mean

ESSENTIAL QUESTIONS

- Why is it important to select at random when choosing a sample from a population?
- How can sample data be used to learn about a population?
- How can sample data be used to compare two populations?

EMBEDDED ASSESSMENTS

These assessments, following activities 25 and 26, will give you an opportunity to demonstrate your understanding of statistics and your ability to use sample data to draw conclusions.

Getting Ready

Write your answers on notebook paper. Show your work.

Use the following information to answer Items 1–6.

Each of the 10 students in Mr. Finn's honors math class was asked how many hours per week he or she spent studying in a typical school week. The responses are given here:

6	9	12	4	7
5	6	10	3	8

1. Calculate the mean and median of this data set.
2. Calculate the first and third quartiles for this data set.
3. Calculate the IQR (interquartile range) for this data set.
4. Draw a box plot for this data set.
5. Draw a dot plot for this data set.
6. Calculate the mean absolute deviation (MAD) for this data set. You can use the table below to help organize your work.

Data value	Distance from the mean
6	
9	
12	
4	
7	
5	
6	
10	
3	
8	

Lesson 24-1 Class Data

My Notes

Learning Targets:

- Determine from what population data has been collected.
- Determine if a data collection is a census.
- Display and analyze data in circle graphs, bar charts, and dot plots.

SUGGESTED LEARNING STRATEGIES: Think-Pair-Share, Summarizing Paraphrasing, Interactive Word Wall

In sixth grade, you collected data about your class and used graphs and statistics (such as the mean and the median) to summarize those data. This was a way to learn about your class, and because it was possible to get data from every student in the class, you could make clear statements about characteristics of your class.

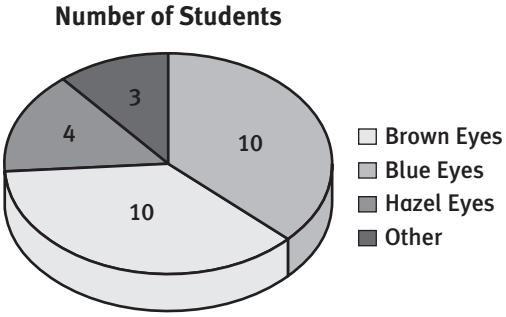
The group that you are interested in learning about is called a ***population***. When you are able to collect data from every individual in the group you are interested in, you have what is called a ***census*** of that group.

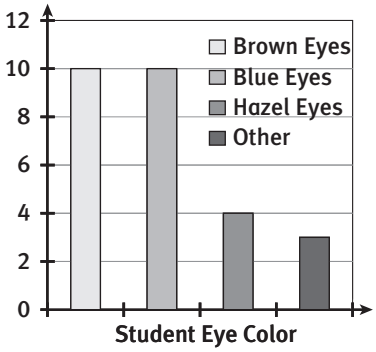
Suppose that you wanted to gather some information about the members of your current class, such as eye color.

1. What is the population in this example?

2. **Reason abstractly.** Explain why choosing five students in your class to survey eye color would not be a census.

Assume that you survey your class and get the following results for eye color: brown eyes, 10; blue eyes, 10; hazel eyes, 4; and other, 3. To analyze this data, you might display it in a circle graph or a bar chart.





3. Look at the circle graph. Calculate the percentage of students with each eye color.

4. Look at the data in the circle graph. Explain whether any of the data is equivalent and why or why not.

ACADEMIC VOCABULARY

A ***population*** is the whole group that you are interested in learning about.

A ***census*** is a study where data is collected from everyone in the whole population. The US census takes place every 10 years.

MATH TIP

Equivalent means to have the same value. Fractions, expressions, and ratios are equivalent when they have the same values.

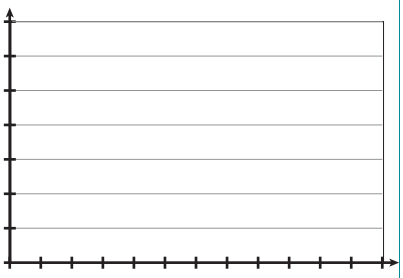
My Notes

MATH TIP

Displaying data visually helps you make **part-to-whole** or **part-to part** comparisons. A part-to-whole comparison is made by comparing one data point to the entire set of data; e.g., by calculating a number as a percentage of the whole. A part-to-part comparison is made by comparing two or more data points.

TECHNOLOGY TIP

If you have access to a computer and programs that create charts, use it to enter your values for birth months and create the pie graph and bar chart.



5. Explain a *part-to-whole* comparison of students with brown eyes to the entire class.

6. Make a *part-to-part* comparison of students with brown eyes and hazel eyes.

7. Now, collect data on your classmates' birth months. Count the number of students in your class born in each month and create a table. Use the My Notes space or notebook paper to record the data for each month.

8. Use the data you collected in Item 6 to determine the number of students born in each month. Create a circle graph, dividing the circle into sectors representing each month. Label each sector with the name of the month and the number of students born in that month.

9. Using the data in your circle graph, make part-to-whole and part-to-part comparisons.

10. Identify whether any data in your circle graph is equivalent and explain why or why not.

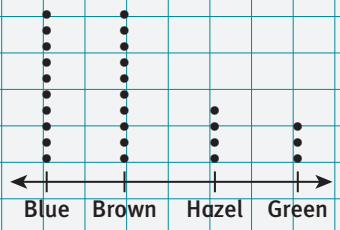
11. Suppose that another class in your school had birthdays as follows: January, 3; February, 4; March, 6; April, 3; May, 2; June, 1; July, 2; August, 0; September, 4; October 3; November, 1; and December, 1. Use this data to create a bar chart.

12. Make part-to-part and part-to-whole comparisons of the class data.

13. Look at the data in your bar chart. Is any of the data equivalent? Explain why or why not.

My Notes

You might also display data in a dot plot. The dot plot on the right shows the eye colors of a different group of students.



14. Survey your class and list the eye colors of your classmates. Record the number of students with each eye color. Create a dot plot to show this data. Use the My Notes space. Then make part-to-part and part-to-whole comparisons for your data.

15. Look at your class data. Is any of the data equivalent? Explain why or why not.

16. Think of the three methods of displaying data that you have studied. Which method do you think is the most useful, and why? Would your answers change as the amount of your data increases? Explain why.

Check Your Understanding

17. Use the following set of data representing the shoe sizes of shoppers who were surveyed to create a bar chart.

$3\frac{1}{2}$	4	$5\frac{1}{2}$	$6\frac{1}{2}$	$5\frac{1}{2}$	5	$4\frac{1}{2}$	$3\frac{1}{2}$
$4\frac{1}{2}$	7	5	4	$3\frac{1}{2}$	7	8	$6\frac{1}{2}$

a. What is the most common shoe size?

b. What equivalents are in this set of data?

c. Make part-to-whole and part-to-part comparisons for this bar chart.

18. Survey your class and list their shoe sizes. Make a circle graph and a dot plot showing the sizes.

a. Write a brief summary of what these charts tell you about the shoe sizes of the students in your class.

b. Include part-to-whole and part-to-part comparisons in your summary. Describe any equivalent data.

LESSON 24-1 PRACTICE

19. **Make sense of problems.** Matt decides to collect data from students on the tennis team about how many texts they send in a day. He asks all the players who come to Wednesday's practice to check their phones and records their responses. Is this a census? Explain.

20. Describe part-to-whole and part-to-part comparisons of data and give examples of each.

21. Compare the circle graph, bar chart, and dot plots as methods of displaying data. Explain which you think best displays data and why.

22. Describe a population for which you can perform a census, and explain why the census clearly represents that population.

My Notes

Learning Targets:

- Understand that the way a sample is selected is important.
- Understand that random sampling is a fair method for selecting a sample.
- Use the random-number digit table to select a random sample.

SUGGESTED LEARNING STRATEGIES: Think-Pair-Share, Create Representations, Look for a Pattern, Summarizing, Paraphrasing, Interactive Word Wall

ACADEMIC VOCABULARY

A ***sample*** is a small part that is representative of a whole. In statistics, the sample is the part of the population that we collect data from.

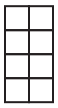
We usually study a sample in order to learn about the population.

Sometimes, instead of collecting data from everyone in the population, we decide to study just a part of the population. For example, instead of collecting data from every student at the school, you might decide to collect data from just 50 students at the school. When we study just a part of the population, the part of the population that we decide to study is called a ***sample.*** The process of choosing a sample from a population is called ***sampling***.

You will start your study of sampling by considering a small population. The population consists of 100 middle school students who signed up for a summer reading program. Suppose that we wanted to learn about the average number of books read by these students over the summer.

The number of books read by each of the 100 students is represented graphically using rectangles on the Reading Program Data page found at the end of this activity. But don't turn to that page yet! First you need to understand how the data are represented.

Each student is represented by a rectangle that is divided up into small squares. For example, one student is represented by this rectangle:



Each small square represents one book, so this rectangle tells us that this student read eight books.

1. What does the following rectangle represent?

My Notes

2. What does the following rectangle represent?

Now, carefully tear out the Reading Program Data page and place it face down on your desk. When your teacher tells you to turn it over, turn it over and look at the population shown on the page. Your teacher will give you a short time to look at the page and think about what the average number of books read by these students is. When you teacher tells you to, turn the page back over and record your guess of the average number of books read for this population in the space provided below.

I think that the average number of books read by the 100 students in the summer reading program is about _____.

Your teacher will provide a poster. Add a dot to the poster to show your guess for the average number of books read. When the dot plot is complete, it will display the guesses for all of the students in the class. Use the class dot plot to answer the following items.

3. Considering all of the guesses from your class, would you say that the guesses are similar or that the guesses showed a lot of variability?

4. After looking at the dot plot, if you had to come up with a single number that would summarize the guesses from the whole class, what number would you use? Is this number close to your original guess?

Rather than guessing at just the population average, you could select a sample from this population and use the data values for the students in the sample to learn about the population. Of course, you would want the students in the sample to be representative of the population.

My Notes

Turn over the Reading Program Data page and circle five rectangles that you think are representative of the population. Each of these rectangles represents one of the students in the summer reading program.

5. For each of the five students you selected, count the number of books that each student read (the number of squares that make up the rectangle) and record those data values here:

6. Calculate the mean (average) of the five data values for the students in your sample. This is called the **sample mean**.

MATH TERMS

The **sample mean** is the average of the data values for a sample.

The **population mean** is the average of the data values for the whole population.

If the sample is selected in a reasonable way, the sample mean can be used as an estimate of the population mean.

Your teacher will provide a poster. Add a dot to the poster to show the mean of the five data values for your sample. When the dot plot is complete, it will display the sample averages for all of the students in the class. Use this dot plot and the earlier dot plot of the class guesses to answer the following items.

7. Considering all of the sample means from your class, would you say that the means were similar or the means showed a lot of variability?

8. **Use appropriate tools strategically.** In what ways are the two dot plots similar? In what ways are they different? Does one of the two dot plots show less variability than the other? If so, which one?

My Notes

9. **Reason quantitatively.** If you were to do a census of the population, you would find that that actual population mean is 7.5. Based on the two dot plots, does it surprise you that the population mean is 7.5? Explain your thinking.

10. If you really wanted to learn about the average number of books read by students in the summer reading program, do you think that just taking a quick overall look at the population and then making a guess is a good strategy? Did it work well when the students in your class did this?

11. If you wanted to learn about the average number of books read by students in the summer reading program by studying a sample of five students from this population, would picking five students that you think are representative of the population and then calculating the sample mean be a good strategy? Did it work well when the students in your class did this?

12. In general, people are not very good at picking a representative sample! Can you think of another way that you could select five students from this population that you think would be more likely to produce a sample that was representative of the population?

So far, you have seen that first impression estimates and other methods of selecting a sample from a population don't work very well! In this example, even the method of picking students who you thought were representative did not work well, because most people tend to overlook the many small rectangles in the population that correspond to students who read only one or two books. People are just not very good at selecting a sample that is representative of the population—even if they are trying to do a good job!

My Notes

MATH TERMS

A sample that is formed by selecting individuals from the population at random is called a **random sample.**

Instead of letting your own personal ideas or biases influence the way a sample is selected, we need a fair method for selecting a sample. We do this by selecting from the population at random. When individuals in a population are selected at random, the result is a ***random sample***.

When individuals are selected at random, every individual in the population has the same chance of being included in the sample. Here are three ways that you could select a random sample of five students from the population of the 100 students in the summer reading program:

Method 1: Write each of the 100 student names on a slip of paper, place all of the slips of paper in a box, mix them well, and then select five slips of paper. The names written on these slips would be the five students you would include in the sample.

Method 2: Number the students in the population from 1 to 100, giving each student a different number. Write each of the numbers from 1 to 100 on a slip of paper, place all of the slips of paper in a box, mix them well, and then select five slips of paper. The students corresponding to the numbers on these slips would be the five students you would include in the sample.

Method 3: Number the students in the population from 00 to 99 (using the two-digit numbers 00, 01, 02, and so on), giving each student a different two-digit number. Using a table of random digits, you could get five two-digit numbers. The students corresponding to these numbers would be the five students you would include in the sample.

13. Which of the three methods described would take the most time to implement? Which would take the least time? Why do you think this?

My Notes

14. Number the rectangles on the Reading Program Data page, starting with the number 00 and ending with the number 99. Make sure that every rectangle gets a number. There are 100 rectangles (corresponding to students) in this population, so you will need to use all of the numbers from 00 to 99.

A table of random digits appears at the end of this activity. This is the same table you used in Activity 23, but it is reproduced here just in case you misplaced that page. Tear out this page.

Toss a paper clip onto the page of random digits. Mark the digit on the page that is closest to the center of the larger loop of the paper clip. Starting with that digit, write 10 digits below, dividing them up into two-digit numbers. (If by chance you get a duplicate two-digit number, just ignore that one and use the next two digits from the table.

My five two-digit numbers:

15. Find the rectangles with these five numbers. For each of these five students, count the number of books that the student read (the number of squares that make up the rectangle) and record those data values here:

16. Calculate the mean of the five data values for the students selected.

17. Is the mean you computed in Item 16 a sample mean or a population mean? Share your response with your group. Incorporate your understanding into your group discussion to confirm your knowledge and use of key mathematical language.

My Notes

Add your sample mean to a class dot plot on a poster that your teacher will provide. This dot plot will display the sample means for the different random samples selected by the students in your class. Use this dot plot and the earlier dot plot of the class means of self-selected samples to answer the following items.

18. **Reason abstractly.** Considering all of the means from the class random samples, would you say that the means were similar to each other or that the means showed a lot of variability?

19. How does the dot plot of the sample means from random samples compare to the class dot plot of means from the self-selected samples? In what ways are the two dot plots similar? In what ways are they different? Does one of the two dot plots show less variability than the other? If so, which one?

20. If you wanted to learn about the average number of books read by students in the summer reading program by studying a sample of five students from this population, is picking five students at random from the population and then calculating the sample mean a good strategy? Did it work well when the students in your class did this?

21. Suppose that instead of a selecting a random sample of size 5 (five students in the sample), each student in your class had selected a random sample of size 10. If the sample means for these samples were used to make a dot plot, how do you think that this dot plot would be different from the dot plot for samples of size 5? Do you think it would be centered in about the same place? Do you think it would show more or less variability?

My Notes

22. Why is selecting a random sample a good way to learn about a population?

Check Your Understanding

Every year, a large school district holds a spelling bee. This year, 400 students have entered the spelling bee. You want to learn about the mean age of students who enter the spelling bee.

23. What is the population of interest to you?

24. If you wanted to do a census of this population, from how many students would you need to collect data?

25. Describe how you could select a random sample of 20 students from this population.

26. Sarah thinks it will be OK to collect data on just the age from the first 20 students eliminated from the spelling bee and then use the mean age of these students to estimate the mean age of all the students entered in the spelling bee. Explain to Sarah why this is not a good idea.

LESSON 24-2 PRACTICE

This Saturday, the Walk-a-thon to raise money for cancer research will be held. Livia wants to learn about the average amount raised by registered participants in the age range of 11 to 15. For the event, 125 participants have registered for this age group.

27. What is the population of interest to Livia?

28. If Livia wanted to do a census of this population, from how many people would Livia need to collect data?

29. **Make sense of problems.** If all the registered walkers in the age range of 11 to 15 attend, describe how you would select a random sample of 12 from this population.

30. Livia decides to collect data from the last 15 walkers to check in at the walkathon. Explain why this may not be a good idea.

31. After the event, Livia adds up all the donations for the registered participants in the age range 11 to 15 and divides by 125 to find the average amount raised. Is this a census? Explain.

ACTIVITY 24 PRACTICE

1. Explain what it means to say that selecting a random sample from a population is a fair way to select a sample.
2. Write a few sentences describing what you learned in Activity 24 about selecting a sample.
3. You are interested in learning about how many text messages are sent by ninth-grade students at Reseda High School. You select 30 students at random from the ninth graders at Reseda High and ask each one how many text messages they send in a typical day.
   a. Are these 30 students the population or a sample?
   b. Before your started, a friend suggested that you collect data from the 30 ninth graders in Mr. Rossman's physics class. Why is it better to select ninth-grade students at random?

Use the information below to answer Items 4 and 5.

Forty students belong to the robotics club at Morro Middle School. The ages of these students are shown in the table below.

Student	Age	Student	Age	Student	Age	Student	Age
01	12	11	14	21	14	31	12
02	12	12	13	22	14	32	12
03	11	13	13	23	13	33	14
04	12	14	13	24	12	34	13
05	12	15	14	25	13	35	14
06	13	16	12	26	12	36	13
07	14	17	13	27	12	37	13
08	12	18	13	28	12	38	12
09	15	19	13	29	11	39	14
10	12	20	13	30	13	40	13

Suppose that you did not know the ages of all of these students and that you planned to select a sample of six students and use the sample mean to estimate the mean age of the students in the robotics club.

4. Explain why it would not be a good idea to just pick the first six students on the list of students in the table above.
5. Below is a sequence of random digits, arranged in groups of two. Use this list to select a random sample of six students from the students in the robotics club. Ignore any two-digit numbers in the list that are greater than 40. If you come across a two-digit number you have already used, skip it and go on to the next two-digit number.

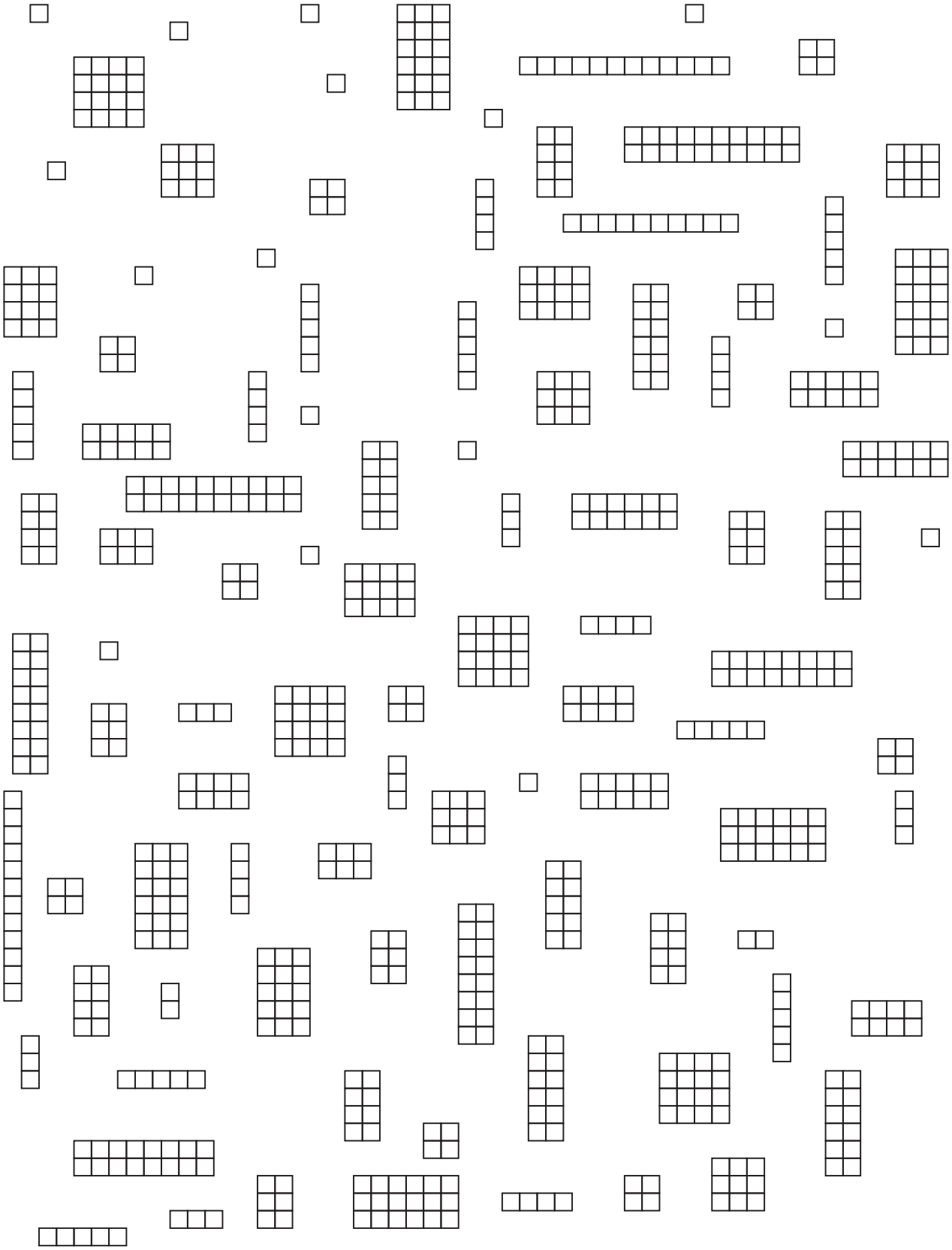
71	92	07	53	88	80	28	58	89	29
92	95	96	67	72	34	45	21	38	53
86	30	30	93	07	30	92	88	70	53

   a. Which students are in the sample?
   b. What are the ages of the students in the sample?
   c. What is the mean of these six ages?
   d. Is the mean you calculated in part c a population mean or a sample mean?

MATHEMATICAL PRACTICES
Use Appropriate Tools Strategically

6. Why is it better to select a random sample than to just decide who will be in the sample by picking people that you think will be representative of the population?

Reading Program Data
Number of Books Read for 100 Students in Summer Reading Program



This page is intentionally blank.

Random Digit Table

Row 1	7 1 1 3 6	4 6 8 8 9	1 9 0 5 4	1 6 7 1 5	1 2 2 6 4
Row 2	6 8 5 7 5	8 8 4 4 9	0 5 6 1 7	5 4 5 7 3	6 7 9 0 8
Row 3	9 3 0 9 8	7 0 5 8 7	5 4 2 3 2	0 1 8 8 1	6 6 8 8 0
Row 4	9 6 7 7 9	6 9 3 1 3	1 9 1 6 8	3 9 0 3 6	8 0 9 0 7
Row 5	1 7 5 4 0	5 9 9 8 2	1 9 7 8 3	9 6 6 0 2	8 6 7 2 4
Row 6	5 5 0 9 4	2 0 6 7 3	6 8 3 3 6	4 1 4 2 4	1 8 4 5 7
Row 7	1 4 5 3 6	1 1 4 1 4	3 9 6 1 1	6 3 0 4 4	6 4 8 8 5
Row 8	1 8 3 8 4	8 0 5 0 4	9 3 8 7 7	7 2 6 3 8	5 4 6 5 2
Row 9	0 0 7 6 6	8 1 1 9 5	6 8 4 1 4	1 3 2 4 2	1 7 3 5 9
Row 10	8 9 9 4 9	2 5 9 0 4	3 8 3 9 1	1 8 0 8 8	2 0 2 9 5
Row 11	8 1 4 2 9	7 4 3 2 7	9 4 5 1 9	9 7 0 7 4	5 7 2 4 3
Row 12	4 4 8 3 1	1 8 5 8 5	3 1 9 5 0	4 4 8 7 9	1 2 9 6 7
Row 13	1 0 1 8 7	0 3 1 9 9	5 5 2 7 5	4 5 3 6 2	6 8 8 8 2
Row 14	2 1 2 5 1	8 1 4 3 0	1 0 8 8 3	2 5 5 3 8	3 5 4 9 1
Row 15	6 4 3 1 0	6 4 3 2 2	7 9 9 7 9	4 1 3 7 6	6 3 2 3 7
Row 16	0 1 0 5 3	0 0 2 1 3	1 1 1 9 8	1 8 4 0 7	7 8 9 7 0
Row 17	8 8 3 6 4	8 5 9 4 5	9 9 6 1 5	4 5 8 4 6	1 2 7 0 6
Row 18	1 0 7 7 7	9 5 8 3 5	8 7 3 5 3	6 4 6 4 9	1 3 2 6 9
Row 19	4 6 4 7 6	7 8 5 7 9	3 2 1 0 3	1 2 8 5 1	8 0 2 6 8
Row 20	9 4 0 3 5	9 4 9 9 6	5 5 9 2 2	9 5 1 1 5	1 2 1 4 7
Row 21	8 8 4 2 7	5 1 9 1 1	1 4 0 9 8	3 1 6 4 0	6 3 0 0 3
Row 22	7 2 6 3 1	3 2 4 2 3	2 2 6 9 9	9 2 1 4 5	8 3 6 7 6
Row 23	7 6 3 8 4	6 5 0 7 8	7 4 1 0 9	3 6 9 5 3	6 1 0 9 6
Row 24	9 9 3 2 4	3 0 4 8 8	0 4 7 1 4	5 3 9 0 4	2 4 9 1 8
Row 25	9 7 9 9 2	6 4 8 4 0	6 0 6 1 5	6 9 9 2 2	3 7 4 0 3
Row 26	7 7 1 0 6	6 8 6 7 2	2 6 9 6 8	3 9 9 5 2	8 8 2 3 8
Row 27	6 2 4 3 2	2 7 9 3 5	1 0 0 7 7	6 5 9 7 2	2 8 8 7 1
Row 28	8 5 3 6 4	7 8 3 1 2	9 4 3 2 7	4 2 4 1 2	8 4 1 9 7
Row 29	5 1 5 7 8	9 1 9 7 5	8 9 1 4 3	7 1 6 9 4	3 6 4 5 5
Row 30	3 5 8 7 0	1 0 8 3 3	5 9 7 6 4	2 2 4 6 0	8 7 1 3 4

This page is intentionally blank.

Exploring Sampling Variability

Sample Speak

Lesson 25-1 Sample Statistic and Sampling Variability

My Notes

Learning Targets:

- Understand the difference between variability in a population and sampling variability.
- Know that increasing the sample size decreases sampling variability.

SUGGESTED LEARNING STRATEGIES: Think-Pair-Share, Create Representations, Look for a Pattern, Summarizing, Paraphrasing, Interactive Word Wall

In Activity 24, each student in your class selected a random sample of size $n = 5$ from the population of the 100 students in a summer reading club. For this population, the mean (average) number of books read was _____.

MATH TIP

The letter n is used to represent the sample size.

1. When you selected a sample of size 5 from this population and calculated the sample mean, was the sample mean you found exactly equal to the population mean?

2. **Reason abstractly.** If you were to take a different random sample of five students from this population, do you think you would get a sample mean exactly equal to the population mean? Do you think the sample mean for this new sample would be the same as the sample mean from the first random sample you selected in Activity 24? Explain why you think this.

3. When the students in your class each selected a random sample from the population of summer reading club students, why did everyone *not* get the same value for their sample means?

My Notes

MATH TERMS

A ***sample statistic*** is a numerical value that is calculated using data from a sample.

MATH TERMS

Sampling variability is the variability in the values of a sample statistic that occurs because different samples include different individuals when random samples are selected from a population.

Suppose that you are going to select a random sample from a population, and you are going to calculate a **sample statistic**, such as the mean or the median, for your sample. The value of your sample statistic will vary depending on the sample that you happen to select, because different samples may include different individuals from the population. This sample-to-sample variability is called **sampling variability**. In this activity, you will explore sampling variability and see why it is important to think about sampling variability when you try to use data from a sample to learn about a population. We will start by revisiting the summer reading club population and then move on to investigating a mock election.

Revisiting the Summer Reading Club Population

To complete this section, you will need the numbered Summer Reading Program Data page and the Random Digit Table that you used in Activity 24.

Before you complete Items 4–7 below, your teacher will assign you a sample size of 10, 15 or 20.

My assigned sample size is _____.

Toss a paper clip onto the page of random digits. Mark the digit on the page that is closest to the center of the larger loop of the paper clip. Starting with that digit and taking two digits at a time, write down enough two-digit numbers to select your sample. (For example, if your assigned sample size is 10, you will need 10 two-digit numbers. Ignore any two-digit numbers that are repeats of previous two-digit numbers.) Write your two-digit numbers below.

4. Find the rectangles corresponding to the numbers you obtained from the random digit table. For each of these selected students, determine the number of books read and record those values below.

5. Calculate the sample mean for your sample.

Mean for random sample 1:

My Notes

6. Repeat this entire process to select another random sample, and calculate the sample mean for this sample.

   Mean for random sample 2:

7. Are the two sample means you calculated in Items 5 and 6 equal? If not, explain why they are different.

Add your two sample means to a dot plot on a poster that your teacher will provide. Be sure to add your dots to the poster that corresponds to the sample size you were assigned.

8. Take a look at the three dot plots. How are the dot plots for random samples of size $n = 10$, $n = 15$, and $n = 20$ similar? How are they different?

9. Do sample means tend to differ more from one sample to another for random samples of size 10 or for random samples of size 20?

10. **Make sense of problems.** If you were going to select a random sample from the summer reading club population and use the sample mean as an estimate of the population mean, would you expect the sample mean to be closer to the population mean if the sample size was 10 or if the sample size was 20? Explain how the dot plots of the sample means that your class constructed support your answer.

My Notes

Check Your Understanding

11. For each of the statements below, decide if it describes variability in a population or if it describes sampling variability.

Statement 1: There is variability in the number of books read, because not every student read the same number of books.

Statement 2: The mean number of books read for the students in one random sample of five summer reading club students may be different from the mean for the students in a different random sample.

12. Does sampling variability increase or decrease if you increase the sample size?

LESSON 25-1 PRACTICE

13. The numbers represent number of books read by 20 students in the summer reading program. Calculate the sample mean for the sample.

1 4 4 9 16 5 10 4 10 12 1 10 3 9 1 4 15 5 12 12

14. The two-digit numbers represent rectangles of books read in the summer reading program for a sample of 15 students. Locate the appropriate rectangle on the sheet and record the books read by each student. Then calculate the sample mean for the sample.

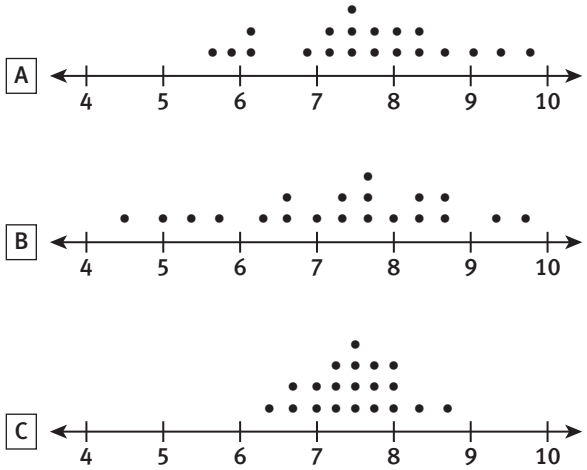
94 03 59 49 96 55 92 29 51 15 12 14 78 84 27

15. Refer to the Random Digit Table to select two-digit random numbers to represent a sample of size 10.

- Record the random numbers in the chart.
- Locate the appropriate rectangle on the sheet to determine how many books were read by each of these students.
- Calculate the sample mean.

My Notes

16. Construct viable arguments. Refer to the three dot plots labeled A, B, and C. If the graphs represent the plots of sample means from samples of size 11, 14, and 18, which dot plot represents which sample size? Explain your reasoning.



17. Think of the population consisting of all of the students who try out for sports teams at your school. For each of the statements below, decide if it describes variability in the population or sampling variability.

Statement 1: If you recorded the number of years each student has been playing sports, there would be variability in these numbers. Not all students have been playing sports for the same number of years.

Statement 2: The mean age of the students in one sample of 10 students would probably be different from the mean age for 10 students in a different sample.

Statement 3: The average height of students who try out for athletic teams will tend to vary from one random sample of 5 students to another.

Statement 4: Not all students weigh the same amount, so there is variability in students' weights.

My Notes

Learning Targets:

- Use data from a random sample to estimate a population characteristic.
- Understand the implications of sampling variability when estimating a population characteristic.
- Use data from a random sample to draw a conclusion about a population.

SUGGESTED LEARNING STRATEGIES: Think-Pair-Share, Create Representations, Look for a Pattern, Summarizing, Paraphrasing, Interactive Word Wall

Predicting Election Results

Suppose that your school has decided to carry out a mock election to see who the students at your school would elect president of the United States. Each student will vote and will choose between the Republican candidate (Candidate 1) and the Democrat candidate (Candidate 2). The school newspaper wants to write an article on the upcoming election and has asked you to predict who will win the election, Candidate 1 or Candidate 2.

1. One way to proceed would be to interview *every* student at your school and ask each one which of the candidates he or she plans to vote for. What is the advantage of this approach? What is one disadvantage of this approach?

Suppose that you decided to select a sample of 20 students from your school. You ask each student which candidate will receive their vote. You calculate the proportion of the students in the sample that plan to vote for Candidate 1 and use this as an estimate of the proportion of all the students at the school who will vote for Candidate 1.

2. The editor of the school paper suggests that you use the 20 students in your first period class for your sample. Explain why this is not a good idea. Share your reasoning with your group members and list any details you may not have considered before. If you do not know the exact words to describe your ideas, use synonyms or request assistance from group members to help you convey your ideas. Use nonverbal clues such as raising your hand to ask for clarification of others' ideas.

My Notes

3. What would be a better way to select the sample?

You plan to select a random sample of 20 students and use the proportion of students in this sample who plan to vote for Candidate 1 to predict the election result. For example, if 12 of the 20 students in the sample say they will vote for Candidate 1, you would estimate that the proportion of students at the school who will vote for Candidate 1 is $\frac{12}{20} = 0.60$, or 60%. Because this is more than 50%, you would predict that Candidate 1 will win the election.

4. **Reason quantitatively.** If you select a random sample of 20 students and predict the election winner, can you be sure that your prediction will be correct? Explain why or why not.

5. Do you think that the chance of making an incorrect prediction will be greater if the actual percentage of students in the entire population who will vote for Candidate 1 is 20%, or if the actual population percentage is 40%? Explain why you think this.

Your teacher will now assign you to one of five groups. Each group will receive a paper bag containing red and white plastic beads. Each bag has a different proportion of red beads. The proportion of red beads in the bag is written on the bag. Pretend that the beads in the bag represent students at your school. A red bead represents a student who will vote for Candidate 1, and a white bead represents a student who will vote for Candidate 2.

My Notes

Record the percentage who will vote for Candidate 1 for your assigned population below.

Percentage who will vote for Candidate 1: _____

6. For your population, will Candidate 1 win the election?

Work with your group to complete Items 7–14.

7. Select a random sample of 20 "students" from your population and calculate the proportion of students in your sample of 20 who will vote for Candidate 1. (For example, if 7 of the 20 beads in your sample are red, the sample proportion would be $\frac{7}{20} = 0.35$, and you would predict that Candidate 1 would lose the election.) Complete the first row of the table that appears on page 355 at the end of this activity. (Carefully tear this page out of your book).

Place the beads back in the bag and mix up the beads in the bag. Then repeat the process of selecting a random sample 24 more times to complete the rest of the table.

8. How many of your 25 predictions were wrong?

9. Based on the last column of the table you created, was the sample proportion usually close to the actual population proportion?

My Notes

10. For your 25 random samples, answer the following:

a. What was the largest prediction error?

b. What was the smallest prediction error?

c. What was a typical prediction error?

Your teacher will provide each group with a piece of chart paper. Your group should make a poster that provides the following information:

- The actual population proportion who will vote for Candidate 1 for your population
- A dot plot of the sample proportions from the 25 random samples
- The number of samples that led to a prediction that was wrong
- The largest prediction error
- The smallest prediction error
- The value of a typical prediction error

After your group has finished making its poster, ask your teacher where to display it along with the posters from the other groups.

Take a few minutes to look at all of the group posters. Then use them to answer the following items.

11. **Use appropriate tools strategically.** How are the centers of the dot plots related to the actual population proportions?

12. Is the sample-to-sample variability in the sample proportions about the same in all five of the dot plots?

My Notes

13. Do sample proportions tend to be closer to the actual value of the population proportion when the population proportion is 0.2 or 0.5?

14. **Reason abstractly.** If the population had consisted entirely of white beads, what would the dot plot of sample proportions for 25 random samples have looked like?

Your teacher will now provide a mystery bag. This bag contains a population of red and white beads and the proportion of red beads in the bag is either 0.2, 0.3, 0.4, 0.5, or 0.6. A student in the class will select a random sample of 20 beads from this population.

For the random sample, record the following:

Sample size:
Number of red beads:
Sample proportion of red beads:

15. Based on the sample proportion, would you rule out any of the proportions 0.2, 0.3, 0.4, 0.5, or 0.6 as possibilities for the population proportion for the mystery bag? Explain your reasoning. (*Hint:* you may want to look at the posters that your class created.)

16. Which of 0.2, 0.3, 0.4, 0.5, or 0.6 do you think is the actual value of the population proportion of red beads for the mystery bag? Explain your reasoning.

My Notes

Check Your Understanding

17. Suppose that you are going to select a random sample of students from your school, measure the heights of the students in your sample, and then find the mean height for your sample. There are many different sets of students who could end up forming your sample. Will the sample mean be the same for all possible samples? If so, explain why. If not, what is the phrase used for the fact the sample means will vary according to the sample selected?

18. Return to the context of the previous item. The sample mean will vary according to what sample you happen to select. If you want to reduce this variation, would you choose to select a sample of size 5 or a sample of size 25?

Use the following to answer Items 19 and 20.

Suppose that 60% of the students at your school favor a proposed school policy that would ban cell phones at school. Also suppose that every student in your class selected a different random sample of 50 students from your school and calculated the proportion of students in the sample who favor the ban.

19. If your class constructed a dot plot of all these sample proportions, where would you expect the dot plot to be centered?

20. If every student had selected a random sample of size 25 instead of a sample of size 50, would you expect a dot plot of the sample proportions to be more spread out or less spread out than in a dot plot of sample proportions for samples of size 50?

My Notes

LESSON 25-2 PRACTICE

Using a deck of standard playing cards, remove the jokers and the red twos, threes, fours, fives, sixes, and sevens. Shuffle (mix) the cards remaining in the deck thoroughly. Select samples without looking at the type of card being selected. Return the cards in the sample to the deck and shuffle thoroughly before selecting a new sample. (If decks of playing cards are not available, you could substitute the random number table using two-digit numbers to represent a selection where selecting 01 to 65 represents choosing a black card and selecting 66 to 00 represents choosing a red card.)

Use this situation to simulate an election in which students are determining whether to travel to an historic site to write about what they learn there or to travel to an amusement park to study the science and mathematics of the rides at the park. Select 20 samples of 10, in which red indicates a vote for the historic site destination and black indicates a trip to the amusement park.

21. **Model with mathematics.** Create a data chart to collect the results of the twenty samples. Use the following headings:
- Sample Number
- Number of Votes for the Amusement Park (black)
- Proportion of Votes in the Sample for the Amusement Park (black)
- Prediction (Win, Lose, Tie for the Amusement Park Trip)
- Prediction Error (Difference Between the Sample Proportion and the Actual Population Proportion, 0.65)

22. Construct a dot plot of the sample proportions from the 20 random samples.

23. How many samples led to a prediction that was wrong?

24. State the largest prediction error.

25. State the smallest prediction error.

26. What is the value of a typical prediction error?

ACTIVITY 25 PRACTICE

1. In your own words, explain the difference between variability in a population and sampling variability.

2. Think of the population consisting of all of the teachers who work in your school district. For each of the statements below, decide if it describes variability in the population or sampling variability.

   Statement 1: If you recorded the number of years each of the teachers has been teaching, there would be variability in these numbers. Not all teachers have been teaching for the same number of years.

   Statement 2: The mean age of the teachers in one sample of 10 teachers would probably be different from the mean age for 10 teachers in a different sample.

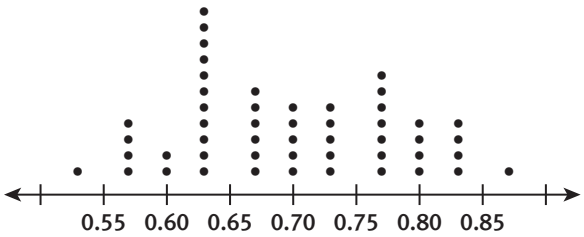
   Statement 3: The proportion of teachers who use public transportation to get to school will tend to vary from one random sample of teachers to another.

   Statement 4: Not all teachers are the same age, so there is variability in teachers' ages.

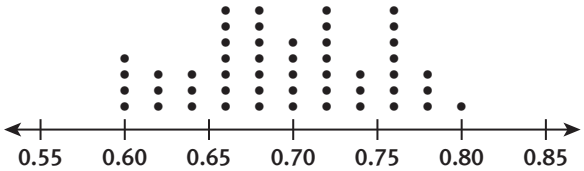
3. Devon selected 50 different random samples of students from her school. For each sample, she determined the proportion of students who prefer rock music to rap music. Frank selected 50 different random samples of students from the same school and determined the proportion who prefer rock music for each of his samples. They produced the dot plots shown below.

   Devon used random samples of size 30. All of Frank's random samples had the same sample size. Do you think that Frank used a sample size that is less than 30, equal to 30, or greater than 30? Explain why you think this.

   Devon's sample proportions:

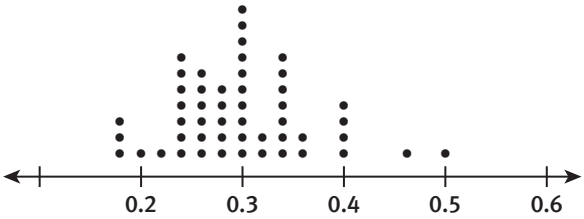


   Frank's sample proportions:

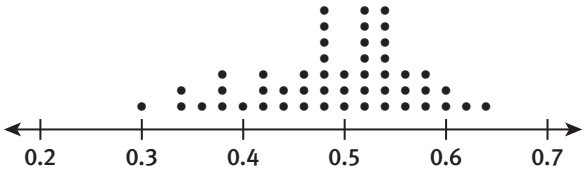


4. Two populations each consist of 1000 beads. In one population, 30% of the beads are red. In the other population, 50% of the beads are red. Many random samples of size 50 were selected from each population, and dot plots were drawn showing the proportions of red beads in the samples. The dot plots are shown below.

   Sample proportions for random samples from a population with 30% red beads:

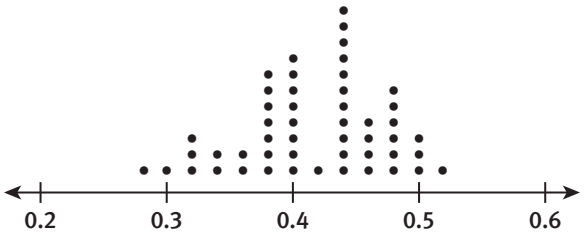


   Sample proportions for random samples from a population with 50% red beads:



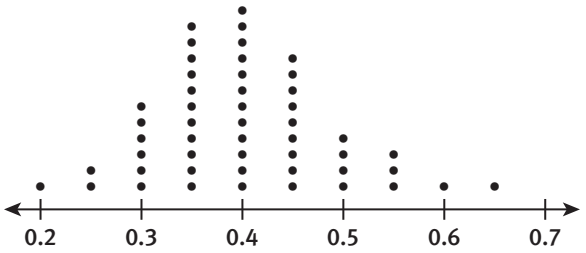
Based on these plots, if you use the sample proportion from a random sample of size 50 as an estimate of the population proportion, do you think your estimate would be closer to the actual value of the population proportion when the population proportion is 0.3 or when the population proportion is 0.5? Explain your choice.

5. Below is a dot plot that shows the sample proportions of red beads for random samples of size 50 from a population of beads that has 40% red beads.

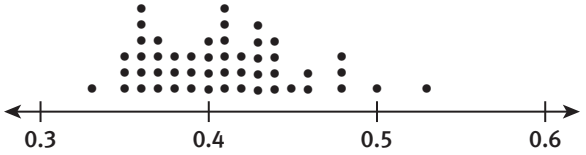


Which of the following do you think is a dot plot of the proportions of red beads for random samples of size 20? Explain your choice.

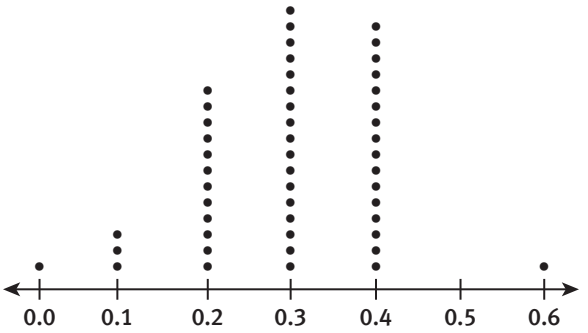
Dot Plot 1



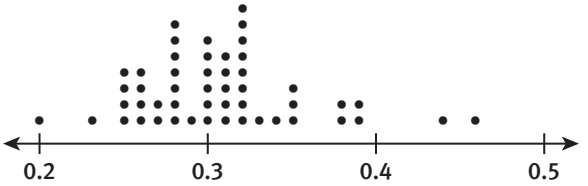
Dot Plot 2



Dot Plot 3



Dot Plot 4



6. Which of the four dot plots from the previous item is most likely to have been generated using sample means from a sample size greater than 50? Explain.

MATHEMATICAL PRACTICES
Model with Mathematics

7. In this activity, you simulated selecting a random sample of 20 students from a population by selecting 20 beads from a bag that contains red beads and white beads. Suppose that 40% of a large population of students would vote for Candidate 1 in a mock election. Can you think of a way to use a random digit table to simulate selecting the random sample that would be used instead of selecting beads from a bag? Describe how you would do this.

Table for Activity 25

Sample	Number in Sample Who Will Vote for Candidate 1	Proportion in Sample Who Will Vote for Candidate 1	Prediction (Candidate 1—Win, Lose or Tie)	Prediction Error Difference Between Sample Proportion and Actual Population Proportion (Sample Proportion Minus Population Proportion)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				

This page is intentionally blank.

Random Sampling and Sampling Variability

SCHOOL POPULATIONS

Embedded Assessment 1

Use after Activity 25

Write your answers on notebook paper or grid paper. Show your work.

1. Each of the following describes a method for selecting a sample of 10 students from the students at your school. For each method described, decide if it would result in a random sample. For purposes of this item, suppose that there are 1000 students at your school and that you have a list of all 1000 student names.

Method 1: Write all 1000 student names on slips of paper. Put these slips in a box, mix them well, and then draw out 10 names.

Method 2: Number the students on the list from 000 to 999. Then use the first 10 students on the list.

Method 3: Number the students on the list from 000 to 999. Then use a table of random digits to obtain 10 different blocks of three digits. Use these 10 three-digit numbers to identify which students to select.

Method 4: Use the first 10 students that arrive at school tomorrow.

Method 5: Use the 10 students who make up the girls' volleyball team at your school.

2. Of the methods described in Item 1 that would result in a random sample, which one do you think would take the least time to implement? Explain why you think this method would be the quickest.

Random Sampling and Sampling Variability

SCHOOL POPULATIONS

Write your answers on notebook paper or grid paper. Show your work. Use the following information to answer Items 3–6.

A population consists of the 1000 students enrolled at Morro Bay High School. Some students live very close to the school and others live farther away. Students who live more than three miles from the school ride the bus to school.

3. You ask each student at the school how far away from school they live and record these numbers. Would there be variability in the distances? Is this variability in a population or is it sampling variability?

4. Fran and Zoe each selected a different random sample of students from this population. They each recorded how far from school the students in their sample lived and calculated the sample mean distance from school. Would you expect the two sample averages to be equal or would you expect them to differ? Is this variability in a population or sampling variability?

5. Fran selected a random sample of 20 students and Zoe selected a random sample of 50 students. Do you think that Fran's sample mean or Zoe's sample mean would be closer to the actual mean distance for the whole population? Explain why you think this.

6. Can you be *certain* that the sample mean that you picked in Item 5 will be the one that is closer? Explain.

Write your answers on notebook paper or grid paper. Show your work.
Use the following information to answer Items 7–10.

A population consists of 600 girls and 400 boys. Jose wanted to investigate what he could expect to happen if he were to take a random sample of 20 people from this population and calculate the proportion of girls in the sample.

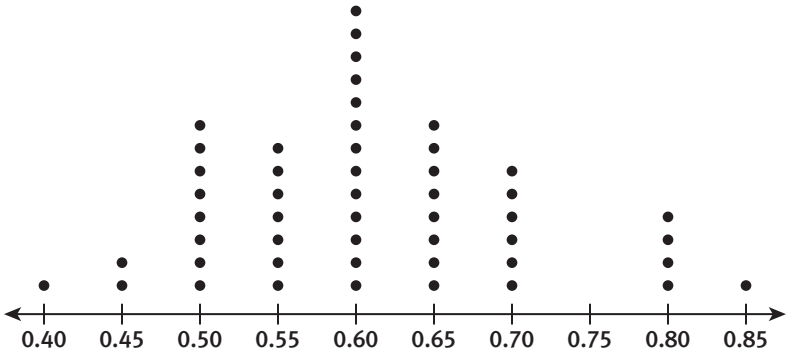
Jose selected 20 students at random and recorded the following data (B = boy and G = girl):

G	G	B	B	G	B	G	B	B	B
G	B	G	G	G	B	G	B	B	G

7. What is the proportion of girls in Jose's sample?

8. Jose's sample proportion was not equal to 0.6, even though 60% of the people in the population are girls. Does this mean that Jose did something wrong when he selected the sample? Explain.

9. Jose decided to take more random samples from this population. He selected 50 different random samples of 20 students. For each of these samples, he calculated the proportion of girls in the sample. A dot plot of Jose's sample proportions is shown below. Did any of Jose's samples result in a sample proportion that was different from the actual population proportion of girls by more than 0.2?



10. There are 1000 students at Jose's school. He selects a random sample of 20 students from his school and six of the students in the sample are girls. Do you think that the proportion of girls at Jose's school is 0.6? Explain why or why not. (*Hint:* Think about the dot plot in Item 9.)

Scoring Guide	Exemplary	Proficient	Emerging	Incomplete
	The solution demonstrates these characteristics:			
Mathematics Knowledge and Thinking (Items 1, 3, 4, 5, 6, 7, 8, 9, 10)	• Clear and accurate understanding of populations, population variability, and sampling variability. • Effective understanding and accuracy in making predictions and drawing conclusions about a population.	• A functional understanding of populations, population variability, and sampling variability. • Making predictions and drawing conclusions about a population that are largely correct.	• Partial understanding of populations, population variability, and sampling variability. • Partially correct predictions and conclusions about a population.	• Inaccurate or incomplete understanding of populations, population variability, and sampling variability. • Inaccurate or incomplete predictions and conclusions about a population.
Problem Solving (Items 7, 9)	• An appropriate and efficient strategy that results in a correct answer.	• A strategy that may include unnecessary steps but results in a correct answer.	• A strategy that results in some incorrect answers.	• No clear strategy when solving problems.
Mathematical Modeling / Representations (Items 1, 3, 4, 5, 6, 8, 9, 10)	• Clear and accurate understanding of sampling a population and obtaining random samples.	• An understanding of population samples and random samples that is largely correct.	• Partial understanding of population samples and random samples.	• Inaccurate or incomplete understanding of population samples and random samples.
Reasoning and Communication (Items 2, 3, 4, 5, 6, 8, 9, 10)	• Precise use of appropriate math terms and language to explain sampling methods, variability, and predictions.	• An adequate explanation of sampling methods, variability, and predictions.	• A misleading or confusing explanation of sampling methods, variability, and predictions.	• An incomplete or inaccurate explanation of sampling methods, variability, and predictions.

Comparative Statistics

Seventh-Grade Students

Lesson 26-1 Two Sample Means

My Notes

Learning Targets:

- Compare the means of two numerical samples.
- Understand that a *meaningful* difference between two sample means is one that is greater than would have been expected due to sampling variability alone.
- Use data from random samples to compare populations.

SUGGESTED LEARNING STRATEGIES: Think-Pair-Share, Create Representations, Summarizing, Paraphrasing, Interactive Word Wall

In this activity, you will spend some time learning about a population of seventh-grade students. The population you will be considering is a large group of seventh graders who participated in the Census at School project.

The Census at School project is an international program to help students develop the ability to think statistically. Students from many countries have participated, including students from the United States, Canada, the United Kingdom, Ireland, Japan, Australia, New Zealand, and South Africa. These students have all completed an online survey to provide information about themselves and their homes, schools, and social lives. Some questions on the survey are asked only in a particular country, but there are 13 questions that are included on the survey in all of the countries that participate in Census at School.

In this activity, you will be looking at random samples from the population of Census at School seventh graders. Because not all seventh graders participate in Census at School, keep in mind that this population is *not* the same as the population of *all* seventh graders. This means that any conclusions you make based on sample data will only be about the population of Census at School seventh graders.

Let's start by focusing only on Census at School seventh graders in the United States. One question on the Census at School survey asked students how many minutes it takes them to travel to school.

My Notes

The U.S. Census at School website (www.amstat.org/censusatschool) has a random sampler that will select a random sample of students from the U.S. Census at School population. The random sampler was used to select a random sample of 20 seventh graders. The data on time to travel to school for these twenty students are shown here:

Sample 1: Time to travel to school (in minutes)

5	7	10	10	15	15	15	15	20	20
20	20	20	20	30	40	45	60	60	60

The 20 times are arranged in order from shortest to longest.

1. For Sample 1, calculate the following:

a. the sample mean

b. the sample median

2. For Sample 1, find the first and third quartiles and the interquartile range (IQR).

3. Draw a box plot for the data of Sample 1.

MATH TIP

Remember that the **sample mean** is the average of the data values, and the **sample median** is the middle data value in a list that has been ordered from smallest to largest. (If there is an even number of data values, the median is the average of the middle two values.)

MATH TIP

Remember that the **first quartile** is the median of the lower half of the data, and the **third quartile** is the median of the upper half of the data.

The **interquartile range** is the distance between the first and third quartiles.

My Notes

A second random sample (Sample 2) of 20 students is selected from the population of U.S. Census at School seventh graders. The travel times for these 20 students (arranged in order from shortest to longest) are:

Sample 2: Time to travel to school (in minutes)

5	7	8	10	10	15	15	20	20	20
20	20	30	30	30	40	45	55	60	60

4. For Sample 2, calculate the following:

a. the sample mean

b. the sample median

c. The first and third quartiles

d. The interquartile range

5. **Model with mathematics.** Draw a box plot for the data of Sample 2.

My Notes

6. Does it surprise you that the box plot for Sample 1 and the box plot for Sample 2 are not identical? Explain why or why not.

7. What term is used to describe differences in sample statistics for different random samples from the same population?

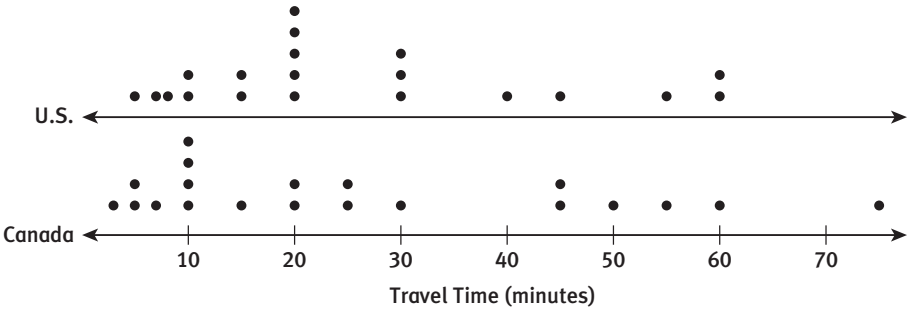
8. Are the mean of Sample 1 and the mean of Sample 2 the same? Is this what you would expect? Explain.

9. Calculate the difference in the two sample means by calculating Sample 1 mean − Sample 2 mean.

Because Sample 1 and Sample 2 were selected from the same population, we know that if there is any difference in the two sample means, it is because of chance differences that occur from one random sample to another. Two random samples from the same population will probably have different sample means. This sampling variability is what makes trying to compare two populations based on sample data tricky! Let's investigate.

My Notes

Suppose that you wanted to know if the average travel time to school for Census at School seventh graders in Canada and the average travel time for Census at School seventh graders in the U.S. are different. A random sample of 20 seventh graders was selected from each of these two populations. Dot plots of the travel times for these samples are shown here:



The sample mean for the U.S. sample is 26.00 minutes, and the sample mean for the Canada sample is 26.25 minutes.

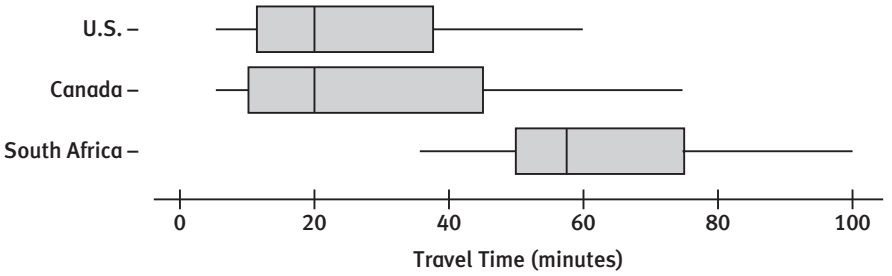
10. Calculate the difference between the U.S. and Canada sample means by calculating U.S. sample mean − Canada sample mean.

11. In Item 9 you calculated the difference in the sample means for Sample 1 and Sample 2, which were random samples from the same population. How does the difference between the U.S. sample mean and the Canada sample mean compare to the difference that you calculated in Item 9?

12. **Reason abstractly.** Based on the U.S. sample mean and the Canada sample mean, do you think it is reasonable to conclude that the U.S. population mean and the Canada population mean are different, or do you think that the two population means might be the same? Explain your reasoning.

My Notes

Take a look at the three box plots shown below. One box plot is from the random sample from the U.S. Census at School population, one is from the random sample from the Canada Census at School population, and one is from a random sample from the South Africa Census at School population.



The sample means are

U.S. sample mean: 26.00 minutes
Canada sample mean: 26.25 minutes
South Africa sample mean: 60.75 minutes

Working with a partner, use the box plots and sample means to answer Items 13–18. Make notes as you listen to your partner. Ask and answer questions clearly to aid comprehension and to ensure understanding of your partner's ideas.

13. How are the box plots for the U.S. sample and the Canada sample similar? How are they different? Are the differences large or small?

My Notes

14. Explain how the box plots and sample means support the following statement:

Based on the U.S. random sample and the Canada random sample, we cannot be sure that the distributions of travel times and the mean travel times for the U.S. population and the Canada population are different.

15. How are the box plots for the U.S. sample and the South Africa sample similar? How are they different? Are the differences large or small?

16. **Construct viable arguments.** Explain how the box plots and sample means support the following statement:

Based on the U.S. random sample and the South Africa random sample, the mean travel times for the U.S. population and the South Africa population are probably different.

17. If there is a lot of overlap in the box plots of random samples from two populations, does this suggest that the difference in the two sample means or medians might be due to sampling variability or does it suggest that the two populations might differ in some important way?

My Notes

18. If there is not a lot of overlap in the box plots of a random sample from each of two populations, does this suggest that the difference in the two sample means or medians might be due to sampling variability or does it suggest that the two populations might differ in some important way?

When comparing two populations based on the means of random samples from the two populations, there are two possibilities to be considered:

Possibility 1: The difference in the sample means is *not meaningful.* The difference is not very big. The two population means might be the same and the sample means may be different only because sample means tend to differ from the population mean (sampling variability).

Possibility 2: The difference in the sample means is *meaningful*. The difference is big enough that we think that it cannot be due to just sampling variability. The two population means are probably different.

At this point, you should be wondering just how big the difference in sample means needs to be for us to choose Possibility 2 and conclude that the population means are different. That is a good question! Completing the rest of this activity will help you to answer this question.

Check Your Understanding

19. Suppose that the mean height for a random sample of 25 U.S. Census at School seventh graders is 163 cm and that the mean height for a random sample of 25 Canada Census at School seventh graders is 165 cm. Explain why this does not tell us that the mean height of *all* U.S. Census at School seventh graders is different from the mean height of *all* Canada Census at School seventh graders.

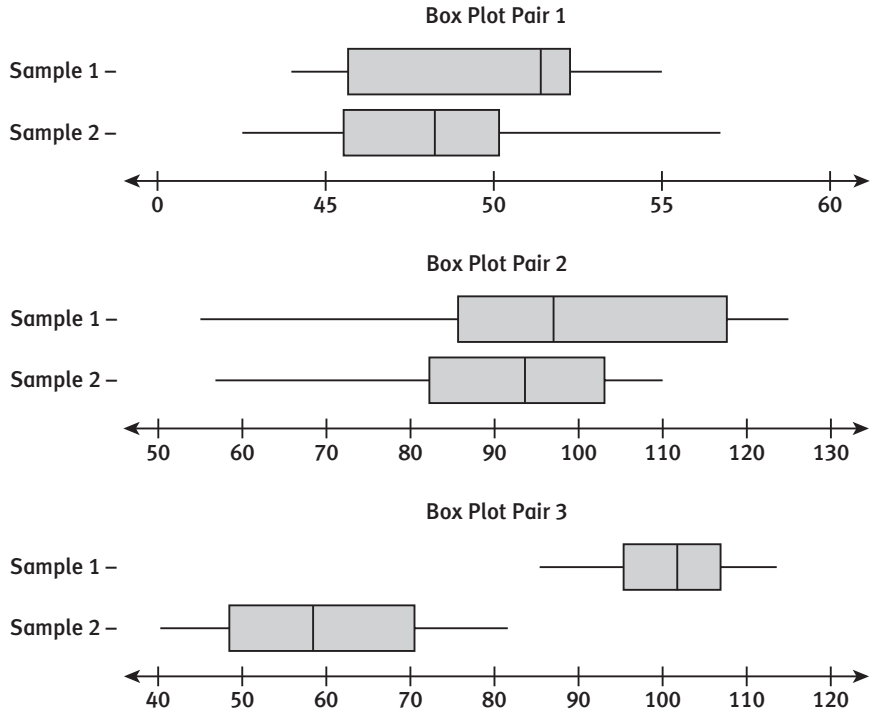
20. For which of the following cases would it be reasonable to conclude that two population means are different based on the sample means from random samples from the populations?

Case 1: The difference in the two sample means is greater than what would be expected due to sampling variability if the population means were equal.

Case 2: The difference in the sample means is consistent with what would be expected due to sampling variability if the population means were equal.

My Notes

21. For which of the following pairs of box plots is it most likely that the two samples were drawn from populations that had different population means? What is it about the box plots that led you to this choice?



My Notes

LESSON 26-1 PRACTICE

Consider the two samples of data from the McKenzie School. The numbers represent the time in seconds that it took each child to cover a distance of 50 meters.

Girls' Times: 8.3, 8.6, 9.5, 9.5, 9.6, 9.8, 9.9, 9.9, 10.0, 10.0, 10.0, 10.1, 10.3, 10.5

Boys' Times: 7.9, 8.0, 8.2, 8.2, 8.4, 8.6, 8.8, 9.1, 9.3, 9.5, 9.8, 9.8, 10.0, 10.1, 10.3

22. Calculate the sample mean and sample median of each data set.

23. Based on the sample means, do you conclude that the distributions of times from the boys' population and girls' population are different? Explain.

24. Calculate the first quartile and third quartile for each data set.

25. **Model with mathematics.** Draw a box plot for each data set on the same scale.

26. Based on a comparison of the box plots, do you conclude that the population means for the boys' times and girls' times are significantly different? Explain.

My Notes

Learning Targets:

- Compare population means for populations with approximately the same amount of variability.
- Express the difference in the sample means in terms of mean absolute deviation (MAD).
- Draw conclusions about population differences based on sample size and the difference in sample means relative to the MAD.

SUGGESTED LEARNING STRATEGIES: Think-Pair-Share, Create Representations, Summarizing, Paraphrasing, Interactive Word Wall

To decide if a difference in two sample means is meaningful, we need to look at what kinds of differences are typical when the population means are equal. In other words, we need to see what kinds of differences are typical of sampling variability alone.

To investigate, we will consider the population of U.S. Census at School seventh-grade boys. Using the Census at School random sampler, a random sample of 10 boys from this population was selected. The number of hours usually spent doing homework each week for each boy in the sample is shown here:

Sample 1 Homework Hours per Week

9	4	2	4	2	2	4	4	3	11

A second random sample of 10 boys from the population resulted in the following 10 homework times:

Sample 2 Homework Hours per Week

9	2	9	4	10	4	1	4	3	7

For the entire population of U.S. Census at School seventh-grade boys, the homework time distribution has population mean = 6.0 hours and population MAD = 3.4 hours. These values were calculated using all of the data values in the entire population.

1. Calculate the following:
   a. mean of Sample 1:
   b. mean of Sample 2:
   c. difference in sample means (Sample 1 − Sample 2):

MATH TIP

A difference in sample means is meaningful if it is bigger than what would be expected from sampling variability alone.

MATH TIP

Remember that the **MAD (mean absolute deviation)** is the average distance of the observations from the mean of the data set. It is a measure of variability in a data set.

My Notes

MATH TIP

To express a difference in sample means in terms of the MAD, calculate:

$$\frac{\text{difference in sample means}}{\text{MAD}}$$

One way to look at the difference in two sample means is to express this difference in terms of the MAD. Here the population MAD was 3.4. Dividing the difference in sample means by the MAD gives

$$\frac{-0.8}{3.4} = -0.24$$

This tells us that the difference in sample means was 0.24 times the MAD for this population.

Expressing the difference in sample means in terms of the MAD allows us to judge the difference in sample means relative to the variability in the population. This is helpful, because a difference of 1 might be considered small or large depending on the context. For example, a difference of 1 hour is small if we are talking about the amount of time that seventh graders spend at school in a year, but a difference of 1 hour is very large if we are talking about how long seventh graders spend getting ready for school each day!

Here are the times spent on homework per week for two more random samples from the population of U.S. Census at School seventh-grade boys:

Sample 3 Homework Hours per Week

9	11	5	11	12	1	1	6	6	3

Sample 4 Homework Hours per Week

6	9	8	12	11	4	1	7	1	10

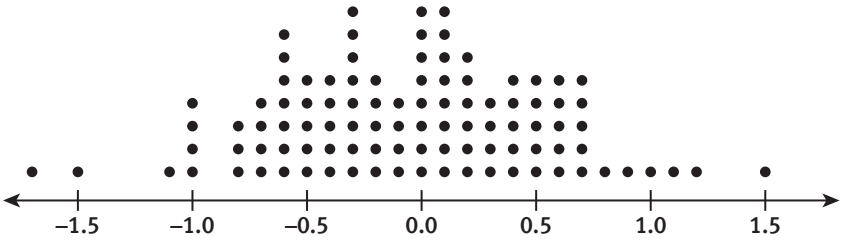
2. Calculate the sample means and the difference in the sample means. Express the difference in sample means in terms of the population MAD.
   a. mean of Sample 3:
   b. mean of Sample 4:
   c. difference in sample means (Sample 3 − Sample 4):
   d. difference in sample means in terms of MAD:

3. **Reason quantitatively.** Based on your answer to Item 2, would you be surprised if two different random samples of ten U.S. Census at School seventh-grade boys had sample mean number of hours spent on homework that differed by 0.1 MAD? Explain your thinking.

My Notes

To understand what kind of differences in sample means result from sampling variability alone, we could continue to select pairs of random samples from the population of U.S. Census at School seventh-grade boys and express each difference in sample means in terms of the population MAD. This was done 98 times, so that with the two that you have already looked at (which had differences of − 0.24 MAD and − 0.12 MAD) there were a total of 100 pairs of samples. The first 10 new pairs are shown in the table that follows. All 100 differences were used to make the dot plot that follows the table.

First Sample Mean	Second Sample Mean	Difference	Difference in Terms of MAD
3.80	6.65	−2.85	−0.84
6.07	6.69	−0.62	−0.18
7.59	2.53	5.06	1.49
6.08	6.24	−0.16	−0.05
7.11	4.92	2.19	0.64
4.71	6.66	−1.95	−0.57
4.69	5.29	−0.60	−0.18
4.19	4.34	−0.15	−0.04
7.30	5.10	2.20	0.65
7.38	6.68	0.70	0.21



Difference in Terms of MAD, $n = 10$

4. Based on the dot plot above, would you be surprised if two different random samples of ten U.S. Census at School seventh-grade boys had sample means that differed by 0.5 MAD? Explain your thinking.

My Notes

5. Based on the dot plot on the previous page, would you be surprised if two different random samples of ten U.S. Census at School seventh-boys had sample means that differed by 2.0 MAD? Explain your thinking.

Here are two samples of 10 seventh graders. One is a random sample from the population of U.S. Census at School seventh-grade boys. The other sample is a random sample from the population of U.S. Census at School seventh-grade girls. Use these samples to answer Items 6–9.

Boys Homework Hours per Week

10	6	2	3	1	9	11	1	8	3

Girls Homework Hours per Week

3	10	10	13	2	10	13	5	3	5

6. Calculate the following:
   a. sample mean for girls:
   b. sample mean for boys:
   c. difference in sample means (girls − boys):

7. Express the difference in the two sample means in terms of the MAD for the population of seventh-grade boys (the MAD for the population of seventh-grade girls was about the same as the MAD for the population of boys). The MAD for the population of boys was 3.4.

8. Locate the value you just found on the horizontal axis given in the dot plot that appears just before Item 4. Do you think that the difference in the sample mean number of homework hours for girls and boys might be just due to sampling variability? Explain why you think this.

My Notes

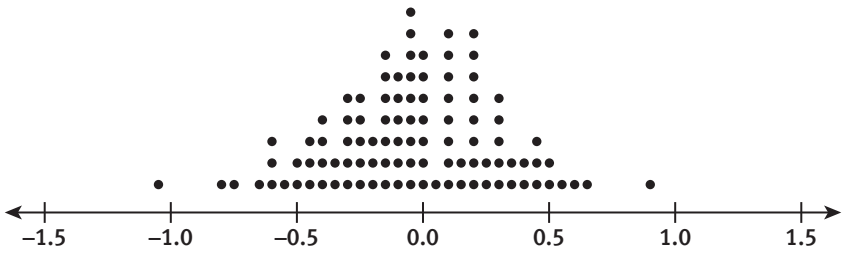
9. Suppose that the difference in sample means for girls and boys had been 2.0 MADs. Would you have answered Item 8 differently? Explain why or why not.

Now take another look at the dot plot of the differences in sample means. Notice that if two random samples of size 10 are taken from the population of U.S. Census at School seventh-grade boys, it is not likely that you would see a difference in sample means of more than 1.5 MADs. This happened in only one out of 100 pairs of random samples.

This suggests that it might be possible to say that if the difference in two sample means is less than 1.5 MADs, the difference might be just sampling variability and the difference is not big enough to conclude that the two samples came from populations with different population means.

While this works for a sample size of 10, will it also work for other sample sizes? Let's see . . .

Consider the population that consists of the number of text messages sent in a typical day for the population of U.S. Census at School seventh-grade girls. This population has a MAD of 72. One hundred pairs of random samples of 20 girls were selected from this population, and the differences in the sample means were expressed in terms of the MAD. These values were used to make the dot plot that follows.

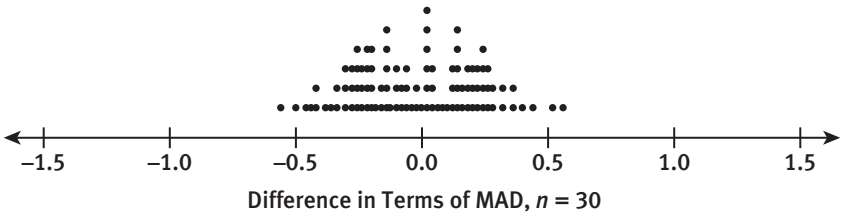


Difference in Terms of MAD, $n = 20$

My Notes

Compare the dot plot for samples of size 20 to the dot plot for samples of size 10. Notice that for samples of size 20, only one pair of samples had a difference in means that was more than 1 MAD. So for samples of size 20, if the sample means differed by 1.1 MADs we would think this is more than just sampling variability. For samples of size 10, a difference of 1.1 MADs might just be sampling variability.

You have already seen box plots for sample sizes of 10 and 20. Consider the dot plot below for pairs of samples where the sample size was 30.



10. Use appropriate tools strategically. Use the dot plot above to complete the following sentence:

For two random samples of size 30, I would think that the population means were not equal if the sample means differed by more than _____ MADs.

My Notes

11. The MAD for the number of text messages sent in a day for the population of U.S. Census at School seventh-grade girls was 72. The MAD for the population of U.S. Census at School boys was also about 72. Suppose that a random sample of 30 girls had a sample mean number of text messages sent of 68 and a random sample of 30 boys had a sample mean number of text messages sent of 53. Is the difference in sample means large enough to conclude that the mean number of text messages sent for the population of girls and the mean for the population of boys are different? Explain your answer.

CONNECT TO AP STATISTICS

In high school, you will see other methods that allow you to compare two populations.

In AP Statistics, you will learn a formal method for comparing two populations, called hypothesis testing.

The three statements in the box below can be used to informally compare two population means, as long as the population MADs are about the same.

Informal Guidelines for Comparing Two Population Means

If two populations have about the same amount of variability, then

1. **For random samples of size 10, two population means are probably different if the sample means differ by more than 1.5 MADs.**
2. **For random samples of size 20, two population means are probably different if the sample means differ by more than 1.0 MADs.**
3. **For random samples of size 30, two population means are probably different if the sample means differ by more than 0.5 MAD.**

My Notes

Check Your Understanding

For each of the following populations, indicate whether you would conclude that *the population means might be equal,* or *the population means are probably different*. Justify your choice.

12. Population 1: U.S. Census at School seventh graders
Population 2: U.S. Census at School eighth graders
Variable of interest: Hours doing homework per week
The two populations both have a MAD of about: 7 hours
Sample size: 30
Mean of random sample from Population 1: 6.3 hours
Mean of random sample from Population 2: 6.6 hours

a. Conclusion and justification:

b. Does this lead you to think that the mean number of hours spent doing homework for seventh graders and the mean for eighth graders are different?

13. Population 1: U.S. Census at School seventh-grade girls
Population 2: U.S. Census at School seventh-grade boys
Variable of interest: Hours of sleep on a school night.
The two populations both have a MAD of about: 1 hour
Sample size: 10
Mean of random sample from Population 1: 8.2 hours
Mean of random sample from Population 2: 7.1 hours

a. Conclusion and justification:

b. Does this lead us to think that the mean number of hours of sleep on a school night for seventh-grade girls and the mean for seventh-grade boys are different?

My Notes

LESSON 26-2 PRACTICE

For the following items, indicate whether you would conclude that *the population means might be equal* or *the population means are probably different*. Justify your choice.

14. Population 1: U.S. Census at School seventh graders
Population 2: South Africa Census at School seventh graders
Variable of interest: Height (in cm)
The two populations both have a MAD of about 9 cm.
Sample size: 20
Mean of random sample from Population 1: 163 cm
Mean of random sample from Population 2: 151 cm

a. Reason quantitatively. Conclusion and justification:

b. Does this lead us to think that the mean heights for U.S. seventh graders and the mean for South Africa seventh graders are different?

15. Population 1: U.S. Census at School seventh-grade boys
Population 2: U.S. Census at School seventh-grade girls
Variable of interest: Hours spent watching TV per week
The two populations both have a MAD of about 2 hours.
Sample size: 30
Mean of random sample from Population 1: 5.8 hours
Mean of random sample from Population 2: 3.3 hours

a. Reason quantitatively. Conclusion and justification:

b. Does this lead us to think that the mean number of hours spent watching TV for seventh-grade boys and the mean for seventh-grade girls are different?

16. Population 1: McKenzie School fifth-grade girls
Population 2: McKenzie School fifth-grade boys
Variable of interest: Times to cover a 50m distance
The two populations both have a MAD of about 0.3 seconds.
Sample size: 10
Mean of random sample from Population 1: 10.4 seconds
Mean of random sample from Population 2: 10.0 seconds

a. Reason quantitatively. Conclusion and justification:

b. Does this lead us to think that the mean number of seconds to cover a distance of 50m for fifth-grade girls and the mean for fifth-grade boys are different?

My Notes

Learning Targets:

- Calculate the mean absolute deviation (MAD)
- Use two random samples to compare population means.
- Draw conclusions about populations with similar amounts of variability based on the difference of two sample means.

SUGGESTED LEARNING STRATEGIES: Activating Prior Knowledge, Think-Pair-Share, Create Representations, Summarizing/Paraphrasing, Interactive Word Wall

There is one last thing to consider. In all of the examples you have seen so far, the population MAD was provided. But it is not realistic to think that you would know the value of the population MAD. The rest of this activity suggests how you might come up with an estimate of the MAD so that you can still use the guidelines on page 377 to compare two population means.

The steps that you can follow are summarized below.

Using Two Random Samples to Compare Population Means

Steps:

1. Construct a dot plot or box plot of each sample to see if the variability in each of the two samples is about the same.

   If the variability in each of the two samples looks to be about the same, continue on to Steps 2–5 below.
2. Calculate the mean and MAD for Sample 1 and the mean and MAD for sample 2.
3. Get an estimate of the common MAD by averaging the two sample MADs.
4. Express the difference in sample means in terms of the common MAD from Step 3.
5. Use the guidelines for comparing two population means (on page 377) to decide whether the difference in sample means is large enough to conclude that the population means are probably different.

My Notes

Now let us consider more data from the Census at School seventh-grade students. Suppose that you wanted to know if the average amount of time spent talking on the telephone is different for seventh-grade girls and seventh-grade boys. A random sample of 10 girls from the population of U.S. Census at School seventh-grade girls was selected using the random sampler on the Census at School website. The numbers of hours spent talking on the phone per week for these 10 girls were:

5	12	12	8	14
6	5	14	3	1

1. Calculate the sample mean number of hours spent talking on the phone for these 10 girls.

2. We can calculate the MAD for this sample by finding the distances from the sample mean for each data value and then calculating the average of these distances. The distance from the sample mean can be calculated by subtracting the sample mean from the data value and then ignoring the sign if the difference is negative. (In other words, what we're calculating here is the absolute value of the difference.) Complete the missing entries in the following table.

Data Value	Distance from the Mean
5	3
12	4
12	
8	
14	
6	
5	
14	
3	
1	

3. Calculate the value of the MAD for this sample.

MAD =

My Notes

A random sample of 10 boys from the population of U.S. Census at School seventh-grade boys was also selected. The numbers of hours spent talking on the phone per week for these 10 boys were:

5	3	1	1	0
13	0	3	10	0

4. **Model with mathematics.** Draw two dot plots of the sample data—one for the sample of girls and one for the sample of boys. Be sure to use the same numerical scale for both dot plots.

5. Does it seem from the dot plots that there is about the same amount of variability in the phone data for girls and the phone data for boys?

6. You have already calculated the following sample statistics:

   sample mean for girls = 8.0 hours
   sample MAD for girls = 4.0 hours

   Now calculate the mean and MAD for the data in the sample of boys. (You can use the table below to help organize your work for computing the MAD.)

   sample mean for boys =
   sample MAD for boys =

Data Value	Distance from the Mean
5	
3	
1	
1	
0	
13	
0	
3	
10	
0	

My Notes

7. Average the two sample MADs to get a common estimate of the population MADs.

   common MAD =

8. Express the difference in the two sample means in terms of the common MAD.

9. Based on the difference in sample means, do you think that the mean number of hours spent talking on the phone for seventh-grade girls and the mean for seventh-grade boys are probably different, or do you think that they might be the same? Justify your answer.

10. **Reason quantitatively.** Suppose the two sample sizes had been 20 instead of 10, but that the sample means and MADs were still the same as the ones you calculated. Would you have answered Item 9 differently? Explain why or why not.

My Notes

Check Your Understanding

11. Suppose that the mean number of hours spent playing sports per week for a random sample of 15 boys selected from the students at Los Osos Middle School was 2.3 hours.

a. If we were to select a second random sample of 15 boys from this school, do you think that the mean for this second sample would also be 2.3 hours? Explain why or why not.

b. Suppose a random sample of 15 girls was selected from this school, and that the mean number of hours spent playing sports for the girls in this sample was 2.2 hours. This sample mean is different from the sample mean for the boys. Explain why this does not necessarily mean that the mean number of hours spent playing sports for the population of *all* boys at the school is different than the mean for *all* of the girls at the school.

12. For each of the following, indicate whether you would conclude that the population means might be equal or if you would conclude that the population means are probably different. Justify your choice.

a. Population 1: seventh graders at Los Osos Middle School
Population 2: eighth graders at Los Osos Middle School
Variable of interest: Hours spent on Facebook per week
The two populations both have a MAD of about: 0.5 hours
Sample size: 30
Mean of random sample from population 1: 1.3 hours
Mean of random sample from population 2: 1.6 hours

b. Population 1: seventh graders at Los Osos Middle School
Population 2: ninth graders at Los Osos High School
Variable of interest: Hours spent on homework per week
The two populations both have a MAD of about: 2 hours
Sample size: 20
Mean of random sample from population 1: 4.3 hours
Mean of random sample from population 2: 6.0 hours

My Notes

LESSON 26-3 PRACTICE

Use the following information to answer Items 13–19.

Suppose that you wanted to know if the average amount of time spent playing computer games and video games is different for seventh-grade girls and seventh-grade boys. A random sample of 10 girls from the population of U.S. Census at School seventh-grade girls was selected using the random sampler on the Census at School website. The numbers of hours spent playing computer and video games per week for these 10 girls were as follows:

20	2	0	0	5
1	3	2	14	4

A random sample of 10 boys from the population of U.S. Census at School seventh-grade boys was also selected. The numbers of hours spent playing computer and video games per week for these 10 boys were as follows:

5	3	1	12	3
2	5	4	5	21

13. **Model with mathematics.** Draw two dot plots of the sample data—one for the sample of girls and one for the sample of boys. Be sure to use the same numerical scale for both dot plots.

14. Based on the dot plots, does it look like there is about the same amount of variability in the phone data for the girls and the phone data for the boys?

My Notes

15. Calculate the sample mean and MAD for each of the two samples.

You can use the tables below to help organize your work for computing the MADs.

Girls	
Data Value	**Distance from the Mean**
20	
2	
0	
0	
5	
1	
3	
2	
14	
4	

Boys	
Data Value	**Distance from the Mean**
5	
3	
1	
12	
3	
2	
5	
4	
5	
21	

16. Average the two sample MADs to get a common estimate of the population MADs.

17. Express the difference in the two sample means in terms of the common MAD.

18. Based on the difference in sample means, do you think that the mean number of hours spent playing computer and video games for seventh-grade girls and the mean for seventh-grade boys are probably different, or do you think that they might be the same? Justify your answer.

19. Suppose the two sample sizes had been 20 instead of 10, but that the sample means and MADs were still the same as the ones you calculated. Would you have answered Item 18 differently? Explain why or why not.

ACTIVITY 26 PRACTICE

Write your answers on notebook paper or grid paper. Show your work.

Students participating in Census at School complete two online activities. In one activity, **reaction time** is measured by having the student click on a stop button as quickly as possible after the screen changes color. The computer measures the time between when the color changes and the student clicks. The second activity is a memory test. In this activity, the student uncovers pairs of pictures. If the pictures match, they stay uncovered. If the pictures don't match, they are covered up again. The computer measures how long it takes to uncover all of the pairs and records this as a **memory test score**. A student with a good memory will be able to complete this activity faster and would have a lower score than a student who is not as good at remembering what pictures they have seen.

If you would like to try out these online activities, you can find them at U.S. Census at School website (www.amstat.org/censusatschool/students.cfm)

1. The sample mean reaction time for a random sample of 15 U.S. Census at School seventh graders was 0.35 seconds. The sample mean for a random sample of 15 Japan Census at School seventh graders was 0.33 seconds. Explain why this does not tell us that the mean reaction time of *all* U.S. Census at School seventh graders is different from the mean reaction time of *all* Japan Census at School seventh graders.

2. Suppose that the MAD for the memory test score for the population of U.S. Census at School seventh graders is about 15. The MAD for Japan Census at School seventh graders is also about 15. For each of the following sample sizes and pairs of sample means, determine if you would conclude that the mean memory test score for U.S. Census at School seventh graders and the mean memory test score for Japan Census at School seventh graders might be the same.

   a. Sample size = 10
   U.S. sample mean = 45
   Japan sample mean = 49

   b. Sample size = 30
   U.S. sample mean = 45
   Japan sample mean = 49

   c. Sample size = 20
   U.S. sample mean = 45
   Japan sample mean = 59

   d. Sample size = 30
   U.S. sample mean = 58
   Japan sample mean = 40

Use the following information to answer Items 3–9.

Suppose that you wanted to know if the mean reaction time is different for U.S. Census at School seventh graders and New Zealand Census at School seventh graders. A random sample of 10 students from the population of U.S. Census at School seventh graders was selected using the random sampler on the Census at School website. The reaction times (in seconds) for these 10 students were:

0.37	0.38	0.36	0.31	0.46
0.27	0.30	0.34	0.37	0.36

A random sample of 10 students from the population of New Zealand Census at School seventh graders was also selected. The reaction times for these 10 students were:

0.38	0.34	0.36	0.37	0.34
0.34	0.31	0.38	0.33	0.45

3. Draw two dot plots of the sample data—one for the sample of U.S. seventh graders and one for the New Zealand sample. Be sure to use the same numerical scale for both dot plots.

4. Do the dot plots suggest that there is about the same amount of variability in the reaction time data for U.S. seventh graders and the reaction time data for New Zealand seventh graders?

5. Calculate the sample mean and MAD for each of the two samples. You can use the tables below to help organize your work for computing the MADs.

United States	
Data Value	**Distance from the Mean**
0.37	
0.38	
0.36	
0.31	
0.46	
0.27	
0.30	
0.34	
0.37	
0.36	

New Zealand	
Data Value	**Distance from the Mean**
0.38	
0.34	
0.36	
0.37	
0.34	
0.34	
0.31	
0.38	
0.33	
0.45	

6. Average the two sample MADs to get a common estimate of the population MADs.

7. Express the difference in the two sample means in terms of the common MAD.

8. Based on the difference in sample means, do you think that the mean reaction time for U.S. seventh graders and the mean for New Zealand seventh graders are probably different, or do you think that they might be the same? Justify your answer.

9. Suppose the two sample sizes had been 20 instead of 10, but that the sample means and MADs were still the same. Would you have answered Item 8 differently? Explain why or why not.

Use the following information to answer Items 10–15.

Suppose that you wanted to know whether the mean memory test score is different for U.S. Census at School seventh graders and Canada Census at School seventh graders. A random sample of 10 students from the population of U.S. Census at School seventh graders was selected using the random sampler on the Census at School website. The memory test scores for these 10 students were as follows:

42	34	48	37	53
40	51	37	45	43

A random sample of 10 students from the population of Canada Census at School seventh graders was also selected. The memory test scores for these 10 students were as follows:

38	32	44	39	39
43	38	28	32	37

10. Draw two dot plots of the sample data—one for the sample of U.S. seventh graders and one for the Canada sample. Be sure to use the same numerical scale for both dot plots.

11. Do the dot plots suggest that there is about the same amount of variability in the memory test score data for U.S. seventh graders and the reaction time data for Canada seventh graders?

12. Calculate the sample mean and MAD for each of the two samples. You can use the tables below to help organize your work for computing the MADs.

United States	
Data Value	**Distance from the Mean**
42	
34	
48	
37	
53	
40	
51	
37	
45	
43	

Canada	
Data Value	**Distance from the Mean**
38	
32	
44	
39	
39	
43	
38	
28	
32	
37	

13. Average the two sample MADs to get a common estimate of the population MADs.

14. Express the difference in the two sample means in terms of the common MAD.

15. Based on the difference in sample means, do you think that the mean memory test score for U.S. seventh graders and the mean for Canada seventh graders are probably different, or do you think that they might be the same? Justify your answer.

16. Suppose the two sample sizes had been 30 instead of 10, but that the sample means and MADs were still the same. Would you have answered Item 15 differently? Explain why or why not.

Use the five step method described on page 380 to answer Items 17 and 18.

17. A random sample of 10 girls and a random sample of 10 boys were selected from the population of U.S. Census at School seventh graders. The data on reaction times for the students in these samples is shown below. Based on these two samples, would you conclude that the population mean reaction time for U.S. Census at School seventh-grade girls is different from the mean for U.S. Census at School seventh-grade boys? Be sure to show all five steps in your answer.

Girls' Reaction Times				
0.36	0.38	0.44	0.42	0.36
0.27	0.32	0.32	0.33	0.33
Boys' Reaction Times				
0.31	0.27	0.28	0.34	0.33
0.22	0.27	0.32	0.31	0.31

18. A random sample of 10 girls and a random sample of 10 boys were selected from the population of U.S. Census at School seventh graders. The data on memory test scores for the students in these samples is shown below. Based on these two samples, would you conclude that the population mean memory test score for U.S. Census at School seventh-grade girls is different from the mean for U.S. Census at School seventh-grade boys? Be sure to show all five steps in your answer.

Girls' Memory Test Scores				
39	38	43	31	27
40	44	32	34	32
Boys' Memory Test Scores				
44	40	34	30	38
39	38	47	39	44

MATHEMATICAL PRACTICES

Construct Viable Arguments and Critique the Reasoning of Others

19. Explain why you cannot automatically conclude that two populations have different means just because random samples from the two populations have different sample means.

Write your answers on notebook paper or grid paper. Show your work.

1. The sample mean arm span (the distance from the middle finger on one hand to the middle finger on the other hand when the arms are extended, measured in cm) for a random sample of 15 U.S. Census at School seventh-grade girls was 155 cm. The sample mean for a random sample of 15 U.S. Census at School eighth-grade girls was 157 cm. Explain why this does not tell us that the mean arm span of *all* U.S. Census at School seventh-grade girls is different from the mean arm span time of *all* U.S. Census at School eighth-grade girls.

Use the following information to answer Items 2–8.

Suppose that you wanted to know whether the mean arm span is different for U.S. Census at School seventh-grade girls and U.S. Census at School seventh-grade boys. A random sample of 10 students from the population of U.S. Census at School seventh-grade girls was selected using the random sampler on the Census at School website. The arm spans (in cm) for these 10 students were as follows:

171	152	176	147	167
165	152	161	152	147

A random sample of 10 students from the population of U.S. Census at School seventh-grade boys was also selected. The arm spans for these 10 students were as follows:

176	168	176	155	154
160	170	172	179	190

2. Draw two dot plots of the sample data—one for the sample of U.S. seventh-grade girls and one for the sample of seventh-grade boys. Be sure to use the same numerical scale for both dot plots.
3. Do the dot plots suggest that there is about the same amount of variability in the arm span data for U.S. seventh-grade girls and the arm span data for U.S. seventh-grade boys?
4. Calculate the sample mean and MAD for each of the two samples. You can use the tables shown to help organize your work for computing the MADs.

Seventh-Grade Girls	
Data Value	**Distance from the Mean**
171	
152	
176	
147	
167	
165	
152	
161	
152	
147	

Seventh-Grade Boys	
Data Value	**Distance from the Mean**
176	
168	
176	
155	
154	
160	
170	
172	
179	
190	

5. Average the two sample MADs to get a common estimate of the population MADs.
6. Express the difference in the two sample means in terms of the common MAD.
7. Based on the difference in sample means, do you think that the mean arm span for U.S. seventh-grade girls and the mean for U.S. seventh-grade boys are probably different, or do you think that they might be the same? Justify your answer.
8. Suppose the two sample sizes had been 20 instead of 10, but that the sample means and MADs were still the same as the ones you calculated. Would you have answered Item 7 differently? Explain why or why not.

Scoring Guide	Exemplary	Proficient	Emerging	Incomplete
	The solution demonstrates these characteristics:			
Mathematics Knowledge and Thinking (Items 1, 2, 3, 4, 5, 6, 7, 8)	• Clear and accurate understanding of variability and mean absolute deviation (MAD).	• A functional understanding of variability and MAD.	• Partial understanding of variability and MAD.	• Inaccurate or incomplete understanding of variability and MAD.
Problem Solving (Items 7, 8)	• An appropriate and efficient strategy that results in a correct answer.	• A strategy that may include unnecessary steps but results in a correct answer.	• A strategy that results in some incorrect answers.	• No clear strategy when solving problems.
Mathematical Modeling / Representations (Items 2, 3, 4, 5, 6)	• Clear and accurate understanding of representing a sample with a dotplot, mean, and MAD.	• Correctly representing a sample with a dotplot, mean, and MAD.	• Partial understanding of representing a sample with a dotplot, mean, and MAD.	• Little or no understanding of representing a sample with a dotplot, mean, and MAD.
Reasoning and Communication (Items 1, 3, 7, 8)	• Precise use of appropriate math terms and language to explain variability, MAD, and conclusions drawn from the MAD.	• An adequate explanation of variability, MAD, and conclusions drawn from the MAD.	• A misleading or confusing explanation of variability, MAD, and conclusions drawn from the MAD.	• An incomplete or inaccurate explanation of variability, MAD, and conclusions drawn from the MAD.