Student Edition & Teacher Guide

IMKH

An IM K–12 Math™ Curriculum

iM v.360 by Kendall Hunt

PRINT SAMPLER

ALGEBRA

A1

iM Illustrative Mathematics CERTIFIED ®

An IM K–12 Math™ Curriculum

All-embracing, all-encompassing, and all-inclusive

IM® v.360, the new version of the IM K-12 Math curriculum has undergone significant upgrades, enhancements, and revisions based upon feedback from school leaders, teachers and students nationwide. This updated version introduces fresh activities, lessons, problems, and titles.

What is different with IM® v.360?

Upgrades to the K–5 curriculum include:

- *NEW!* Language Learning Goals, End of Unit Guidance, Checklist Guidance
- Strengthened representations of diverse cultures
- Revisions to the Course Guide content, Instructional Routines, and blackline masters
- 2 lessons added in Kinder for number writing/sense (previously found in centers but do direct lesson)
- More blackline masters included in SE so teachers don't need to copy and distribute (alleviates lift)
- Reviewing activities that could create stress (especially food/recipes - when scarcity is a real issue in urban districts)

Upgrades to the 6–12 curriculum include:

- NEW! Narrative Structures, Section-level Assessments (Checkpoints), Instructional Goals, and Teacher Reflection Questions
- Embedded guidance for building a classroom community
- Embedded Math Language Routines and revised Instructional Routine language, including for 5 Practices activities
- Revised context and activity launches to invite more students into the mathematics, including more representations of diverse cultures
- Revised lesson contexts to align with the California framework, including environmental literacy enhancements
- Unit Narratives being revised for accuracy, clarity, and length
- More guidance around BLM's which to laminate and reuse
- More blackline masters included in SE so teachers don't need to copy and distribute (alleviates lift)

ALGEBRA 1

Student Edition

UNIT 1

Book 1

Certified by Illustrative Mathematics®

ISBN 9798385123049

HS_vII

20240203

Table of Contents

UNIT 1 ONE-VARIABLE STATISTICS

UNIT 2 LINEAR EQUATIONS AND SYSTEMS

UNIT 3 TWO-VARIABLE STATISTICS

Getting to Know You

Let's work together to collect data and explore statistical questions.

1.1 Types of Questions

Use the four questions to make a group of two related questions. What property do the items in the group share that the others do not?

- Question A: How many potato chips are in this bag of chips?
- Question B: What is the typical number of chips in a bag of chips?
- Question C: What type of chips are these?
- Question D: What type of chips do students in this class prefer?

1.2 Representing Data About You and Your Classmates

Your teacher will assign you a set of 3 questions.

1. Write an additional question of interest that requires data collected from the class to answer.

2. For each of the 4 questions of interest, write a survey question that will help you collect data from the class that can be analyzed to answer the question of interest. Ask the 4 survey questions to 15 classmates, and record their responses to collect data. Then return your group.

Sec A

3. Summarize the data for each question in a sentence or two, and share the results with your group.

4. With your group, decide what the responses for the questions numbered 1 have in common. Then do the same for questions numbered 2 and 3.

5. Does the question you wrote fit best with the questions numbered 1, 2, or 3? Explain your reasoning.

responder's name	question 1 response	question 2 response	question 3 response	my question response

Are you ready for more?

1. Find a news article that uses numerical data to discuss a statistical question.
2. Find a news article that uses categorical data to discuss a statistical question.

Sec A

Lesson 1 Summary

Statistics is about using data to solve problems or make decisions. There are two types of data:

- **Numerical data** are expressed using numbers that can be put in order. For example, the question "How tall are the students in this class?" would involve measuring the height of each student, resulting in numerical data.
- **Categorical data** are expressed using characteristics. For example, the question "What brand of phones do people use?" would involve surveying several people, with their answers resulting in categorical data.

The question that you ask determines the type of data that you collect and whether or not there is *variability* in the data collected. In earlier grades, you learned that there is variability in a data set if not all of the values in the data set are the same. These are examples of **statistical questions** because they are answered by collecting data that have variability:

- "What is the average class size at this school?" would produce numerical data with some variability.
- "What are the favorite colors of students in this class?" would produce categorical data with some variability.

These are examples of **non-statistical questions** because they are answered by collecting data that does not vary:

- "How many students are on the roster for this class?" has only one possible answer. There is only one value in the data set, so there is no variability.
- "What color is this marker?" has only one answer. There is only one value in the data set, so there is no variability.

Glossary

- categorical data
- non-statistical question
- numerical data
- statistical question

Practice Problems

1 Write a question of interest for which you would expect to collect numerical data.

Sec A

2 Write a question of interest for which you would expect to collect categorical data.

3 Select **all** the statistical questions.

A. What is the typical amount of rainfall for the month of June in the Galapagos Islands?

B. How much did it rain yesterday at the Mexico City International Airport?

C. Why do you like to listen to music?

D. How many songs does the class usually listen to each day?

E. How many songs did you listen to today?

F. What is the capital of Canada?

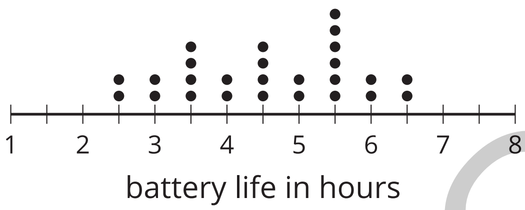
G. How long does it typically take for 2nd graders to walk a lap around the track?

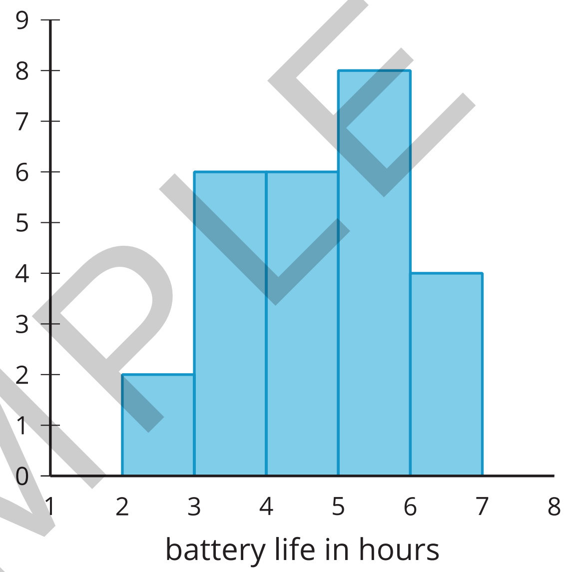
Data Representations

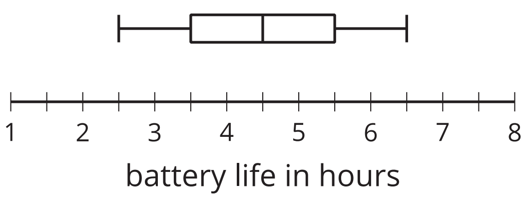
Let's represent and analyze data using dot plots, histograms, and box plots.

2.1 Notice and Wonder: Battery Life

The dot plot, histogram, and box plot summarize the hours of battery life for 26 cell phones that are constantly streaming video. What do you notice? What do you wonder?







2.2 Tomato Plants: Histogram

A histogram can be used to represent the distribution of numerical data.

1. The data represent the number of days it takes for different tomato plants to produce tomatoes. Use the information to complete the frequency table.

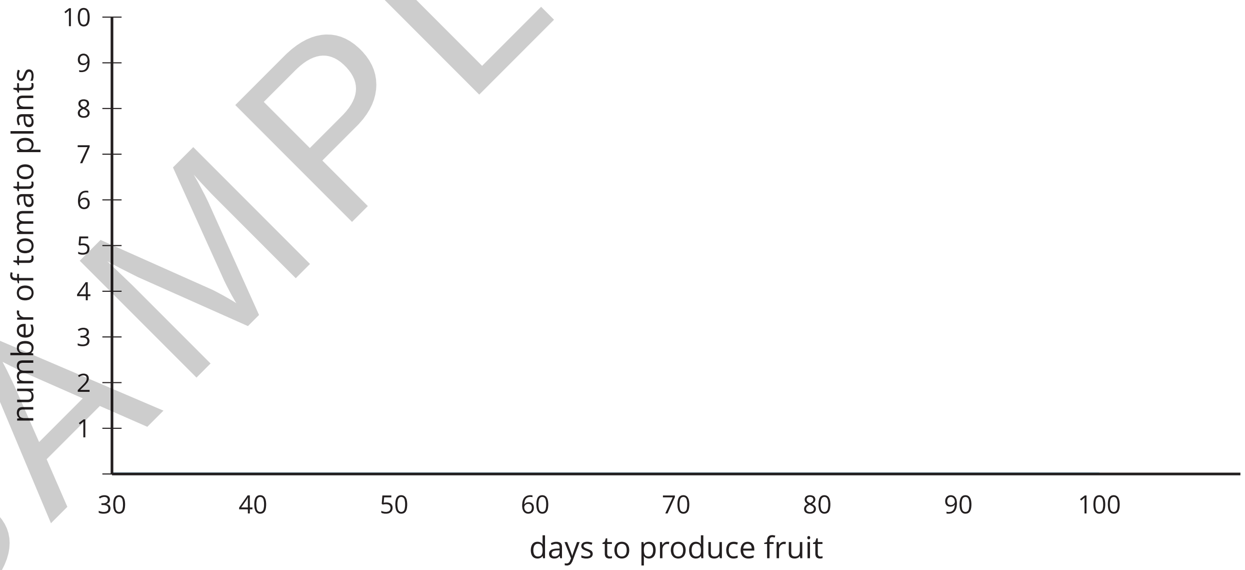
   47 52 53 55 57 60 61 62 63 65 65

   65 65 68 70 72 72 75 75 75 76 77

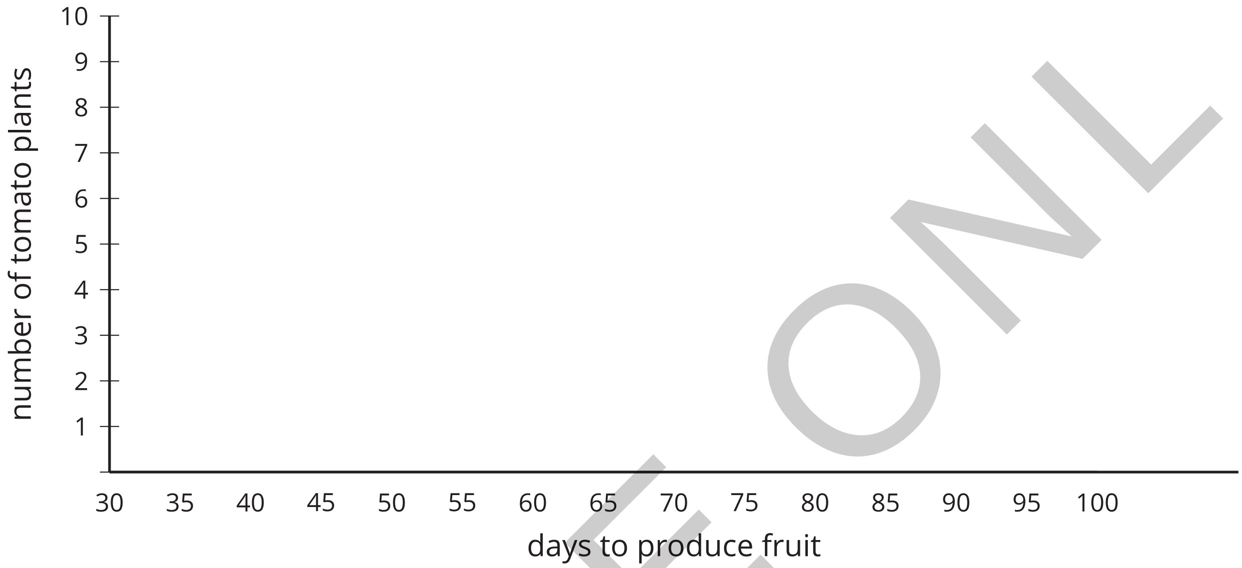
   78 80 81 82 85 88 89 90

days to produce fruit	frequency
40–50	
50–60	
60–70	
70–80	
80–90	
90–100	

2. Use the set of axes and the information in your table to create a histogram.



3. The histogram you created has intervals of width 10 (like 40–50 and 50–60). Use the set of axes and data to create another histogram with an interval of width 5. How does this histogram differ from the other one?



Are you ready for more?

It often takes some playing around with the interval widths to figure out which gives the best sense of the shape of the distribution.

1. What might be a problem with using interval widths that are too large?

2. What might be a problem with using interval widths that are too small?

3. What other considerations might go into choosing the width of an interval?

2.3 Tomato Plants: Box Plot

A box plot can also be used to represent the distribution of numerical data.

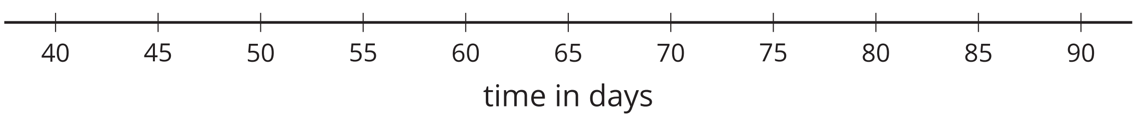
minimum	Q1	median	Q3	maximum

1. Using the same data as in the previous activity for tomato plants, find the median, and add it to the table. What does the median represent for these data?

2. Find the median of the least 15 values to split the data into the first and second quarters. This value is called the first quartile. Add this value to the table under Q1. What does this value mean in this situation?

3. Find the value (the third quartile) that splits the data into the third and fourth quarters, and add it to the table under Q3. Add the minimum and maximum values to the table.

4. Use the **five-number summary** to create a box plot that represents the number of days it takes for these tomato plants to produce tomatoes.



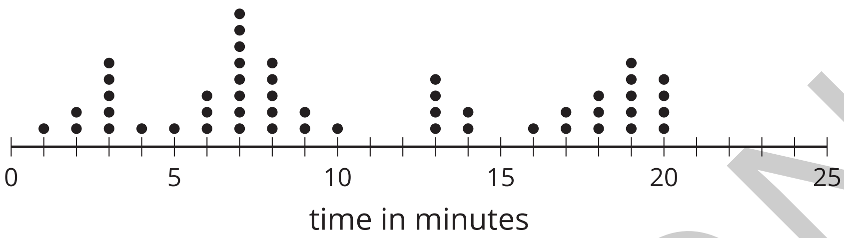
Lesson 2 Summary

The table shows a list of the number of minutes people could intensely focus on a task before needing a break. Fifty people of different ages are represented.

19 7 1 16 20 2 7 19 9 13 3 9 18 13 20 8 3 14 13 2

8 5 17 7 18 17 8 8 7 6 2 20 7 7 10 7 6 19 3 18 8

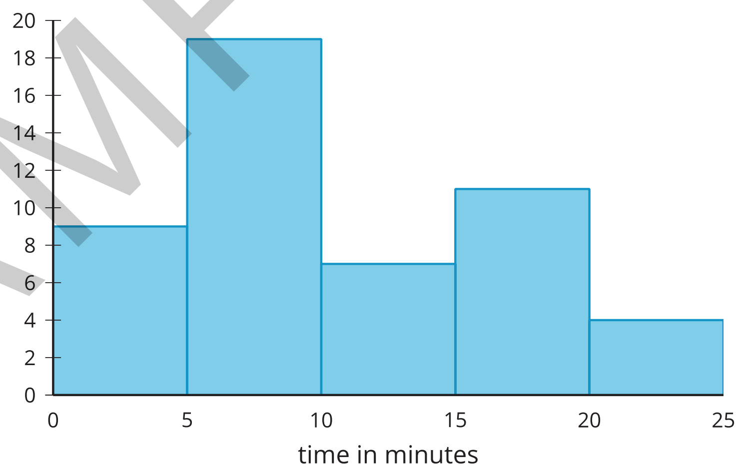
19 7 13 20 14 6 3 19 4



In a situation like this, it is helpful to represent the data graphically to better notice any patterns or other interesting features in the data. A dot plot can be used to see the shape and **distribution** of the data.

There were quite a few people that lost focus at around 3, 7, 13, and 19 minutes, and nobody lost focus at 11, 12, or 15 minutes. Dot plots are useful when the data set is not too large and shows all of the individual values in the data set. In this example, a dot plot can easily show all of the data. If the data set is very large (more than 100 values, for example), or if there are many different values that are not exactly the same, it may be hard to see all of the dots on a dot plot.

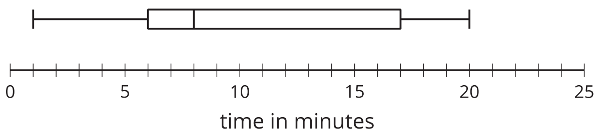
A histogram is another representation that shows the shape and distribution of the same data.



Most people lost focus between 5 and 10 minutes or between 15 and 20 minutes, while only 4 of the 50 people got distracted between 20 and 25 minutes. When creating histograms, each interval includes the number at the lower end of the interval but not the number at the upper end.

For example, the tallest bar displays values that are greater than or equal to 5 minutes but less than 10 minutes. In a histogram, values that are in an interval are grouped together. Although the individual values get lost with the grouping, a histogram can still show the shape of the distribution.

Here is a box plot that represents the same data.



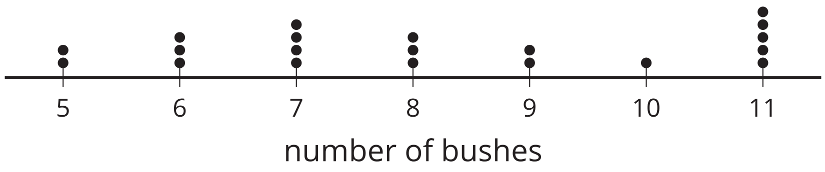
Box plots are created using a **five-number summary**. For a set of data, the five-number summary consists of these five statistics: the minimum value, the first quartile, the median, the third quartile, and the maximum value. These values split the data into four sections, each representing approximately one-fourth of the data. The median of this data is indicated at 8 minutes, and about 25% of the data fall in the short second quarter of the data between 6 and 8 minutes. Similarly, approximately one-fourth of the data are between 8 and 17 minutes. Like the histogram, the box plot does not show individual data values, but other features such as quartiles, range, and median are seen more easily. Dot plots, histograms, and box plots provide three different ways to look at the shape and distribution while highlighting different aspects of the data.

Glossary

- distribution
- five-number summary

Practice Problems

1 The dot plot displays the number of bushes in the yards for houses in a neighborhood. What is the median?



2 The data set represents the shoe sizes of 19 students in a fifth grade physical education class.

4 5 5 5 6 6 6 6 7 7 7 7 7.5 7.5 8 8 8.5 8.5 9

Create a box plot to represent the distribution of the data.

3 The data set represents the number of pages in the last book read by each of 20 students over the summer.

163 170 171 173 175 205 220 220 220 253 267 281 305

305 305 355 371 388 402 431

Create a histogram to represent the distribution of the data.

4

from Unit 1, Lesson 1

Each set of data was collected from surveys to answer statistical questions. Select **all** of the data sets that represent numerical data.

A. {1, 1.2, 1.4, 1.4, 1.5, 1.6, 1.8, 1.9, 2, 2, 2.1, 2.5}

B. {Red, Red, Yellow, Yellow, Blue, Blue, Blue}

C. {45, 60, 60, 70, 75, 80, 85, 90, 90, 100, 100, 100}

D. {-7, -5, -3, -1, -1, -1, 0}

E. {98.2, 98.4, 98.4, 98.6, 98.6, 98.6, 98.6, 98.7, 98.8, 98.8}

F. {Yes, Yes, Yes, Yes, Maybe, Maybe, No, No, No}

G. {A, A, A, B, B, B, C, C, C}

5

from Unit 1, Lesson 1

Is "What is the typical distance a moped can be driven on a single tank of gas?" a statistical question? Explain your reasoning.

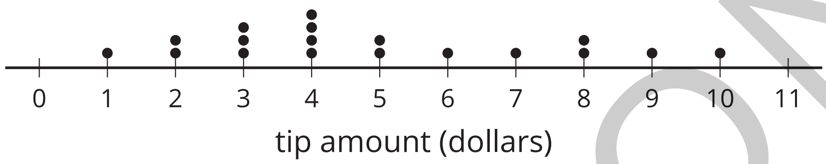
A Gallery of Data

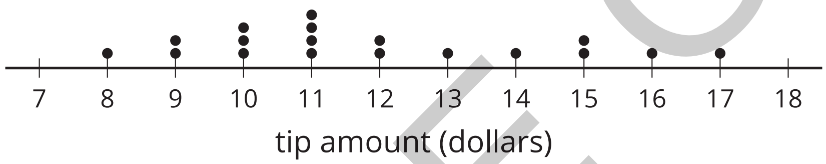
Let's make, compare, and interpret data displays.

3.1 Notice and Wonder: Dot Plots

The dot plots represent the distribution of the amount of tips, in dollars, left at 2 different restaurants on the same night.

What do you notice? What do you wonder?





3.2 Data Displays

Your teacher will assign your group a statistical question. As a group:

1. Create a dot plot and a box plot to display the distribution of the data.

2. Write 3 comments that interpret the data. Pause here for instructions on visiting the other displays.

3. Visit each display and leave a note or question about the information in the display.

Are you ready for more?

Choose one of the more interesting questions that you or a classmate asked, and collect data from a larger group—such as more students from the school. Create a data display, and compare results from the data collected in class.

Lesson 3 Summary

We can represent a distribution of data in several different forms, including lists, dot plots, histograms, and box plots. A list displays all of the values in a data set and can be organized in different ways. This list shows the pH for 30 different water samples.

5.9 7.6 7.5 8.2 7.6 8.6 8.1 7.9 6.1 6.3 6.9 7.1 8.4 6.5 7.2 6.8

7.3 8.1 5.8 7.5 7.1 8.4 8.0 7.2 7.4 6.5 6.8 7.0 7.4 7.6

Here is the same list organized in order from least to greatest.

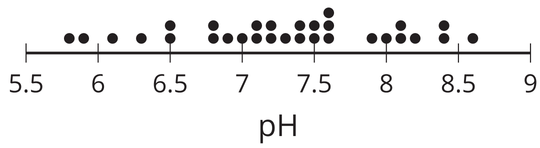
5.8 5.9 6.1 6.3 6.5 6.5 6.8 6.8 6.9 7.0 7.1 7.1 7.2 7.2 7.3 7.4

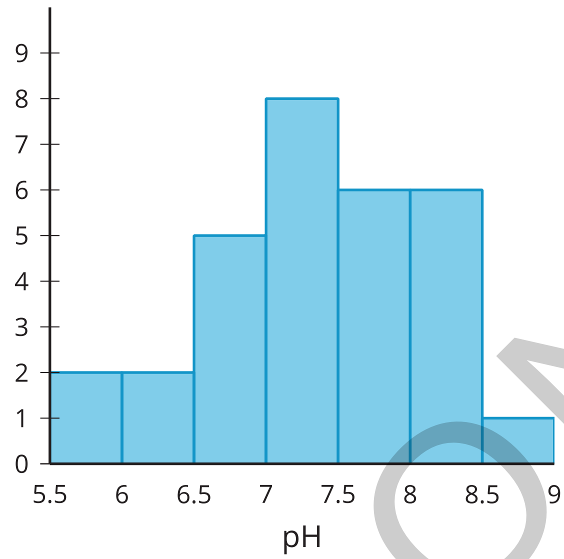
7.4 7.5 7.5 7.6 7.6 7.6 7.9 8.0 8.1 8.1 8.2 8.4 8.4 8.6

With the list organized, you can more easily:

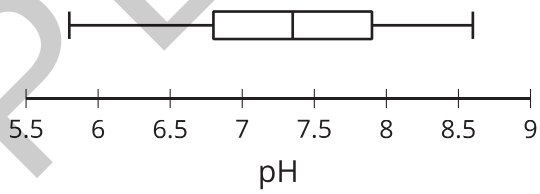
- Interpret the data
- Calculate the values of the five-number summary
- Estimate or calculate the mean
- Create a dot plot, box plot, or histogram

Here are a dot plot and histogram representing the distribution of the data in the list.





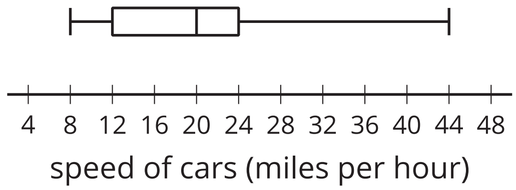
A dot plot is created by putting a dot for each value above the position on a number line. For the pH dot plot, there are 2 water samples with a pH of 6.5 and 1 water sample with a pH of 7. A histogram is made by counting the number of values from the data set in a certain interval and drawing a bar over that interval at a height that matches the count. In the pH histogram, there are 5 water samples that have a pH between 6.5 and 7 (including 6.5, but not 7). Here is a box plot representing the distribution of the same data as represented by the dot plot and histogram.



To create a box plot, you need to find the minimum, first quartile, median, third quartile, and maximum values for the data set. These 5 values are sometimes called the *five-number summary*. Drawing a vertical mark and then connecting the pieces as in the example creates the box plot. For the pH box plot, we can see that the minimum is about 5.8, the median is about 7.4, and the third quartile is around 7.9.

Practice Problems

1 The box plot represents the distribution of speeds, in miles per hour, of 100 cars as they passed through a busy intersection.



a. What is the smallest value in the data set? Interpret this value in the situation.

b. What is the largest value in the data set? Interpret this value in the situation.

c. What is the median? Interpret this value in the situation.

d. What is the first quartile (Q1)? Interpret this value in the situation.

e. What is the third quartile (Q3)? Interpret this value in the situation.

2 The data set represents the number of eggs produced by a small group of chickens each day for ten days.

7 7 7 7 7 8 8 8 8 9

Select **all** the values that could represent the typical number of eggs produced in a day.

A. 7.5 eggs

B. 7.6 eggs

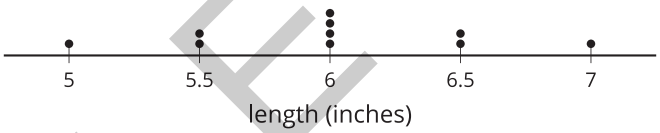
C. 7.7 eggs

D. 8 eggs

E. 9 eggs

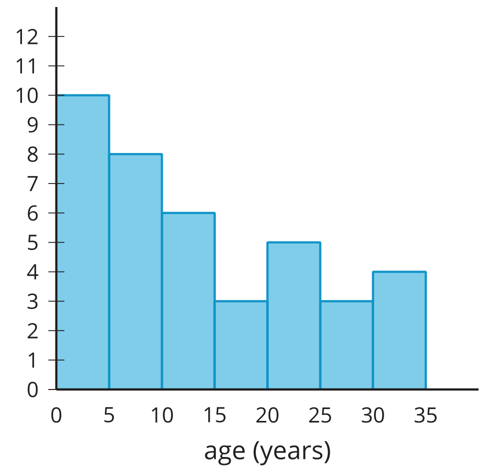
3 from Unit 1, Lesson 2

The dot plot displays the lengths of pencils (in inches) used by students in a class. What is the mean?



4 from Unit 1, Lesson 2

The histogram represents ages of 39 people at a store that sells children's clothes. Which interval contains the median?



A. The interval from 0 to 5 years.

B. The interval from 5 to 10 years.

C. The interval from 10 to 15 years.

D. The interval from 15 to 20 years.

5 from Unit 1, Lesson 2

The data set represents the responses, in degrees Fahrenheit, collected to answer the question “How hot is the sidewalk during the school day?”.

92 95 95 95 98 100 100 100 103 105 105 111 112 115

115 116 117 117 118 119 119 119 119 119 119

a. Create a dot plot to represent the distribution of the data.

b. Create a histogram to represent the distribution of the data.

c. Which display gives you a better overall understanding of the data? Explain your reasoning.

6 from Unit 1, Lesson 1

Is “What is the area of the floor in this classroom?” a statistical question? Explain your reasoning.

Unit 1, Lesson 4

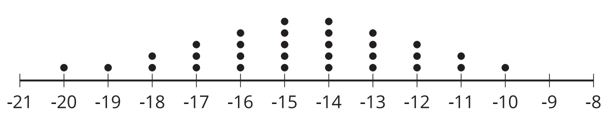
The Shape of Distributions

Let's explore data and describe distributions.

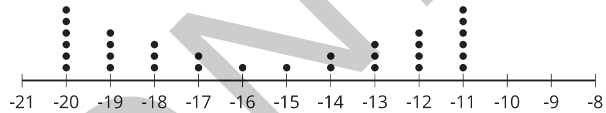
4.1 Which Three Go Together: Distribution Shape

Which three go together? Why do they go together?

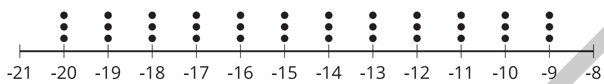
A.



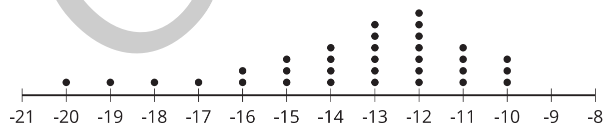
B.



C.



D.



4.2 Card Sort: Matching Distributions

Take turns with your partner matching 2 different data displays that represent the distribution of the same set of data.

1. For each set that you find, explain to your partner how you know it's a match.
2. For each set that your partner finds, listen carefully to their explanation. If you disagree, discuss your thinking and work to reach an agreement.
3. When finished with all ten matches, describe the shape of each distribution.

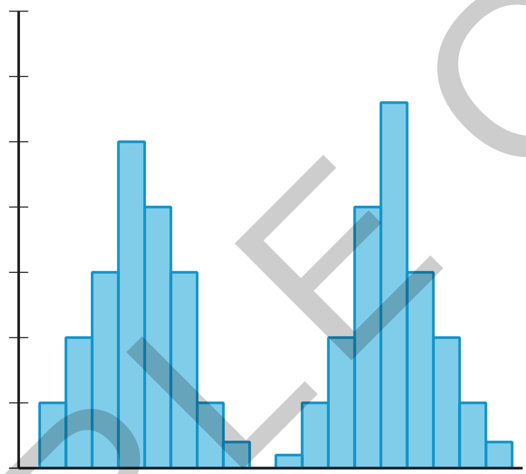
4.3 Where Did The Distribution Come From?

Your teacher will assign you some of the matched distributions. Using the information provided in the data displays, make an educated guess about the question that produced this data. Be prepared to share your reasoning.

Sec B

Are you ready for more?

This distribution shows the length in inches of fish caught and released from a nearby lake.

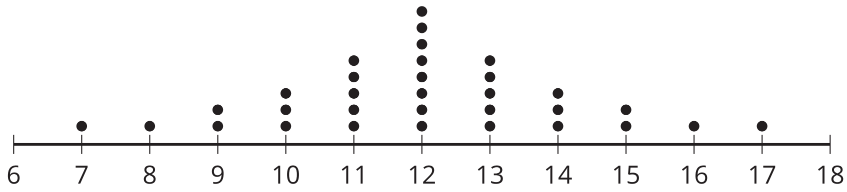


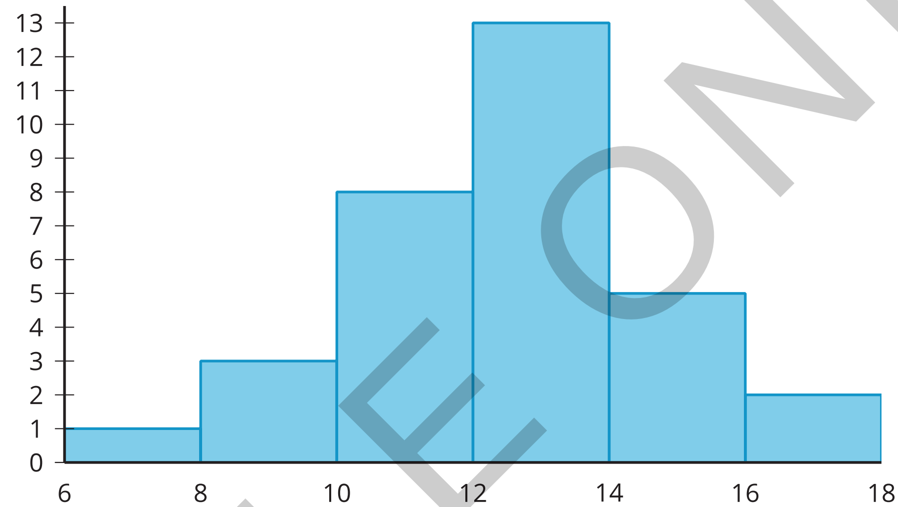
1. Describe the shape of the distribution.

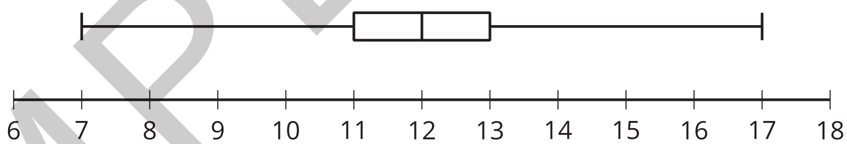
2. Make an educated guess about what could cause the distribution to have this shape.

Lesson 4 Summary

We can describe the shape of distributions as *symmetric*, *skewed*, *bell-shaped*, *bimodal*, or *uniform*. The dot plot, histogram, and box plot shown here represent the distribution of the same data set. This data set has a symmetric distribution.

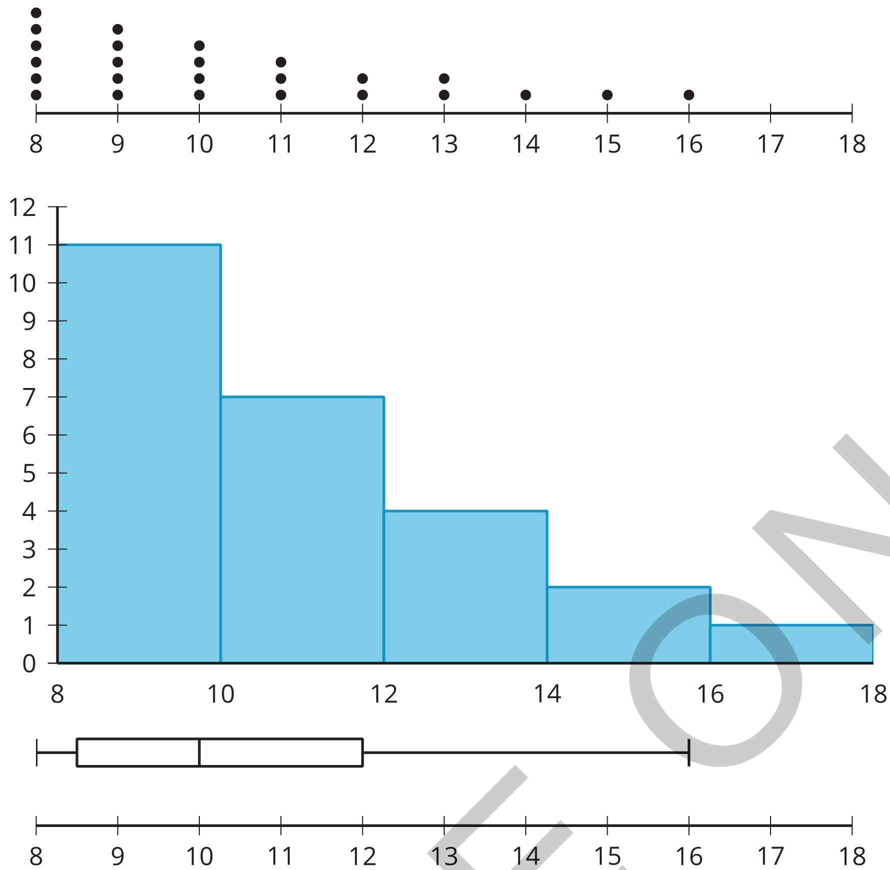






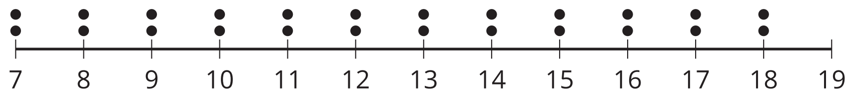
In a **symmetric distribution**, the mean is equal to the median and there is a vertical line of symmetry in the center of the data display. The histogram and the box plot both group data together. Since histograms and box plots do not display each data value individually, they do not provide information about the shape of the distribution to the same level of detail that a dot plot does. This distribution, in particular, can also be called bell-shaped. A **bell-shaped distribution** has a dot plot that takes the form of a bell with most of the data clustered near the center and fewer points farther from the center. This makes the measure of center a very good description of the data as a whole. Bell-shaped distributions are always symmetric or close to it.

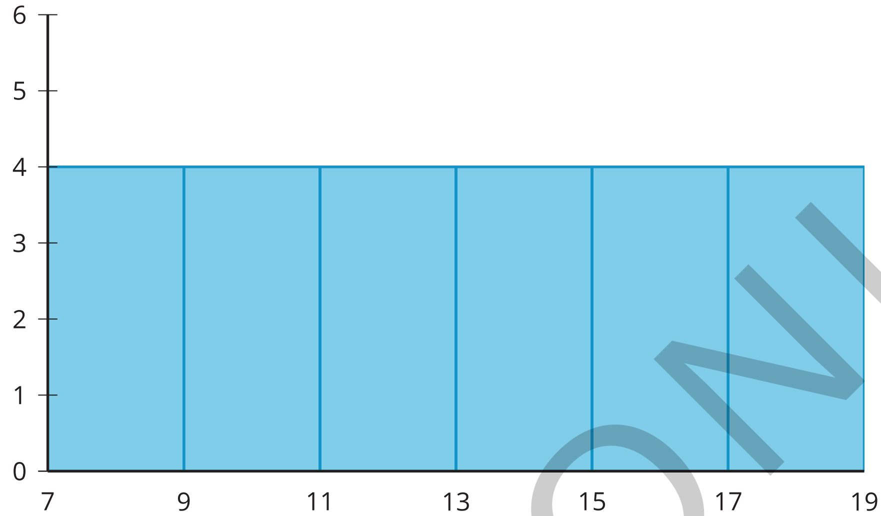
The dot plot, histogram, and box plot shown here represent a skewed distribution.

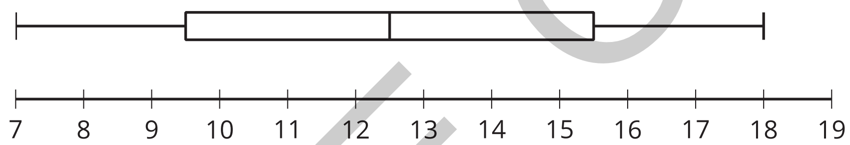


In a **skewed distribution**, one side of the distribution has more values farther from the bulk of the data than the other side has. This results in the mean and median not being equal. In this skewed distribution, the data are skewed to the right because most of the data are near the 8 to 10 interval, but there are many points to the right. The mean is greater than the median. The large data values to the right cause the mean to shift in that direction while the median remains with the bulk of the data. So, the mean is greater than the median for distributions that are skewed to the right. In a data set that is skewed to the left, a similar effect happens but to the other side. Again, the dot plot provides a greater level of detail about the shape of the distribution than do either the histogram or the box plot.

A **uniform distribution** has the data values evenly distributed throughout the range of the data. This causes the distribution to look like a rectangle.

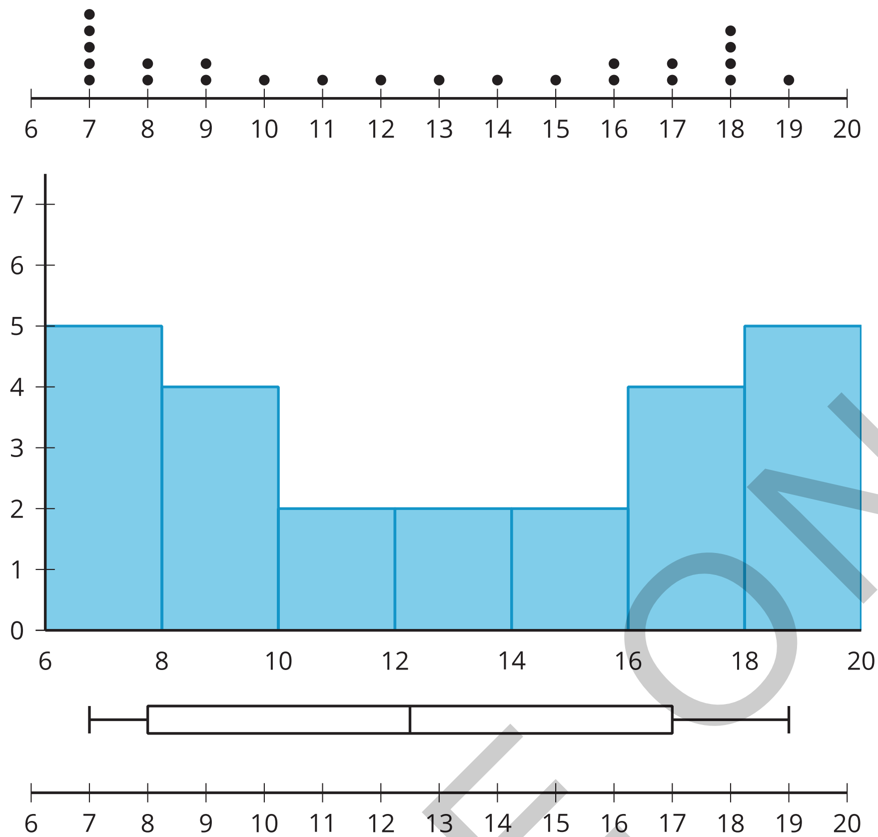






In a uniform distribution, the mean is equal to the median because a uniform distribution is also a symmetric distribution. The box plot does not provide enough information to describe the shape of the distribution as uniform, though the even length of each quarter does suggest that the distribution may be approximately symmetric.

A **bimodal distribution** has two very common data values seen in a dot plot or histogram as distinct peaks.



Sometimes, a bimodal distribution has most of the data clustered in the middle of the distribution. In these cases, the center of the distribution does not describe the data very well. Bimodal distributions are not always symmetric. For example, the peaks may not be equally spaced from the middle of the distribution, or other data values may disrupt the symmetry.

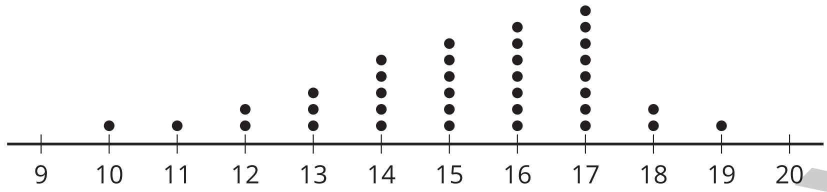
Glossary

- bell-shaped distribution
- bimodal distribution
- skewed distribution
- symmetric distribution
- uniform distribution

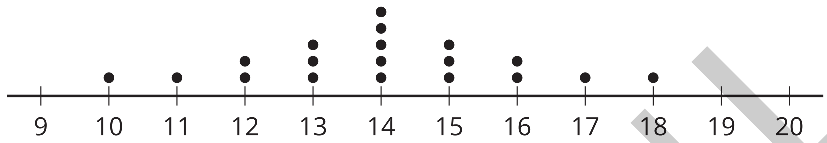
Practice Problems

1 Which of the dot plots shows a symmetric distribution?

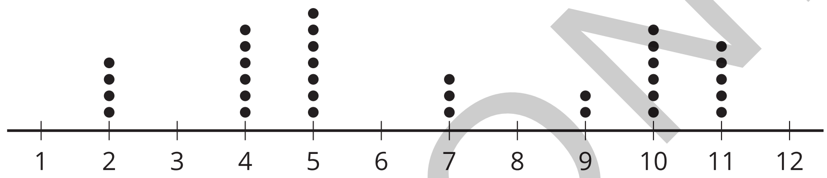
A.



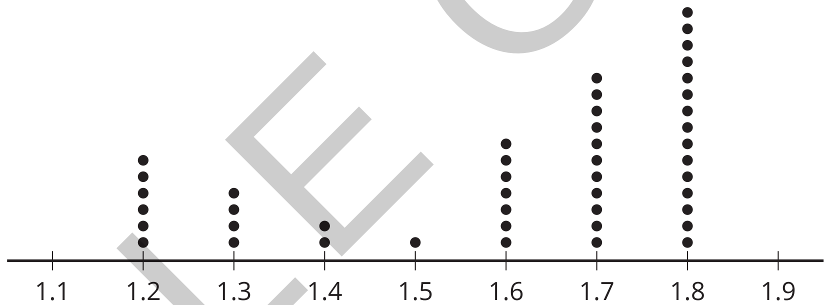
B.



C.

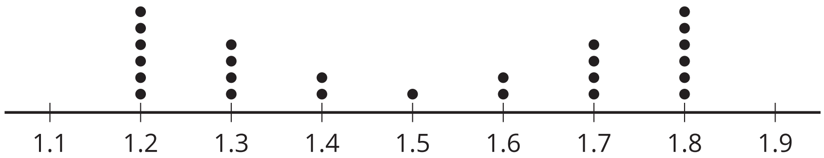


D.

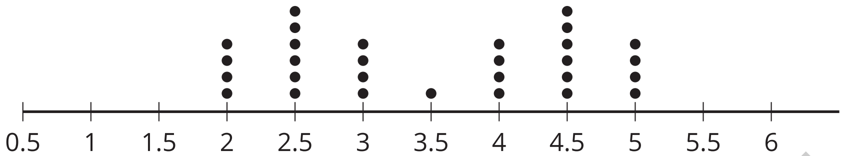


2 Which of the dot plots shows a skewed distribution?

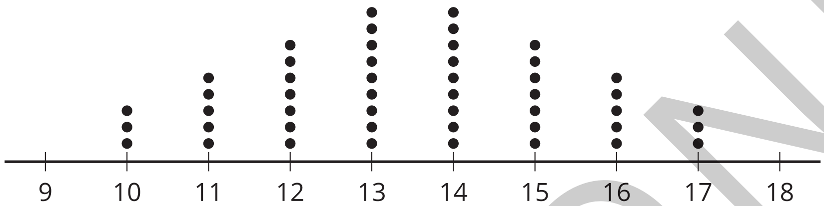
A.



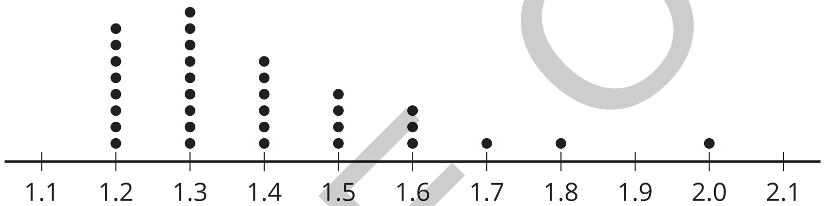
B.



C.



D.



3 Create a dot plot showing a uniform distribution.

4 from Unit 1, Lesson 3

The data represent the number of ounces of water that 26 students drank the day before a test at school:

8 8 8 16 16 16 32 32 32 32 32 32 64 64 64 64 64

64 64 80 80 80 80 88 88 88

a. Create a dot plot for the data.

b. Create a box plot for the data.

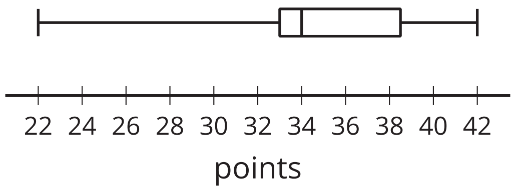
c. What information about the data is provided by the box plot that is not provided by the dot plot?

d. What information about the data is provided by the dot plot that is not provided by the box plot?

e. It was recommended that students drink 48 or more ounces of water. How could you use a histogram to easily display the number of students who drank the recommended amount?

5 from Unit 1, Lesson 2

The box plot represents the distribution of the number of points scored by a cross country team at 12 meets.



a. If possible, find the mean. If not possible, explain why not.

b. If possible, find the median. If not possible, explain why not.

c. Did the cross country team ever score 30 points at a meet?

Unit 1, Lesson 5

Calculating Measures of Center and Variability

Let's calculate measures of center and measures of variability and know which are most appropriate for the data.

5.1 Calculating Centers

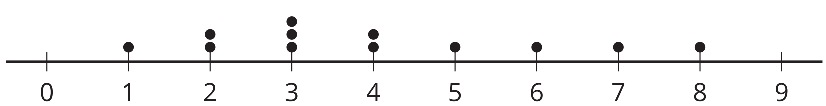
Decide if each situation is true or false. Explain your reasoning.

1. The mean can be found by adding all the numbers in a data set and dividing by the number of numbers in the data set.

2. The mean of the data in the dot plot is 4.

3. The median of the data set is 9 for the data: 4, 5, 9, 1, 10.

4. The median of the data in the dot plot is 3.5.



5.2 Heartbeats: Part 1

The heart rates of eight high school students are listed in beats per minute:

72 75 81 76 76 77 79 78

1. What is the interquartile range?

2. How many values in the data set are:
   a. less than Q1?
   b. between Q1 and the median?
   c. between the median and Q3?
   d. greater than Q3?

3. A pod of dolphins contains 800 dolphins of various ages and lengths. The median length of dolphins in this pod is 5.8 feet. What information does this tell you about the length of dolphins in this pod?

4. The same vocabulary test with 50 questions is given to 600 students from fifth to tenth grades and the number of correct responses is collected for each student in this group. The interquartile range is 40 correct responses. What information does this tell you about the number of correct responses for students taking this test?

5.3 Heartbeats: Part 2

1. Calculate the MAD using the same data from the previous activity by finding the average distance from each data value to the mean. You may find it helpful to organize your work by completing the table provided.

data values	mean	deviation from the mean (data value - mean)	absolute deviation |deviation|
72			
75			
81			
76			
76			
77			
79			
78			

MAD:

2. For another data set, all of the values are either 3 beats per minute above the mean or 3 beats per minute below the mean. Is that enough information to find the MAD for this data set? If so, find the MAD. If not, what other information is needed? Explain your reasoning.

3. Several pennies are placed along a meter stick, and the position in centimeters of each penny is recorded. The mean position is the 50 centimeter mark and the MAD is 10 centimeters. What information does this tell you about the position of the pennies along the meter stick?

Are you ready for more?

Suppose there are 6 pennies on a meter stick so that the mean position is the 50 centimeter mark and the MAD is 10 centimeters.

1. Find possible locations for the 6 pennies.

2. Find a different set of possible locations for the 6 pennies.

Lesson 5 Summary

The *mean absolute deviation*, or MAD, and the *interquartile range*, or IQR, are measures of variability. Measures of variability tell you how much the values in a data set tend to differ from one another. A greater measure of variability means that the data are more spread out, while a smaller measure of variability means that the data are more consistent and are closer to the measure of center.

To calculate the MAD of a data set:

1. Find the mean of the values in the data set.
2. Find the distance between each data value and the mean (on the number line):

$$|\text{data value} - \text{mean}|$$

3. Find the mean of the distances. This value is the MAD.

To calculate the IQR, subtract the value of the first quartile from the value of the third quartile. Recall that the first and third quartile are included in the five-number summary.

Practice Problems

1 The data set represents the number of errors on a typing test.

5 6 8 8 9 9 10 10 10 12

a. What is the median? Interpret this value in the situation.

b. What is the IQR?

2 The data set represents the heights, in centimeters, of ten model bridges made for an engineering competition.

13 14 14 16 16 16 16 18 18 19

a. What is the mean?

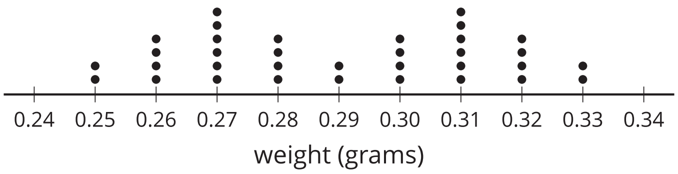
b. What is the MAD?

3 from Unit 1, Lesson 4

Describe the shape of the distribution shown in the dot plot. The dot plot displays the golf scores from a golf tournament.

4 from Unit 1, Lesson 4

The dot plot shows the weight, in grams, of several different rocks. Select **all** the terms that describe the shape of the distribution.



A. bell-shaped

B. bimodal

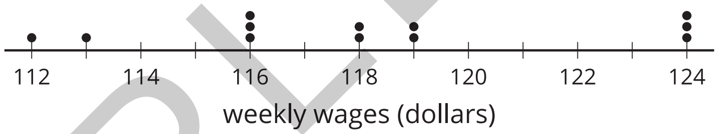
C. skewed

D. symmetric

E. uniform

5 from Unit 1, Lesson 3

The dot plot represents the distribution of wages earned during a one-week period by 12 college students.



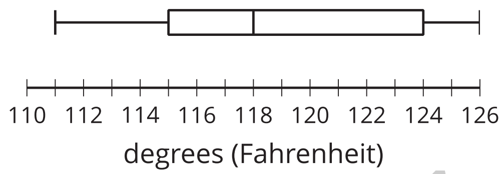
a. What is the mean? Interpret this value based on the situation.

b. What is the median? Interpret this value based on the situation.

c. Would a box plot of the same data have allowed you to find both the mean and the median?

6 from Unit 1, Lesson 2

The box plot displays the temperature of saunas in degrees Fahrenheit. What is the median?



Unit 1, Lesson 6

Mystery Computations

Let's explore spreadsheets.

6.1 Make 24

Your teacher will give you 4 numbers. Use these numbers, along with mathematical operations like addition and multiplication, to make 24.

6.2 Mystery Operations

Navigate to this activity in the digital version of the materials, or to ggbm.at/fjcybyqf.

Input different numbers in column A, and try to predict what will happen in column B. (Do not change anything in column B.)

1. How is the number in cell B2 related to all or some of the numbers in cells A2, A3, A4, and A5?

2. How is the number in cell B3 related to all or some of the numbers in cells A2, A3, A4, and A5?

3. How is the number in cell B4 related to all or some of the numbers in cells A2, A3, A4, and A5?

4. How is the number in cell B5 related to all or some of the numbers in cells A2, A3, A4, and A5?

6.3 More Spreadsheets!

Navigate to this activity in the digital version of the materials or to ggbm.at/wu9t7kkd.

1. Change the spreadsheet so that B2 contains =A2+A4. To edit the formula in B2, you may have to click it twice.
2. Change the numbers in A2 through A5. Make sure that your new formula does what it is supposed to do by doing a mental calculation and checking the result in B2.
3. Change the contents of B3 so that B3 does something different.
4. Before trading with a partner, make sure your new formula is not visible by clicking in a different cell.
5. Trade with your partner.
6. Change the numbers in Column A to try and figure out your partner's new rule.

Lesson 6 Summary

Spreadsheets are useful mathematical and statistical tools. Here is an example of a spreadsheet. Each cell in the spreadsheet can be named with its column and row. For example, cell B2 contains the value 99. Cell A4 contains the value -17. Cell D1 is selected.

	A	B	C	D	E
1					
2		99			
3					
4	-17				
5				0.25	
6					
7					

It is possible for the value in a cell to depend on the value in other cells. Let's type the formula $= B2 - D5$ into cell D1.

	A	B	C	D	E
1				=B2-D5	
2		99			
3					
4	-17				
5				0.25	
6					
7					

When we press enter, D1 will display the result of subtracting the number in cell D5 from the number in cell B2.

	A	B	C	D	E
1				98.75	
2		99			
3					
4	-17				
5				0.25	
6					
7					

If we type new numbers into B2 or D5, the number in D1 will automatically change.

	A	B	C	D	E
1				79	
2		99			
3					
4	-17				
5				20	
6					
7					

Practice Problems

1 What could be the formula used to compute the value shown in cell B3?

	A	B
1	change these	what happens here?
2	7	20
3	0	350
4	13	0
5	50	69
6	-1	

A. = B3 * B4

B. = A2 + A5

C. = A2 * A5

D. = Sum(A2:A6)

2 What number will appear in cell B2 when the user presses Enter?

	A	B
1	change these	what happens here?
2	10	=Sum(A3:A5)
3	5	
4	0	
5	-7	

Sec C

3 Select **all** the formulas that could be used to calculate the value in cell B4.

	A	B
1	change these	what happens here?
2	7	20
3	0	350
4	13	0
5	50	69
6	-1	

A. =Product(A2:A6)

B. =Sum(A2:A6)

C. = A2 + A3

D. = A2 * A3

E. = A3 * A4 * A5

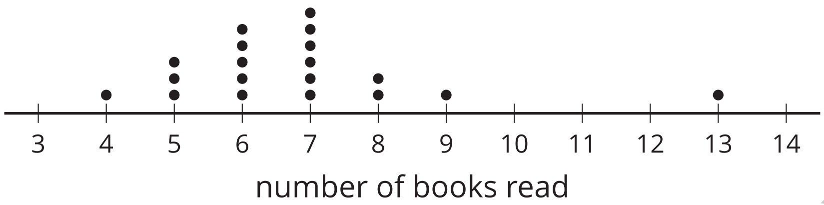
F. = A3 + A4 + A5

4 The formula in cell B2 is = Product(A2 : A5). Describe a way to change the contents of column A so that the value in cell B2 becomes -70.

	A	B
1	change these	what happens here?
2	10	0
3	5	
4	0	
5	-7	

5 from Unit 1, Lesson 5

The dot plot displays the number of books read by students during the semester.

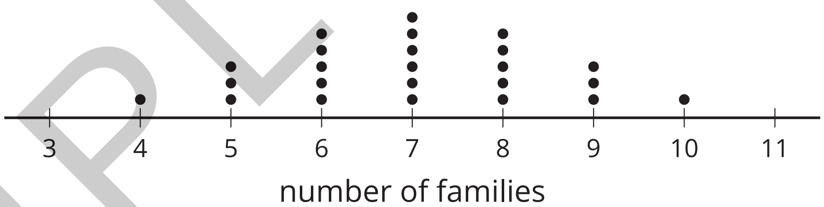


a. Which measure of center would you use given the shape of the distribution in the dot plot? Explain your reasoning.

b. Which measure of variability would you use? Explain your reasoning.

6 from Unit 1, Lesson 5

The dot plot displays the number of families living in different blocks of a town.



a. Which measure of center would you use, given the shape of the distribution in the dot plot? Explain your reasoning.

b. Which measure of variability would you use? Explain your reasoning.

Spreadsheet Computations

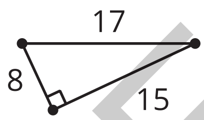
Let's use spreadsheets as calculators.

7.1 Dust Off Those Cobwebs

1. A person walks 4 miles per hour for 2.5 hours. How far does the person walk?

2. A rectangle has an area of 24 square centimeters. What could be its length and width?

3. What is the area of this triangle?



7.2 A Spreadsheet Is a Calculator

Use a spreadsheet to compute each of the following. Type each computation in a new cell, instead of erasing a previous computation.

- $2 + 7$
- $2 - 7$
- $7 \cdot 2$
- 7^2
- $7 \div 2$
- $\frac{1}{7}$ of 91
- $0.1 \cdot 2 + 3$
- $0.1(2 + 3)$
- $13 \div \frac{1}{7}$
- The average of 2, 7, 8, and 11

7.3 Use the Contents of a Cell in a Calculation

1. Type any number in cell A1, and another number in cell A2. Then in cell A3, type =A1+A2. What happens?
2. In cell A4, compute the product of the numbers in A1 and A2.
3. In cell A5, compute the number in A1 raised to the power of the number in A2.
4. Now, type a new number in cell A1. What happens?
5. Type a new number in cell A2. What happens?
6. Use nearby cells to label the contents of each cell. For example, in cell B3, type "the sum of A1 and A2." (This is a good habit to get into. It will remind you and anyone else using the spreadsheet what each cell means.)

7.4 Solve Some Problems

For each problem:

- Estimate the answer before calculating anything.
- Use the spreadsheet to calculate the answer.
- Write down the answer and the formula that you used in the spreadsheet to calculate it.

1. The speed limit on a highway is 110 kilometers per hour. How much time does it take a car to travel 132 kilometers at this speed?

2. In a right triangle, the lengths of the sides that make a right angle are 98.7 cm and 24.6 cm. What is the area of the triangle?

3. A recipe for fruit punch uses 2 cups of seltzer water, $\frac{1}{4}$ cup of pineapple juice, and $\frac{2}{3}$ cup of cranberry juice. How many cups of fruit punch are in 5 batches of this recipe?

4. Check in with a partner, and resolve any discrepancies with your answer to the last question. Next, type 2, $\frac{1}{4}$, $\frac{2}{3}$, and 5 in separate cells. (You may find it helpful to label cells next to them with the meaning of each number.) In a blank cell, type a formula for the total amount of fruit punch that uses the values in the other four cells. Now you should be able to easily figure out:
   a. How much in 7.25 batches?
   b. How much in 5 batches if you change the recipe to 1.5 cups of seltzer water per batch?
   c. Change the ratio of the ingredients in the fruit punch so that you would like the flavor. How many total cups are in $\frac{1}{2}$ batch?

Lesson 7 Summary

A spreadsheet can be thought of as a type of calculator. For example, in a cell, you could type $= 2 + 3$, and then the sum of 5 is displayed in the cell. You can also perform operations on the values in other cells. For example, if you type a number in A1 and a number in A2, and then in A3 type $= A1 + A2$, cell A3 will display the sum of the values in cells A1 and A2.

Familiarize yourself with how your spreadsheet software works on your device.

- On some spreadsheet programs, an = symbol must be typed before the expression in the cell. (On others, it does not matter if your expression begins with =.)
- Know how to "submit" the expression so the computation takes place. If your device has a keyboard, it's likely the Enter key. On a touchscreen device, you may have to tap a check mark.
- Learn symbols to use for various operations, and how to find them on your keyboard. Here are the symbols used for some typical operations:
    - + for add
    - - for subtract or for a negative number (this symbol does double duty in most spreadsheets)
    - * for multiply
    - / for divide
    - a / b for the fraction $\frac{a}{b}$
    - ^ for an exponent
    - . for a decimal point
    - ( ) to tell it what to compute first (often needed around fractions)

Practice Problems

1 Write a formula that you could type into a spreadsheet to compute the value of each expression.

a. $(19.2) \cdot 73$

b. 1.1^5

c. $2.34 \div 5$

d. $\frac{91}{7}$

2 A long-distance runner jogs at a constant speed of 7 miles per hour for 45 minutes. Which spreadsheet formula would give the distance she traveled?

A. = 7 * 45

B. = 7 / 45

C. = 7 * (3 / 4)

D. = 7 / (3 / 4)

3 In a right triangle, the lengths of the sides that make a right angle are 3.4 meters and 5.6 meters. Select **all** the spreadsheet formulas that would give the area of this triangle.

A. = 3.4 * 5.6

B. = 3.4 * 5.6 * 2

C. = 3.4 * 5.6 / 2

D. = 3.4 * 5.6 * (1/2)

E. = (3.4 * 5.6) / 2

4 This spreadsheet should compute the total ounces of sparkling grape juice based on the number of batches, ounces of grape juice in a single batch, and ounces of sparkling water in a single batch.

	A	B
1	number of batches	4
2	ounces of grape juice in 1 batch	3
3	ounces of sparkling water in 1 batch	7
4	total ounces	

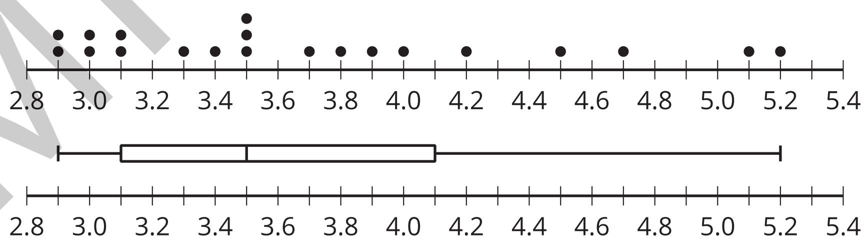
a. Write a formula for cell B4 that uses the values in cells B1, B2, and B3, to compute the total ounces of sparkling grape juice.

b. How would the output of the formula change if the value in cell B1 was changed to 10?

c. What would change about the sparkling grape juice if the value in B3 was changed to 10?

5 from Unit 1, Lesson 5

The dot plot and the box plot represent the same distribution of data.



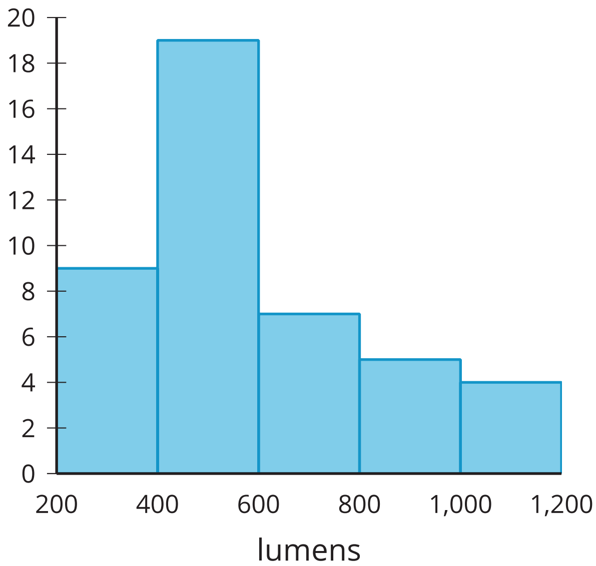
a. How does the median change when the highest value, 5.2, is removed?

b. How does the IQR change when the highest value, 5.2, is removed?

Sec C

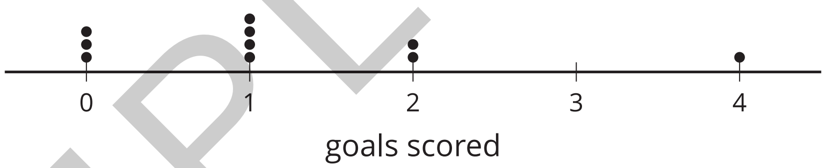
6 from Unit 1, Lesson 4

Describe the shape of the distribution shown in the histogram, which displays the light output, in lumens, of various light sources.



7 from Unit 1, Lesson 2

The dot plot represents the distribution of the number of goals scored by a soccer team in 10 games.



a. If possible, find the mean. If not possible, explain why not.

b. If possible, find the median. If not possible, explain why not.

c. Did the soccer team ever score exactly 3 goals in one of the games?

Sec C

Spreadsheet Shortcuts

Let's explore recursive formulas in spreadsheets.

8.1 Tables of Equivalent Ratios

Here is a table of equivalent ratios:

a	b
3	15
10	50
6	30
1	
	80

1. Complete the table with the missing values.
2. Explain what it means to say that the pairs of numbers are equivalent ratios.

8.2 The Birthday Trick

Navigate to this activity in the digital version of the materials or to ggbm.at/djcz6fjf.

1. In cell B4, we want to enter = B1 * 5 to multiply the month by 5. Begin by entering the "=" sign, but when you are about to type B1, instead, click on cell B1. Then type in the "*5." This shortcut can be used any time: Click on a cell instead of typing its address.
2. Practice this technique as you program each cell in B5 through B10 to perform the right computation.
3. When you are finished, does cell B10 show a number that contains the month and day of your birthday? If not, troubleshoot your computations.
4. Try changing the month and day in cells B1 and B2. The rest of the computations should automatically update. If not, troubleshoot your computations.

Are you ready for more?

Why does this trick work? Try using m for the month and d for the day, and writing the entire computation as an algebraic expression. Can you see why the resulting number contains the month and day?

Sec C

8.3 Using Spreadsheet Patterns

Navigate to this activity in the digital version of the materials or to the URL, ggbm.at/wu9t7kkd.

The spreadsheet contains a table of equivalent ratios.

1. Use spreadsheet calculations to continue the pattern in columns A and B, down to row 5. Pause for discussion.
2. Click on cell A5. See the tiny blue square in the bottom right corner of the cell? Click it and drag it down for several cells and let go.
3. Repeat this, starting with cell B5.

Lesson 8 Summary

Sometimes you want to create a list of numbers based on a rule. For example, let's say that the cost of a gym membership is a $25 sign-up fee followed by monthly dues of $35. We may want to know how much the membership will cost over the course of 6 months. We could use a spreadsheet and set it up this way:

	A	B	C
1	sign-up fee	25	
2	total cost after 1 month	=B1+35	
3	total cost after 2 months		
4	total cost after 3 months		
5	total cost after 4 months		
6	total cost after 5 months		
7	total cost after 6 months		
8			
9			

Which results in:

	A	B	C
1	sign-up fee	25	
2	total cost after 1 month	60	
3	total cost after 2 months		
4	total cost after 3 months		
5	total cost after 4 months		
6	total cost after 5 months		
7	total cost after 6 months		
8			
9			

See the little square on the lower-right corner of cell B2? If we click and drag that down, it will keep adding 35 to the value above to find the value in the next row. Drag it down far enough, and we can see the total cost after 6 months.

	A	B	C
1	sign-up fee	25	
2	total cost after 1 month	60	
3	total cost after 2 months	95	
4	total cost after 3 months	130	
5	total cost after 4 months	165	
6	total cost after 5 months	200	
7	total cost after 6 months	235	
8			
9			

Anytime you need to repeat a mathematical operation several times, continuing a pattern by dragging in a spreadsheet might be a good choice.

Practice Problems

1 *Technology required.* Open a blank spreadsheet. Use "fill down" to recreate this table of equivalent ratios. You should not need to type anything in rows 3–10.

	A	B
1	3	7
2	6	14
3	9	21
4	12	28
5	15	35
6	18	42
7	21	49
8	24	56
9	27	63
10	30	70

2 A list of numbers is made with this pattern: Start with 11, and subtract 4 to find the next number.

Here is the beginning of the list: 11, 7, 3, . . .

Explain how you could use "fill down" in a spreadsheet to find the tenth number in this list. (You do *not* need to actually find this number.)

3 Here is a spreadsheet showing the computations for a different version of the birthday trick:

	A	B
1	month	7
2	day	4
3		
4	multiply month by 50	
5	add 30	
6	multiply by 2	
7	add the day	
8	subtract 60	
9		
10		

Explain what formulas you would enter in cells B4 through B8 so that cell B8 shows a number representing the month and day. (In this example, cell B8 should show 704.) If you have access to a spreadsheet, try your formulas with a month and day to see whether it works.

4 from Unit 1, Lesson 7

Write a formula that you could type into a spreadsheet to compute the value of each expression.

a. $\frac{2}{5}$ of 35

b. $25 \div \frac{5}{3}$

c. $\left(\frac{1}{11}\right)^4$

d. The average of 0, 3, and 17

5 from Unit 1, Lesson 5

The values represent the number of cars in a town given a speeding ticket each day for 10 days.

2 4 5 5 7 7 8 8 8 12

a. What is the median? Interpret this value in the situation.

b. What is the IQR?

6 from Unit 1, Lesson 5

The values represent the most recent sale price, in thousands of dollars, of ten homes on a street.

85 91 93 99 99 99 102 108 110 115

a. What is the mean?

b. What is the MAD?

Unit 1, Lesson 9

Technological Graphing

Let's use technology to represent data.

9.1 It Begins With Data

Open a spreadsheet window and enter the data so that each value is in its own cell in column A.

1. How many values are in the spreadsheet? Explain your reasoning.

2. If you entered the data in the order that the values are listed, the number 7 is in the cell at position A1 and the number 5 is in cell A5. List all of the cells that contain the number 13.

3. In cell C1 type the word "Sum," in C2 type "Mean," and in C3 type "Median." You may wish to double-click or drag the vertical line between columns C and D to allow the entire words to be seen.

	A
1	7
2	8
3	4
4	13
5	5
6	15
7	14
8	8
9	12
10	2

	A
11	8
12	13
13	12
14	13
15	6
16	1
17	9
18	4
19	9
20	15

9.2 Finding Spreadsheet Statistics

Using the data from the *Warm-up*, we can calculate a few **statistics** and look at the data.

- Next to the word Sum, in cell D1, type =Sum(A1:A20)
- Next to the word Mean, in cell D2, type =Mean(A1:A20)
- Next to the word Median, in cell D3, type =Median(A1:A20)

1. What are the values for each of the statistics?

2. Change the value in A1 to 8. How does that change the statistics?

3. What value can be put into A1 to change the mean to 10.05 and the median to 9?

We can also use Geogebra to create data displays.

- Click on the letter A for the first column so that the entire column is highlighted.
- Click on the button that looks like a histogram to get a new window labeled One Variable Analysis .
- Click Analyze to see a histogram of the data.

1. Click the button Σx to see many of the statistics.
   a. What does the value for n represent?

   b. What does the value for Σx represent?

   c. What other statistics do you recognize?

2. Adjust the slider next to the word Histogram. What changes?

3. Click on the button to the right of the slider to bring in another window with more options. Then, click the box next to Set Classes Manually and set the Width to 5. What does this do to the histogram?

4. Click the word Histogram, and look at a box plot and dot plot of the data. When looking at the box plot, notice there is an x on the right side of the box plot. This represents a data point that is considered an outlier. Click on the button to the right of the slider, and uncheck the box labeled Show Outliers to include this point in the box plot. What changes? Why might you want to show outliers? Why might you want to include or exclude outliers?

9.3 Making Digital Displays

Use the data that you collected from the numerical, statistical question from a previous lesson. Use technology to create a dot plot, boxplot, and histogram for your data. Then find the mean, median, and interquartile range for the data.

Are you ready for more?

A stem and leaf plot is a table in which each data point is indicated by writing the first digit(s) on the left (the stem) and the last digit(s) on the right (the leaves). Each stem is written only once and shared by all data points with the same first digit(s). For example, the values 31, 32, and 45 might be represented like:

3	1 2
4	5

Key: 3 | 1 means 31

A class took an exam and earned the scores:

86	73	85	86	72	94	88	98
87	86	85	93	75	64	82	95
99	76	84	68				

1. Use technology to create a stem and leaf plot for this data set.

2. How can we see the shape of the distribution from this plot?

3. What information can we see from a stem and leaf plot that we cannot see from a histogram?

4. What do we have more control of in a histogram than in a stem and leaf plot?

Lesson 9 Summary

Data displays (like histograms or box plots) are very useful for quickly understanding a large amount of information, but often take a long time to construct accurately using pencil and paper. Technology can help create these displays as well as calculate useful *statistics* much faster than doing the same tasks by hand. Especially with very large data sets (in some experiments, millions of pieces of data are collected), technology is essential for putting the information into forms that are more easily understood.

A **statistic** is a quantity that is calculated from sample data as a measure of a distribution. *Mean* and *median* are examples of statistics that are measures of center. *Mean absolute deviation (MAD)* and *interquartile range (IQR)* are examples of statistics that are measures of variability. Although the interpretation must still be done by people, using the tools available can improve the accuracy and speed of doing computations and creating graphs.

Glossary

- statistic

Practice Problems

1 *Technology required.* The data represent the average customer ratings for several items sold online.

0.5 1 1.2 1.3 2.1 2.1 2.1 2.3 2.5 2.6 3.5 3.6 3.7 4 4.1 4.1

4.2 4.2 4.5 4.7 4.8

a. Use technology to create a histogram for the data with intervals 0–1, 1–2, and so on.

b. Describe the shape of the distribution.

c. Which interval has the highest frequency?

2 *Technology required.* The data represent the amount of corn, in bushels per acre, harvested from different locations.

133 133 134 134 134 135 135 135 135 135 135 136 136

136 137 137 138 138 139 140

a. Use technology to create a dot plot and a box plot.

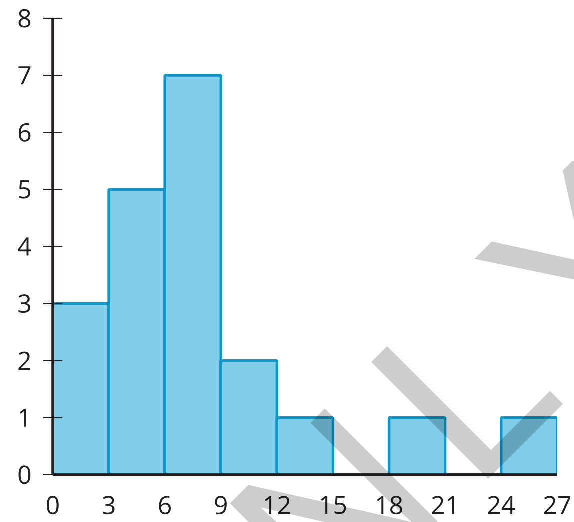
b. What is the shape of the distribution?

c. Compare the information displayed by the dot plot and box plot.

3

from Unit 1, Lesson 4

a. Describe the shape of the distribution.



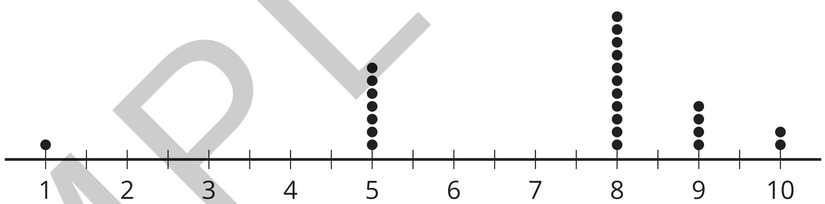
b. How many values are represented by the histogram?

c. Write a statistical question of interest that could have produced the data set summarized in the histogram.

4

from Unit 1, Lesson 3

The dot plot represents the distribution of satisfaction ratings for a landscaping company on a scale of 1 to 10. Twenty-five customers were surveyed.



On average, what was the satisfaction rating of the landscaping company?

The Effect of Extremes

Let's see how statistics change with the data.

10.1 At Least I Did Well

Several online video games match players with other players at random to compete in a team game.

1. What information could you use to determine the top players in a team game like this? Explain your reasoning.

2. There are two games of a similar type. One person claims that the best players play game A. Another person claims that game B has better players. How could you display data to help inform their discussion? Explain your reasoning.

10.2 Separated by Skew

1. Use technology to create a dot plot that represents the distribution of the data, then describe the shape of the distribution.

   6 7 8 8 9 9 9 10 10 10 10 11 11 11 12 12 13 14

2. Find the mean and median of the data.

3. Find the mean and median of the data with 2 additional values included as described.

   a. Add 2 values to the original data set that are greater than 14.

   b. Add 2 values to the original data set that are less than 6.

   c. Add 1 value that is greater than 14 and 1 value that is less than 6 to the original data set.

   d. Add the two values, 50 and 100, to the original data set.

4. Share your work with your group. What do you notice is happening with the mean and median based on the additional values?

5. Change the values so that the distribution fits the description given to you by your teacher, then find the mean and median.

6. Find another group that created a distribution with a different description. Explain your work and listen to their explanation, then compare your measures of center.

10.3 Plots Matching Measures

Create a possible dot plot with at least 10 values for each of the conditions listed. Each dot plot must have at least 3 values that are different.

1. a distribution that has both mean and median of 10

2. a distribution that has both mean and median of -15

3. a distribution that has a median of 2.5 and a mean greater than the median

4. a distribution that has a median of 5 and a median greater than the mean

Are you ready for more?

The mean and the median are by far the most common measures of center for numerical data. There are other measures of center, though, that are sometimes used. For each measure of center, list some possible advantages and disadvantages. Be sure to consider how it is affected by extremes.

1. *Interquartile mean*: The mean of only those points between the first quartile and the third quartile.

2. *Midhinge*: The mean of the first quartile and the third quartile.

3. *Midrange*: The mean of the minimum and maximum value.

4. *Trimean:* The mean of the first quartile, the median, the median again, and the third quartile. So we are averaging four numbers because the median is counted twice.

Lesson 10 Summary

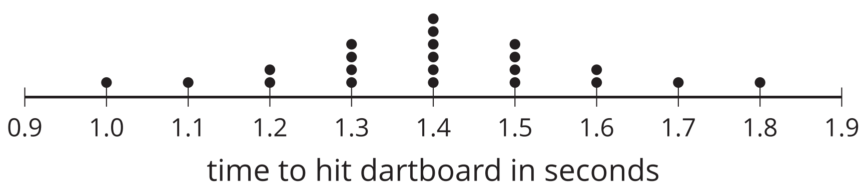
Is it better to use the mean or median to describe the center of a data set?

The mean gives equal importance to each value when finding the center. The mean usually represents the typical values well when the data have a symmetric distribution. On the other hand, the mean can be greatly affected by changes to even a single value.

The median tells you the middle value in the data set, so changes to a single value usually do not affect the median much. So, the median is more appropriate for data that are not very symmetrically distributed.

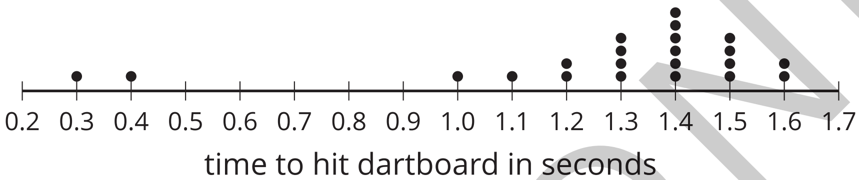
We can look at the distribution of a data set and draw conclusions about the mean and the median.

Here is a dot plot showing the amount of time a dart takes to hit a target in seconds. The data produce a symmetric distribution.



When a distribution is symmetric, the median and mean are both found in the middle of the distribution. Since the median is the middle value (or the mean of the two middle values) of a data set, you can use the symmetry around the center of a symmetric distribution to find it easily. For the mean, you need to know that the sum of the distances away from the mean of the values greater than the mean is equal to the sum of the distances away from the mean of the values less than the mean. Using the symmetry of the symmetric distribution you can see that there are four values 0.1 second above the mean, two values 0.2 seconds above the mean, one value 0.3 seconds above the mean, and one value 0.4 seconds above the mean. Likewise, you can see that there are the same number of values the same distances below the mean.

Here is a dot plot using the same data, but with two of the values changed, resulting in a skewed distribution.



When you have a skewed distribution, the distribution is not symmetric, so you are not able to use the symmetry to find the median and the mean. The median is still 1.4 seconds since it is still the middle value. The mean, on the other hand, is now about 1.273 seconds. The mean is less than the median because the lower values (0.3 and 0.4) result in a smaller value for the mean.

The median is usually more resistant to extreme values than is the mean. For this reason, the median is the preferred measure of center when a distribution is skewed or if there are extreme values. When using the median, you would also use the IQR as the preferred measure of variability. In a more symmetric distribution, the mean is the preferred measure of center, and the MAD is the preferred measure of variability.

Practice Problems

1 Select **all** the distribution shapes for which it is most often appropriate to use the mean.

A. bell-shaped

B. bimodal

C. skewed

D. symmetric

E. uniform

2 For which distribution shape is it usually appropriate to use the median when summarizing the data?

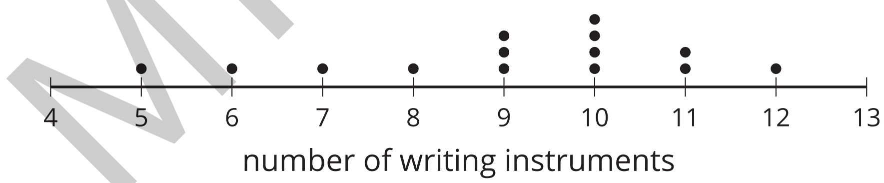
A. bell-shaped

B. skewed

C. symmetric

D. uniform

3 The number of writing instruments in some teachers' desks is displayed in the dot plot. Which is greater, the mean or the median? Explain your reasoning, using the shape of the distribution.



Sec D

4 from Unit 1, Lesson 9

A student has these scores on ten assignments. The teacher is considering dropping a lowest score. What effect does eliminating the lowest value, 0, from the data set have on the mean and median?

0 40 60 70 75 80 85 95 95 100

5 from Unit 1, Lesson 9

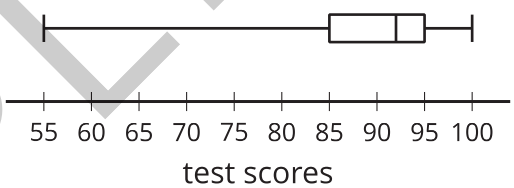
a. What is the five-number summary for the data?

2 2 4 4 5 5 6 7 9 15

b. When the maximum, 15, is removed from the data set, what is the five-number summary?

6 from Unit 1, Lesson 4

The box plot summarizes the test scores for 100 students:



Which term best describes the shape of the distribution?

A. bell-shaped

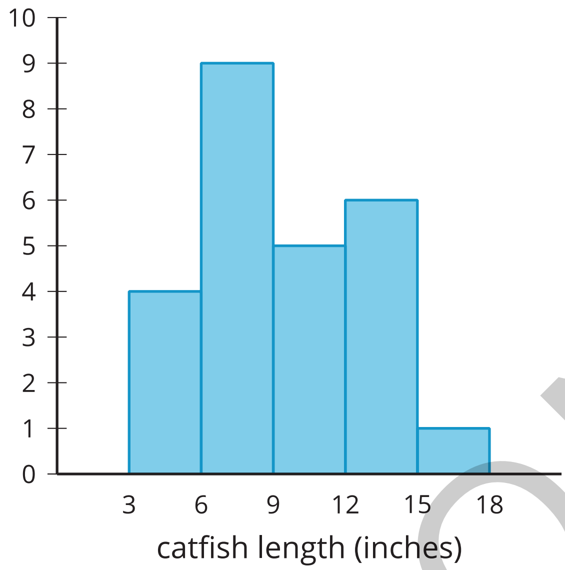
B. uniform

C. skewed

D. symmetric

7 from Unit 1, Lesson 2

The histogram represents the distribution of lengths, in inches, of 25 catfish caught in a lake.



a. If possible, find the mean. If not possible, explain why not.

b. If possible, find the median. If not possible, explain why not.

c. Were any of the fish caught 12 inches long?

d. Were any of the fish caught 19 inches long?

Comparing and Contrasting Data Distributions

Let's investigate variability using data displays and summary statistics.

11.1 Math Talk: Mean

Evaluate the mean of each data set mentally.

- 27 30 33
- 61 71 81 91 101
- 0 100 100 100 100
- 0 5 7 12

11.2 Card Sort: Describing Data Distributions

1. Your teacher will give you a set of cards. Take turns with your partner to match a data display with a written statement.
   a. For each match that you find, explain to your partner how you know it's a match.
   b. For each match that your partner finds, listen carefully to their explanation. If you disagree, discuss your thinking and work to reach an agreement.
2. After matching, determine if the mean or median is more appropriate for describing the center of the data set based on the distribution shape. Discuss your reasoning with your partner. If it is not given, calculate (if possible) or estimate the appropriate measure of center. Be prepared to explain your reasoning.

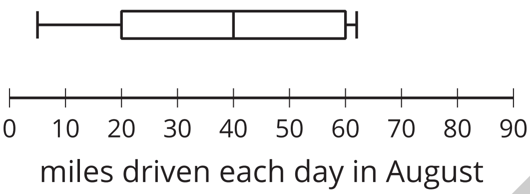
11.3 Visual Variability and Statistics

1. Each box plot summarizes the number of miles driven each day for 30 days in each month. The box plots represent, in order, the months of August, September, October, November, and December.

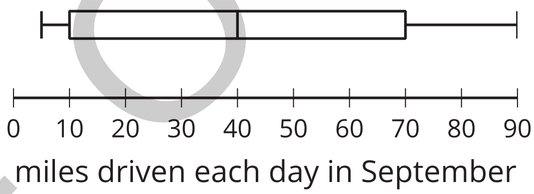
   a. The five box plots have the same median. Explain why the median is more appropriate for describing the center of the data set than the mean for these distributions.

   b. List the box plots in order of least variability to greatest variability. Check with another group to see if they agree.

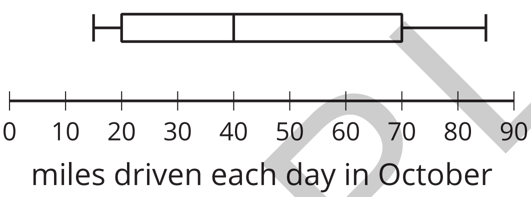
A



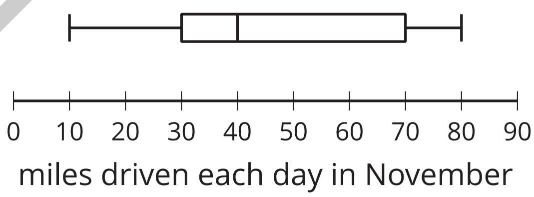
B



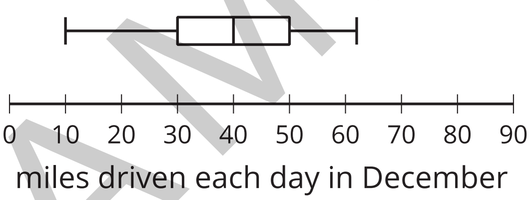
C



D



E

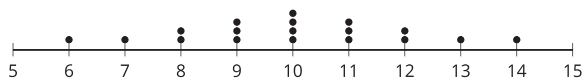


2. These five dot plots represent the number of books on different shelves of the same section of the library.

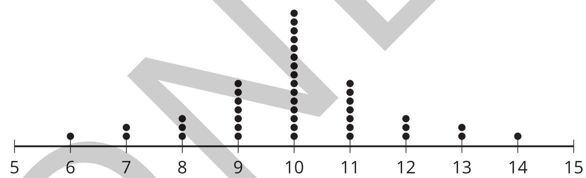
   a. The five dot plots have the same mean. Explain why the mean is more appropriate for describing the center of the data set than the median.

   b. List the dot plots in order of least variability to greatest variability. Check with another group to see if they agree.

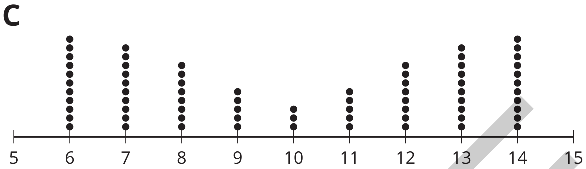
A



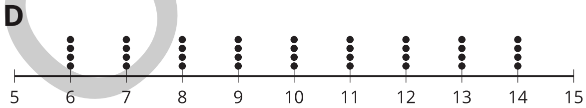
B



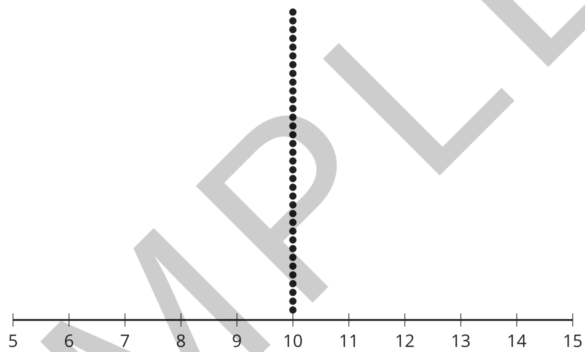
C



D

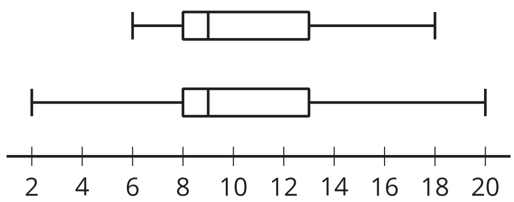


E

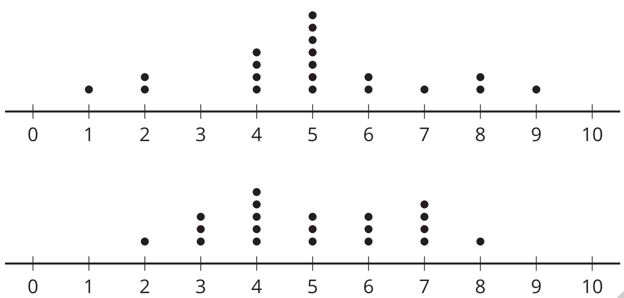


Are you ready for more?

1. These two box plots have the same median and the same IQR. How could we compare the variability of the two distributions?

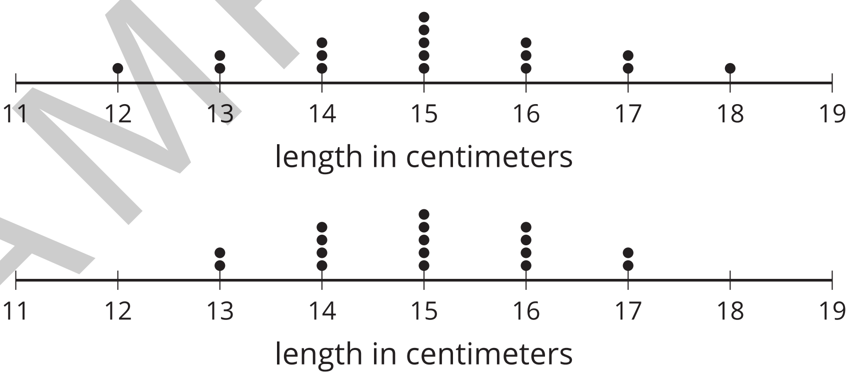


2. These two dot plots have the same mean and the same MAD. How could we compare the variability of the two distributions?



Lesson 11 Summary

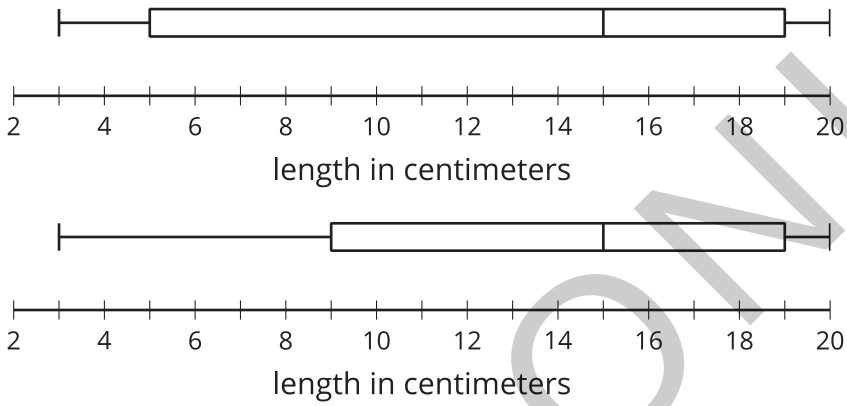
The mean absolute deviation, or MAD, is a measure of variability that is calculated by finding the mean distance from the mean of all the data points. Here are two dot plots, each with a mean of 15 centimeters, displaying the length of sea scallop shells in centimeters.



Notice that both dot plots show a symmetric distribution so the mean and the MAD are appropriate choices for describing center and variability. The data in the first dot plot appear to be more spread apart than the data in the second dot plot, so you can say that the first data set appears to have greater variability than does the second data set. This is confirmed by the MAD.

The MAD of the first data set is 1.18 centimeters and the MAD of the second data set is approximately 0.94 cm. This means that the values in the first data set are, on average, about 1.18 cm away from the mean, and the values in the second data set are, on average, about 0.94 cm away from the mean. The greater the MAD of the data, the greater the variability of the data.

The interquartile range, IQR, is a measure of variability that is calculated by subtracting the value for the first quartile, Q1, from the value for the third quartile, Q3. These two box plots represent the distributions of the lengths in centimeters of a different group of sea scallop shells, each with a median of 15 centimeters.



Notice that neither of the box plots have a symmetric distribution. The median and the IQR are appropriate choices for describing center and variability for these data sets. The middle half of the data displayed in the first box plot appear to be more spread apart, or show greater variability, than the middle half of the data displayed in the second box plot. The IQR of the first distribution is 14 cm, and the IQR is 10 cm for the second data set. The IQR measures the difference between the median of the second half of the data, Q3, and the median of the first half, Q1, of the data, so it is not affected by the minimum or the maximum value in the data set. It is a measure of the spread of the middle 50% of the data.

The MAD is calculated using every value in the data set, and the IQR is calculated using only the values for Q1 and Q3.

Practice Problems

1 In science class, Clare and Lin estimate the mass of eight different objects that actually weigh 2,000 grams each. Some summary statistics:

Clare

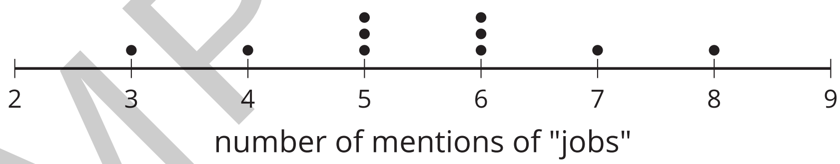
- mean: 2,000 grams
- MAD: 275 grams
- median: 2,000 grams
- IQR: 500 grams

Lin

- mean: 2,000 grams
- MAD: 225 grams
- median: 1,950 grams
- IQR: 350 grams

Which student was better at estimating the mass of the objects? Explain your reasoning.

2 A reporter counts the number of times a politician talks about jobs in campaign speeches. What is the MAD of the data represented in the dot plot?



A. 1.1 mentions

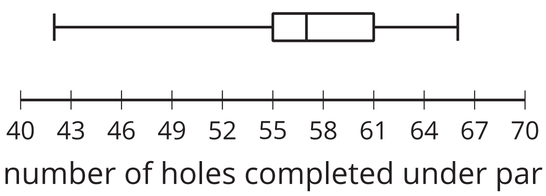
B. 2 mentions

C. 2.5 mentions

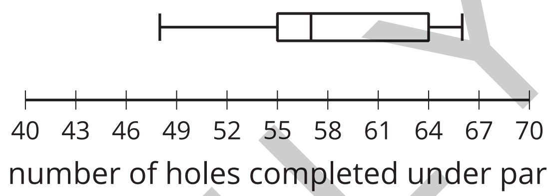
D. 5.5 mentions

3 Four amateur miniature golfers attempt to finish 100 holes under par several times. After each round of 100 holes, the number of holes they successfully complete under par is recorded. Due to the presence of extreme values, box plots were determined to be the best representation for the data. List the four box plots in order of variability from least to greatest.

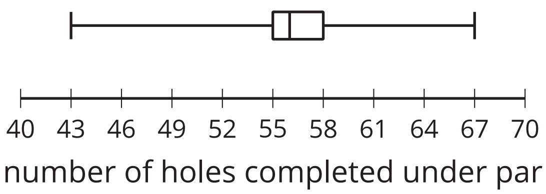
player a



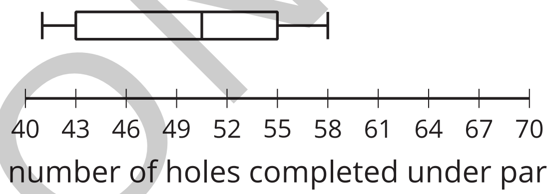
player b



player c



player d



4 from Unit 1, Lesson 10

Select **all** the distribution shapes for which the median *could be* much less than the mean.

A. symmetric

B. bell-shaped

C. skewed left

D. skewed right

E. bimodal

5 from Unit 1, Lesson 9

a. What is the five-number summary for the data?

0 2 2 4 5 5 5 5 7 11

b. When the minimum, 0, is removed from the data set, what is the five-number summary?

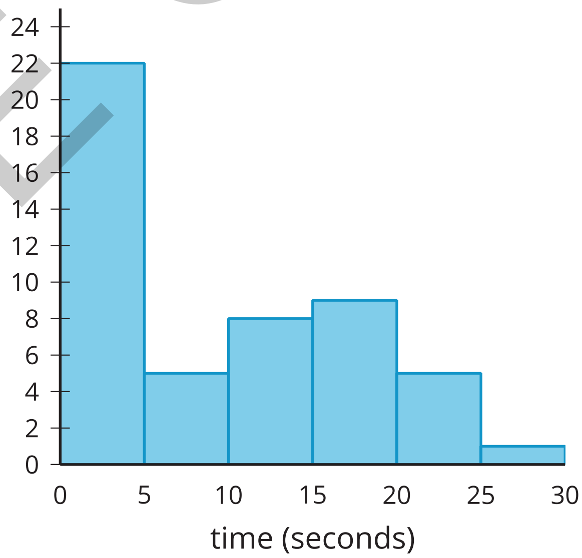
6 from Unit 1, Lesson 9

What effect does eliminating the highest value, 180, from the data set have on the mean and median?

25 50 50 60 70 85 85 90 90 180

7 from Unit 1, Lesson 3

The histogram represents the distribution of the number of seconds it took for each of 50 students to find the answer to a trivia question using the internet. Which interval contains the median?



A. 0 to 5 seconds

B. 5 to 10 seconds

C. 10 to 15 seconds

D. 15 to 20 seconds

Unit 1, Lesson 12

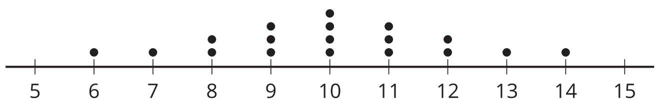
Standard Deviation

Let’s learn about standard deviation, another measure of variability.

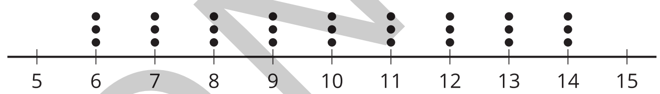
12.1 Notice and Wonder: Measuring Variability

What do you notice? What do you wonder?

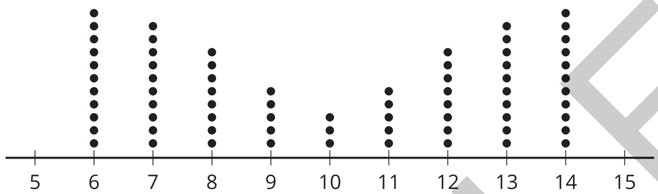
mean: 10, MAD: 1.56, standard deviation: 2



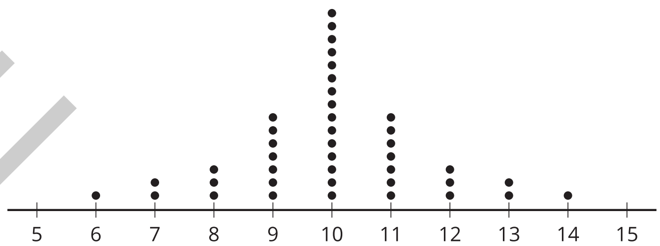
mean: 10, MAD: 2.22, standard deviation: 2.58



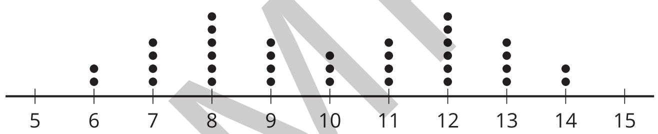
mean: 10, MAD: 2.68, standard deviation: 2.92



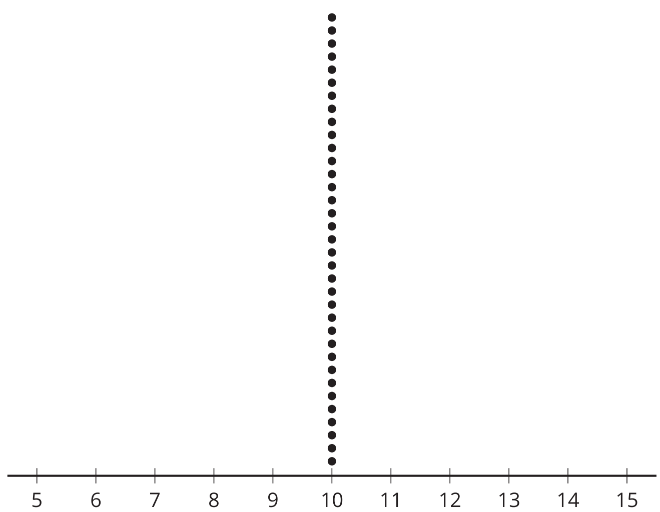
mean: 10, MAD: 1.12, standard deviation: 1.61



mean: 10, MAD: 2.06, standard deviation: 2.34



mean: 10, MAD: 0, standard deviation: 0



Sec D

12.2 Investigating Standard Deviation

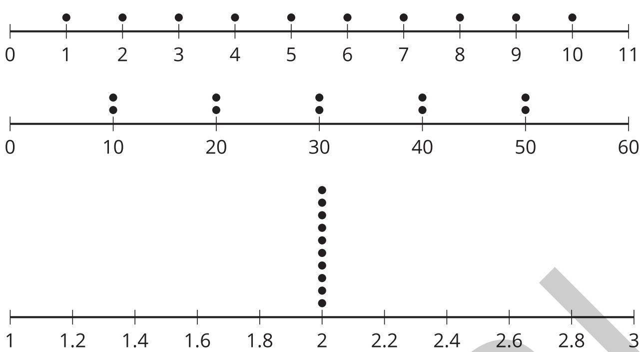
Use technology to find the mean and the standard deviation for the data in the dot plots.

1. What do you notice about the mean and standard deviation that you and your partner found for the three dot plots?

2. Invent some data that fit the conditions. Be prepared to share your data set and reasoning for choice of values.

Partner 1

Dot plots:

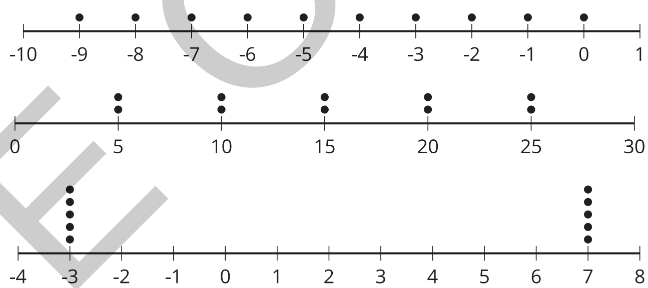


Conditions:

- 10 numbers with a standard deviation equal to the standard deviation of your first dot plot with a mean of 6.
- 10 numbers with a standard deviation three times greater than the data in the first row.
- 10 different numbers with a standard deviation as close to 2 as you can get in 1 minute.

Partner 2

Dot plots:



Conditions:

- 10 numbers with a standard deviation equal to the standard deviation of your first dot plot with a mean of 12.
- 10 numbers with a standard deviation four times greater than the data in the first row.
- 10 different numbers with a standard deviation as close to 2 as you can get in 1 minute.

Sec D

12.3 Investigating Variability

Begin with the data:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

1. Use technology to find the mean, standard deviation, median, and interquartile range.

2. How do the standard deviation and mean change when you remove the greatest value from the data set? How do they change if you add a value to the data set that is twice the greatest value?

3. What do you predict will happen to the standard deviation and mean when you remove the least value from the data set? Check to see if your prediction was correct.

4. What happens to the standard deviation and mean when you add a value to the data set equal to the mean? Add a second value equal to the mean. What happens?

5. Add, change, and remove values from the data set to answer the question: What appears to change more easily, the standard deviation or the interquartile range? Explain your reasoning.

Are you ready for more?

How is the standard deviation calculated? We have seen that the standard deviation behaves a lot like the mean absolute deviation, and that is because the key idea behind both is the same.

1. Using the original data set, calculate the deviation of each point from the mean by subtracting the mean from each data point.

2. If we just tried to take a mean of those deviations what would we get?

3. There are two common ways to turn negative values into more-useful positive values: take the absolute value or square the value. To find the MAD we find the absolute value of each deviation, then find the mean of those numbers. To find the standard deviation we square each of the deviations, then find the mean of those numbers. Then finally take the square root of that mean. Compute the MAD and the standard deviation of the original data set.

Lesson 12 Summary

We can describe the variability of a distribution using the **standard deviation**. The standard deviation is a measure of variability that is calculated using a method that is similar to the one used to calculate the MAD, or mean absolute deviation.

A deeper understanding of the importance of standard deviation as a measure of variability will come with a deeper study of statistics. For now, know that the standard deviation is mathematically important and will be used as the appropriate measure of variability when the mean is an appropriate measure of center.

Like the MAD, the standard deviation is large when the data set is more spread out, and the standard deviation is small when the variability is small. The intuition you gained about MAD will also work for the standard deviation.

Glossary

- standard deviation

Practice Problems

1 The shoe size for all the pairs of shoes in a person's closet are recorded.

7 7 7 7 7 7 7 7 7 7

a. What is the mean?

b. What is the standard deviation?

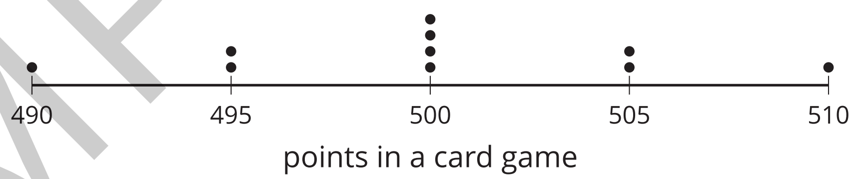
2 Here is a data set:

1 2 3 3 4 4 4 4 5 5 6 7

a. What happens to the mean and standard deviation of the data set when the 7 is changed to a 70?

b. For the data set with the value of 70, why would the median be a better choice for the measure of center than the mean?

3 Which of these best estimates the standard deviation of points in a card game?



A. 5 points

B. 20 points

C. 50 points

D. 500 points

4 from Unit 1, Lesson 11

The mean of data set A is 43.5 and the MAD is 3.7. The mean of data set B is 12.8 and the MAD is 4.1.

a. Which data set shows greater variability? Explain your reasoning.

b. What differences would you expect to see when comparing the dot plots of the two data sets?

Sec D

5 from Unit 1, Lesson 10

Select **all** the distribution shapes for which the mean and median *must be* about the same.

A. bell-shaped

B. bimodal

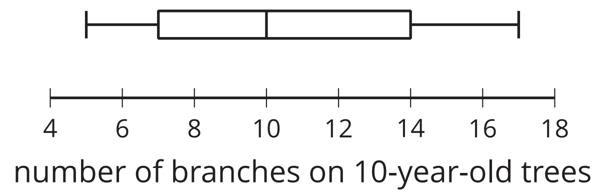
C. skewed

D. symmetric

E. uniform

6 from Unit 1, Lesson 11

What is the IQR?



A. 5 branches

B. 7 branches

C. 10 branches

D. 12 branches

7 from Unit 1, Lesson 9

The data represent the number of cans collected by different classes for a service project.

12 14 22 14 18 23 42 13 9 19 22 14

a. Find the mean.

b. Find the median.

c. Eliminate the greatest value, 42, from the data set. Explain how the measures of center change.

Unit 1, Lesson 13

More Standard Deviation

Let's continue to interpret standard deviation.

13.1 What Do You Want to Know?

100 captive Asian and 100 wild Asian elephants are weighed. There is a meaningful difference between the masses of the 2 groups if the measures of center are at least twice as far apart as the measure of variability. Is there a meaningful difference between the masses of these 2 groups of elephants? Explain your reasoning.

Sec D

13.2 Information Gap: African and Asian Elephants

Your teacher will give you either a *problem card* or a *data card*. Do not show or read your card to your partner.

If your teacher gives you the *problem card*:

1. Silently read your card and think about what information you need to answer the question.
2. Ask your partner for the specific information that you need. “Can you tell me _____?”
3. Explain to your partner how you are using the information to solve the problem. “I need to know _____ because . . . .”

   Continue to ask questions until you have enough information to solve the problem.
4. Once you have enough information, share the problem card with your partner, and solve the problem independently.
5. Read the data card, and discuss your reasoning.

If your teacher gives you the *data card*:

1. Silently read your card. Wait for your partner to ask for information.
2. Before telling your partner any information, ask, “Why do you need to know _____?”
3. Listen to your partner’s reasoning and ask clarifying questions. Only give information that is on your card. Do not figure out anything for your partner!

   These steps may be repeated.
4. Once your partner says they have enough information to solve the problem, read the problem card, and solve the problem independently.
5. Share the data card, and discuss your reasoning.

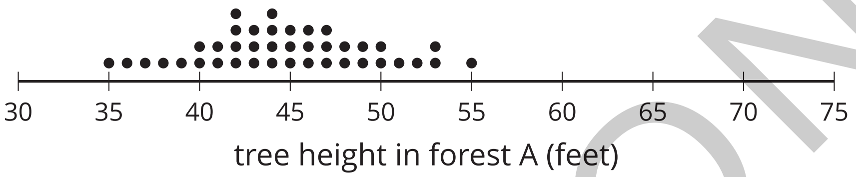
13.3 Interpreting Measures of Center and Variability

For each situation, you are given two graphs of data, a measure of center for each, and a measure of variability for each.

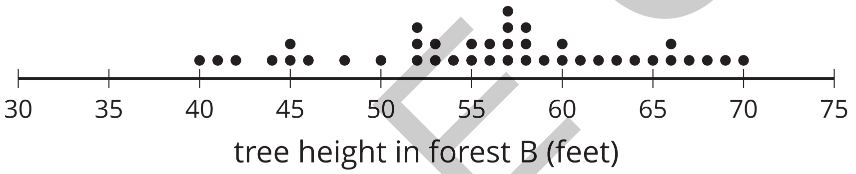
- Interpret the measure of center in terms of the situation.
- Interpret the measure of variability in terms of the situation.
- Compare the two data sets.

1. The heights of the 40 trees in each of two forests are collected.

mean: 44.8 feet, standard deviation: 4.72 feet

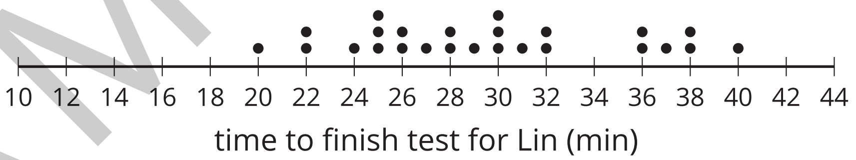


mean: 56.03 feet, standard deviation: 7.87 feet

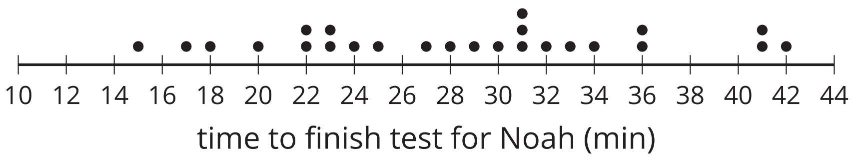


2. The number of minutes it takes Lin and Noah to finish their tests in German class is collected for the year.

mean: 29.48 minutes, standard deviation: 5.44 minutes

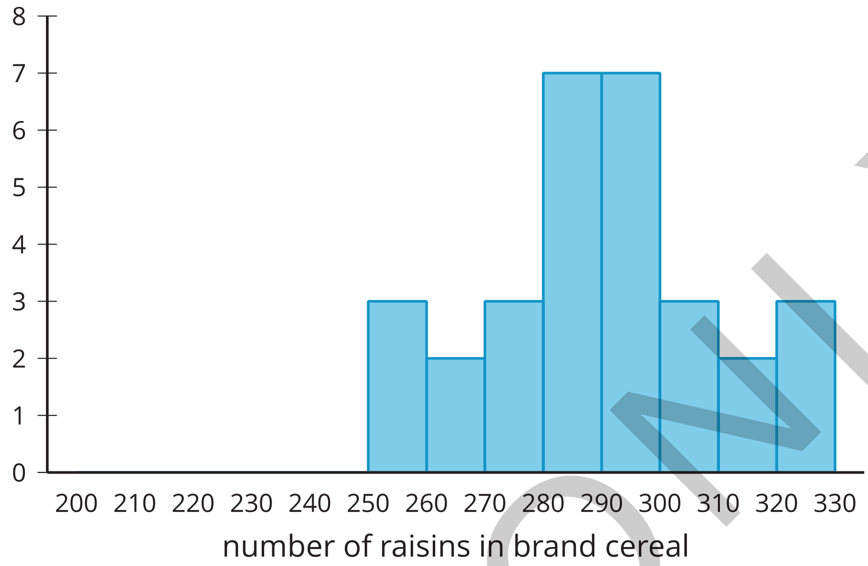


mean: 28.44 minutes, standard deviation: 7.40 minutes

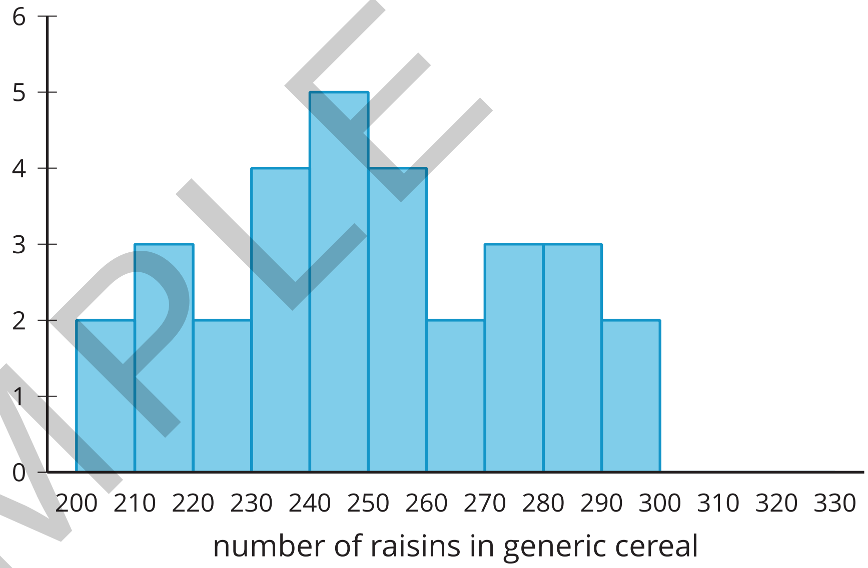


3. The number of raisins in a cereal with a name brand and the generic version of the same cereal are collected for several boxes.

mean: 289.1 raisins, standard deviation: 19.8 raisins



mean: 249.17 raisins, standard deviation: 26.35 raisins



Sec D

Are you ready for more?

One use of standard deviation is that it gives a natural scale as to how far above or below the mean a data point is. This is incredibly useful for comparing points from two different distributions.

For example, there is a saying that you cannot compare apples and oranges, but here is a way. The average weight of a granny smith apple is 128 grams with a standard deviation of about 10 grams. The average weight of a navel orange is 140 grams with a standard deviation of about 14 grams. If we have a 148 gram granny smith apple and a 161 gram navel orange, we might wonder which is larger for its species even though they are both about 20 grams above their respective mean. We could say that the apple, which is 2 standard deviations above its mean, is larger for its species than the orange, which is only 1.5 standard deviations above its mean.

1. How many standard deviations above the mean height of a tree in forest A is its tallest tree?

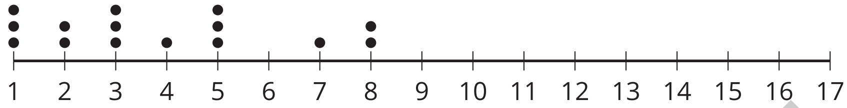
2. How many standard deviations above the mean height of a tree in forest B is its tallest tree?

3. Which tree is relatively taller in its forest?

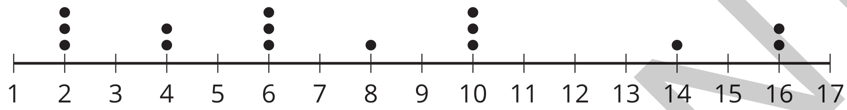
Lesson 13 Summary

The more variation a distribution has, the greater the standard deviation. A more compact distribution will have a lesser standard deviation.

The first dot plot shows the number of points that a player on a basketball team made during each of 15 games. The second dot plot shows the number of points scored by another player during the same 15 games.



points for player one



points for player two

The data in the first plot have a mean of approximately 3.87 points and standard deviation of about 2.33 points. The data in the second plot have a mean of approximately 7.73 points and a standard deviation of approximately 4.67 points. The second distribution has greater variability than the first distribution because the data are more spread out. This is shown in the standard deviation for the second distribution being greater than the standard deviation for the first distribution.

Standard deviation is calculated using the mean, so it makes sense to use it as a measure of variability when the mean is appropriate to use for the measure of center. In cases where the median is a more appropriate measure of center, the interquartile range is still a better measure of variability than standard deviation.

Practice Problems

1 Three drivers compete in the same fifteen races. The mean and standard deviation for the race times of each of the drivers are given.

Driver A has a mean race time of 4.01 seconds and a standard deviation of 0.05 seconds.

Driver B has a mean race time of 3.96 seconds and a standard deviation of 0.12 seconds.

Driver C has a mean race time of 3.99 seconds and a standard deviation of 0.19 seconds.

a. Which driver has the fastest typical race time?

b. Which driver's race times are the most variable?

c. Which driver do you predict will win the next race? Support your prediction using the mean and standard deviation.

Sec D

2 The widths, in millimeters, of fabric produced at a ribbon factory are collected. The mean is approximately 23 millimeters, and the standard deviation is approximately 0.6 millimeters.

Interpret the mean and standard deviation in the context of the problem.

3 Select **all** the statements that are true about standard deviation.

A. It is a measure of center.

B. It is a measure of variability.

C. It is the same as the MAD.

D. It is calculated using the mean.

E. It is calculated using the median.

4 from Unit 1, Lesson 12

The number of different species of plants in some gardens is recorded.

1 2 3 4 4 5 5 6 7 8

a. What is the mean?

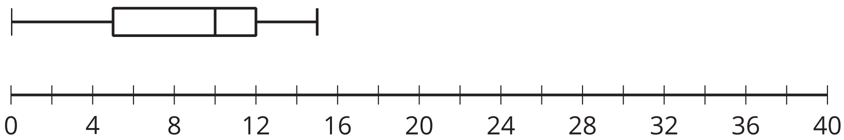
b. What is the standard deviation?

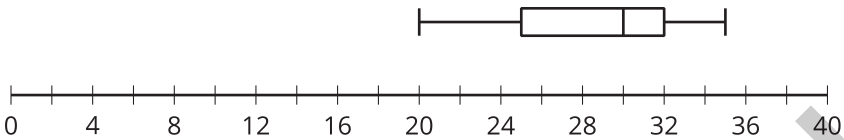
5 from Unit 1, Lesson 12

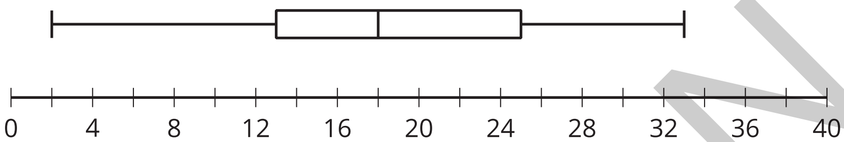
A set of data has ten numbers. The mean of the data is 12 and the standard deviation is 0. What values could make up a data set with these statistics?

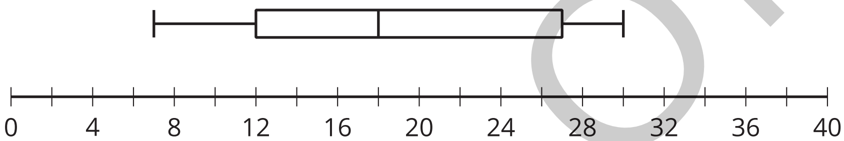
6 from Unit 1, Lesson 11

Which box plot has the largest interquartile range?

A. 

B. 

C. 

D. 

7 from Unit 1, Lesson 9

a. What is the five-number summary for?

1 3 3 3 4 8 9 10 10 17

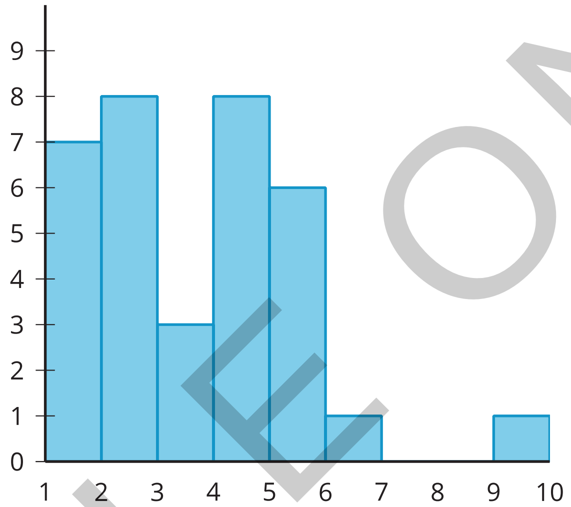
b. When the maximum, 17, is removed from the data set, what is the five-number summary?

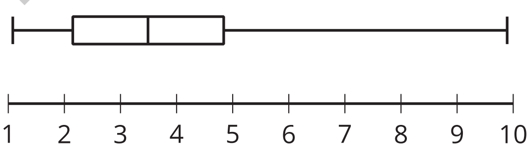
Outliers

Let's investigate outliers and how to deal with them.

14.1 Health Care Spending

The histogram and box plot show the average amount of money, in thousands of dollars, spent on each person in the country (per capita spending) for health care in 34 countries.





1. One value in the set is an **outlier**. Which one is it? What is its approximate value?

2. By one rule for deciding, a value is an outlier if it is more than 1.5 times the IQR greater than Q3. Show on the box plot whether or not your value meets this definition of outlier.

14.2 Investigating Outliers

Here is the data set used to create the histogram and box plot from the warm-up.

1.0803	1.0875	1.4663	1.7978	1.9702	1.9770	1.9890	2.1011	2.1495	2.2230
2.5443	2.7288	2.7344	2.8223	2.8348	3.2484	3.3912	3.5896	4.0334	4.1925
4.3763	4.5193	4.6004	4.7081	4.7528	4.8398	5.2050	5.2273	5.3854	5.4875
5.5284	5.5506	6.6475	9.8923						

1. Use technology to find the mean, standard deviation, and five-number summary.

2. The maximum value in this data set represents the spending for the United States. Should the per capita health spending for the United States be considered an outlier? Explain your reasoning.

3. Although outliers should not be removed without considering their cause, it is important to see how influential outliers can be for various statistics. Remove the value for the United States from the data set.

   a. Use technology to calculate the new mean, standard deviation, and five-number summary.

   b. How do the mean, standard deviation, median, and interquartile range of the data set with the outlier removed compare to the same summary statistics of the original data set?

14.3 Origins of Outliers

1. The number of passenger electric vehicles registered is collected for the 39 counties of Washington State.

    - mean: 3,589.3 cars
    - minimum: 3 cars
    - Q1: 170 cars
    - median: 506 cars
    - Q3: 1,560 cars
    - maximum: 73,996 cars

    16223 337 460 60 467 73996 8368 1556 222 806 1736 3 238

    3444 165 706 3497 10657 37 278 186 424 170 45 4425 4601

    36 677 554 25 858 12 796 773 1560 44 194 841 506

    a. Are any of the values outliers? Explain or show your reasoning.

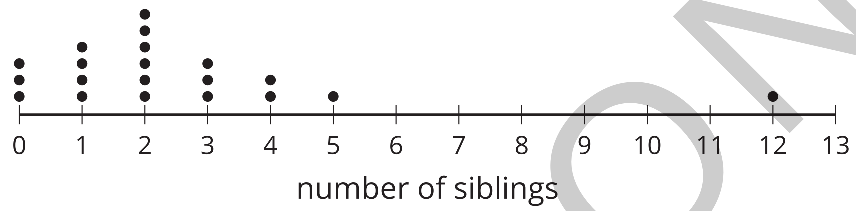
    b. If there are any outliers, why do you think they might exist? Should they be included in an analysis of the data?

2. The situations described here each have an outlier. For each situation, how would you determine if it is appropriate to keep or remove the outlier when analyzing the data? Discuss your reasoning with your partner.

   a. A number cube has sides labeled 1–6. After rolling 15 times, Tyler records his data:

      1 1 1 1 2 2 3 3 4 4 5 5 5 6 20

   b. The dot plot represents the distribution of the number of siblings reported by a group of 20 people.



   c. In a science class, 11 groups of students are synthesizing biodiesel. At the end of the experiment, each group recorded the mass in grams of the biodiesel they synthesized. The masses of biodiesel are

      0 1.245 1.292 1.375 1.383 1.412 1.435 1.471 1.482 1.501

      1.532

Are you ready for more?

Look back at some of the numerical data that you and your classmates collected in the first lesson of this unit.

1. Are any of the values outliers? Explain or show your reasoning.

2. If there are any outliers, why do you think they might exist? Should they be included in an analysis of the data?

Lesson 14 Summary

In statistics, an **outlier** is a data value that is unusual in that it differs quite a bit from the other values in the data set.

Outliers occur in data sets for a variety of reasons including, but not limited to:

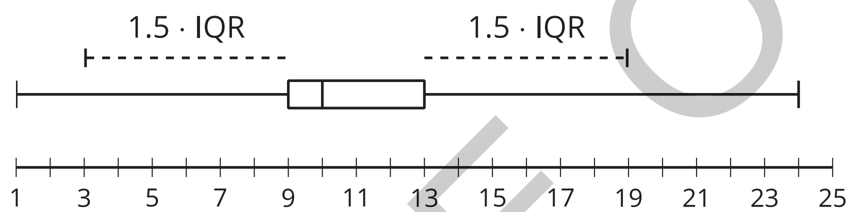
- Errors in the data that result from the data collection or data entry process.
- Results in the data that represent unusual values that occur in the population.

Outliers can reveal cases worth studying in detail or errors in the data collection process. In general, they should be included in any analysis done with the data.

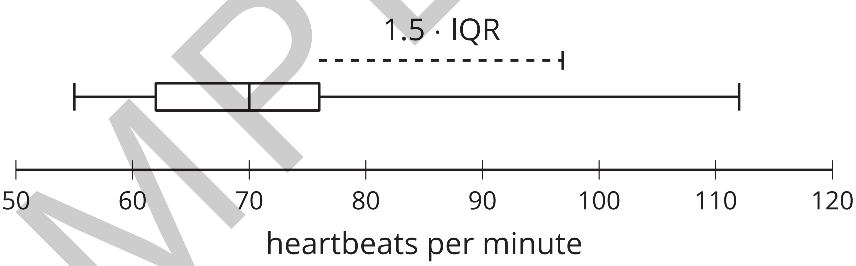
A value is an outlier if it is

- More than 1.5 times the interquartile range greater than Q3 (if $x > \text{Q3} + 1.5 \cdot \text{IQR}$).
- More than 1.5 times the interquartile range less than Q1 (if $x < \text{Q1} - 1.5 \cdot \text{IQR}$).

In this box plot, the minimum and maximum are at least two outliers.



It is important to identify the source of outliers because outliers can affect measures of center and variability in significant ways. The box plot displays the resting heart rate, in beats per minute (bpm), of 50 athletes taken five minutes after a workout.



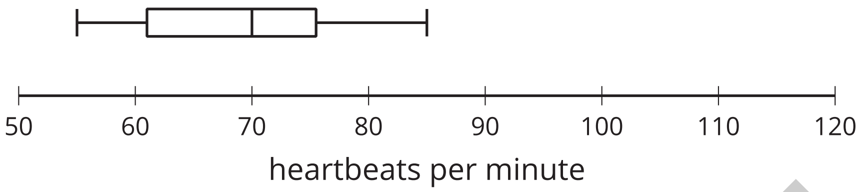
Some summary statistics include:

- mean: 69.78 bpm
- standard deviation: 10.71 bpm
- minimum: 55 bpm
- Q1: 62 bpm
- median: 70 bpm
- Q3: 76 bpm
- maximum: 112 bpm

It appears that the maximum value of 112 bpm may be an outlier. Beacuse the interquartile range is 14 bpm ($76 - 62 = 14$) and $\text{Q3} + 1.5 \cdot \text{IQR} = 97$, we should label the maximum value as an outlier. Searching through the actual data set, it could be confirmed that this is the only outlier.

After reviewing the data collection process, it is discovered that the athlete with the heart rate measurement of 112 bpm was taken one minute after a workout instead of five minutes after. The outlier should be deleted from the data set because it was not obtained under the right conditions.

Once the outlier is removed, the box plot and summary statistics are:



- mean: 68.92 bpm
- standard deviation: 8.9 bpm
- minimum: 55 bpm
- Q1: 61 bpm
- median: 70 bpm
- Q3: 75.5 bpm
- maximum: 85 bpm

The mean decreased by 0.86 bpm and the median remained the same. The standard deviation decreased by 1.81 bpm which is about 17% of its previous value. Based on the standard deviation, the data set with the outlier removed shows much less variability than the original data set containing the outlier. Because the mean and standard deviation use all of the numerical values, removing one very large data point can affect these statistics in important ways.

The median remained the same after the removal of the outlier and the IQR increased slightly. These measures of center and variability are much more resistant to change than the mean and standard deviation are. The median and IQR measure the middle of the data based on the number of values rather than the actual numerical values themselves, so the loss of a single value will not often have a great effect on these statistics.

The source of any possible errors should always be investigated. If the measurement of 112 beats per minute was found to be taken under the right conditions and merely included an athlete whose heart rate did not slow as much as the other athletes' heart rate, it should not be deleted so that the data reflect the actual measurements. If the situation cannot be revisited to determine the source of the outlier, it should not be removed. To avoid tampering with the data and to report accurate results, data values should not be deleted unless they can be confirmed to be an error in the data collection or data entry process.

Glossary

- outlier

Practice Problems

1 The number of letters received in the mail over the past week is recorded.

2 3 5 5 5 15

Which value appears to be an outlier?

A. 2

B. 3

C. 5

D. 15

2 Elena collects 112 specimens of beetle and records their lengths for an ecology research project. When she returns to the laboratory, Elena finds that she incorrectly recorded one of lengths of the beetles as 122 centimeters (about 4 feet). What should she do with the outlier, 122 centimeters, when she analyzes her data?

3 Mai took a survey of students in her class to find out how many hours they spend reading each week. Here are some summary statistics for the data that Mai gathered:

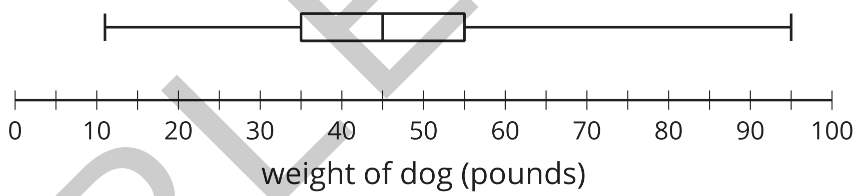
mean: 8.5 hours — standard deviation: 5.3 hours

median: 7 hours — Q1: 5 hours — Q3: 11 hours

a. Give an example of a number of hours larger than the median which would be an outlier. Explain your reasoning.

b. Are there any outliers below the median? Explain your reasoning.

4 The box plot shows the statistics for the weight, in pounds, of some dogs.



Are there any outliers? Explain how you know.

5 from Unit 1, Lesson 13

The mean exam score for the first group of twenty examinees applying for a security job is 35.3 with a standard deviation of 3.6.

The mean exam score for the second group of twenty examinees is 34.1 with a standard deviation of 0.5. Both distributions are close to symmetric in shape.

a. Use the mean and standard deviation to compare the scores of the two groups.

b. The minimum score required to get an in-person interview is 33. Which group do you think has more people get in-person interviews?

6 from Unit 1, Lesson 13

A group of pennies made in 2018 are weighed. The mean is approximately 2.5 grams with a standard deviation of 0.02 grams.

Interpret the mean and standard deviation in terms of the context.

7 from Unit 1, Lesson 12

These values represent the expected number of paintings a person will produce over the next 10 days.

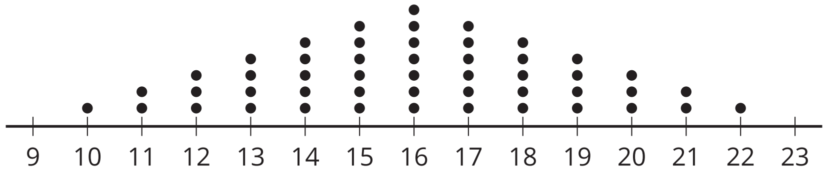
0 0 0 1 1 1 2 2 3 5

a. What are the mean and standard deviation of the data?

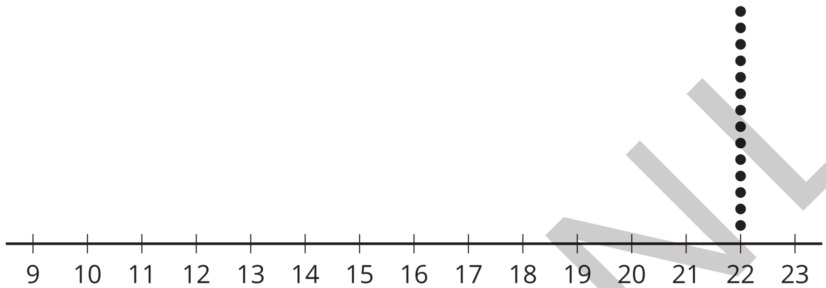
b. The artist is not pleased with these statistics. If the 5 is increased to a larger value, how does this affect the median, mean, and standard deviation?

8 from Unit 1, Lesson 11

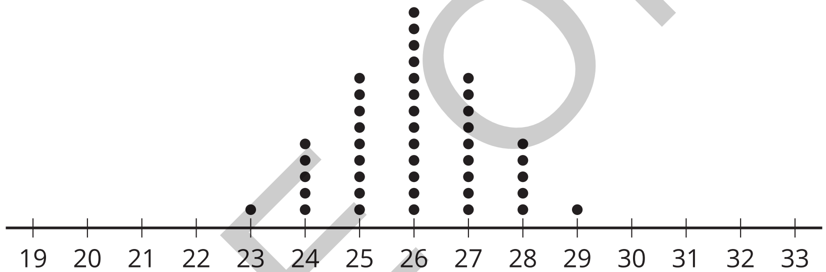
List the four dot plots in order of variability from least to greatest.



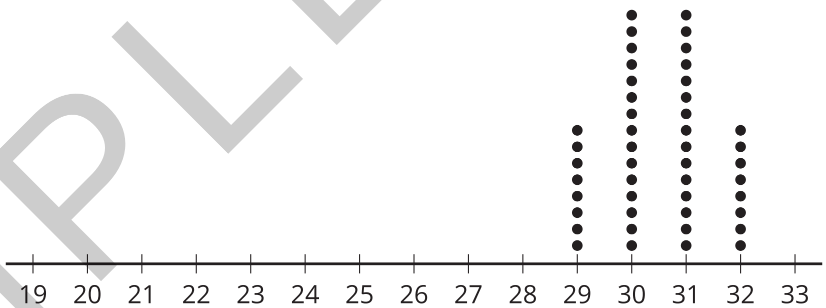
a.



b.



c.



d.

Comparing Data Sets

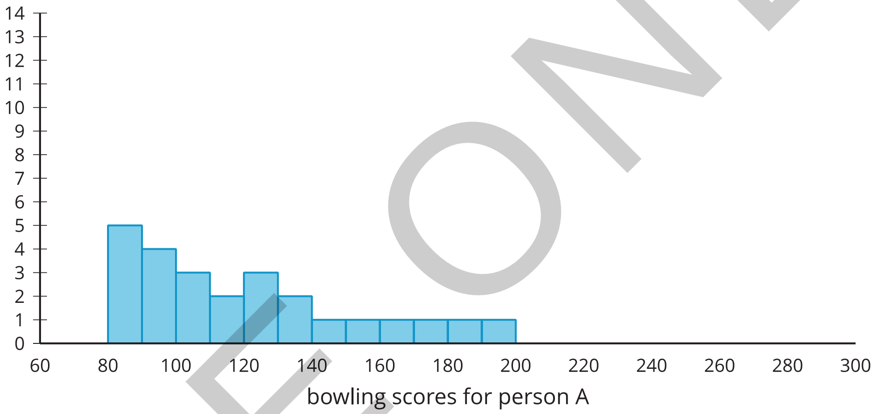
Let's compare statistics for data sets.

15.1 Bowling Partners

Each histogram shows the bowling scores for the last 25 games played by each person. Choose 2 of these people to join your bowling team. Explain your reasoning.

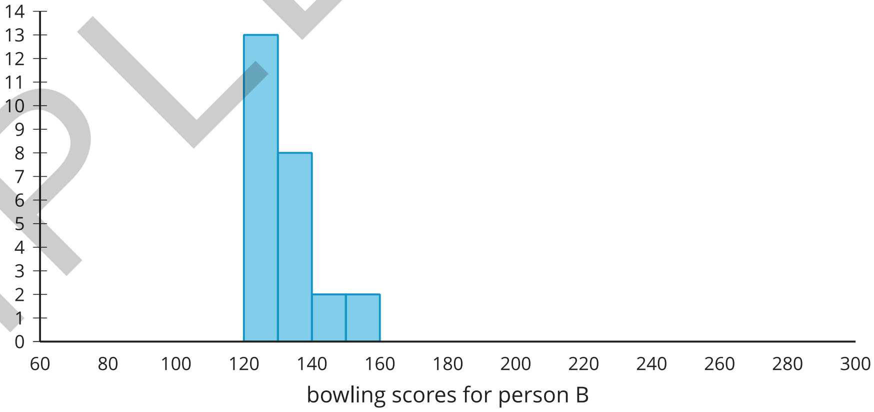
Person A

- mean: 118.96
- median: 111
- standard deviation: 32.96
- interquartile range: 44



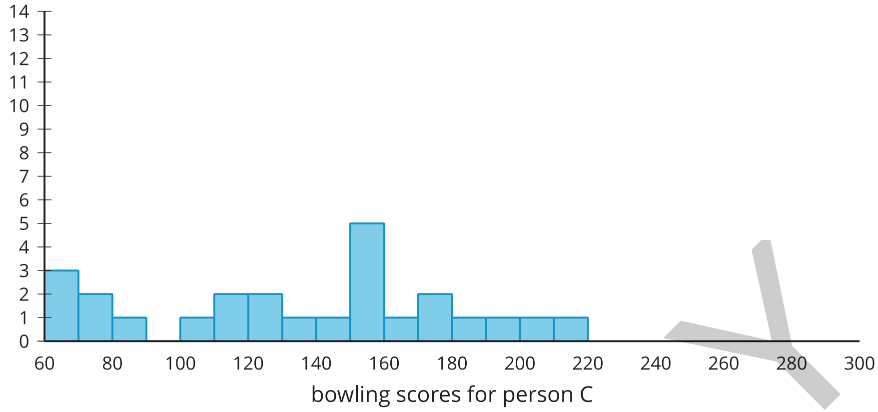
Person B

- mean: 131.08
- median: 129
- standard deviation: 8.64
- interquartile range: 8



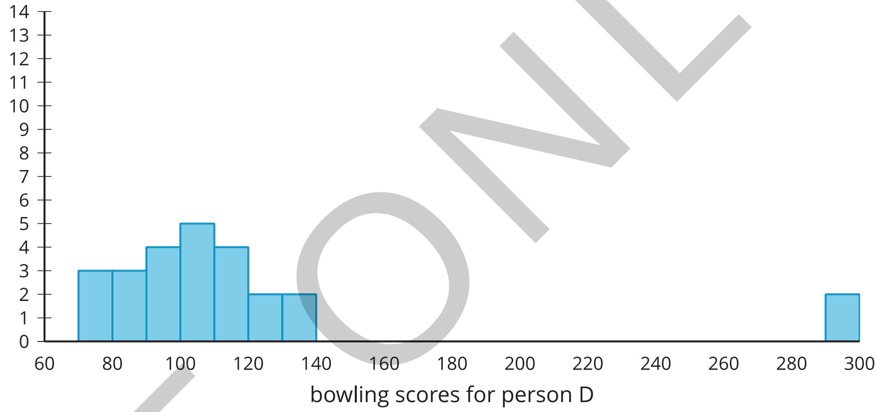
Person C

- mean: 133.92
- median: 145
- standard deviation: 45.04
- interquartile range: 74



Person D

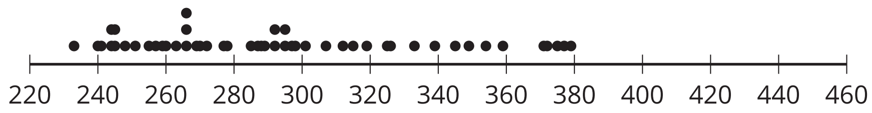
- mean: 116.56
- median: 103
- standard deviation: 56.22
- interquartile range: 31.5



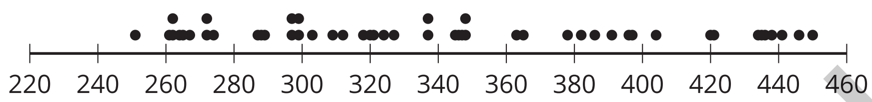
Sec D

15.2 Comparing Marathon Times

All of the marathon runners from each of two different age groups have their finishing times represented in the dot plot.



ages 30 – 39 marathon finish times (minutes)



ages 40 – 49 marathon finish times (minutes)

1. Which age group tends to take longer to run the marathon? Explain your reasoning.

2. Which age group has more variable finish times? Explain your reasoning.

Are you ready for more?

1. How do you think finish times for a 20–29 age range will compare to these two distributions?

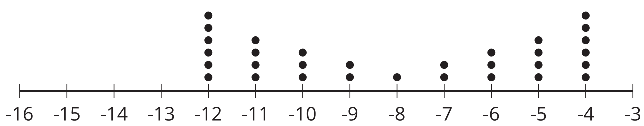
2. Find some actual marathon finish times for this group and make a dot or box plot of your data to help compare.

15.3 Comparing Measures

For each group of data sets,

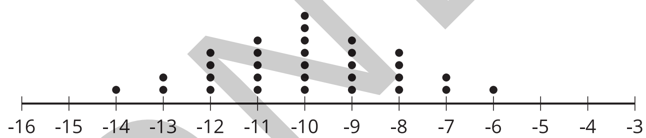
- Determine the best measure of center and measure of variability to use based on the shape of each distribution.
- Determine which set has the greatest measure of center.
- Determine which set has the greatest measure of variability.
- Be prepared to explain your reasoning.

1a



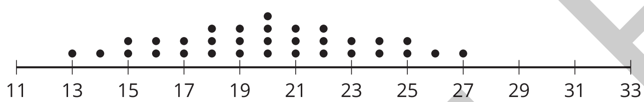
distribution 1a

1b



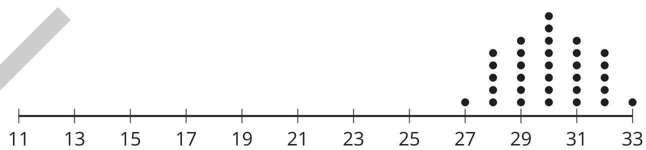
distribution 1b

2a



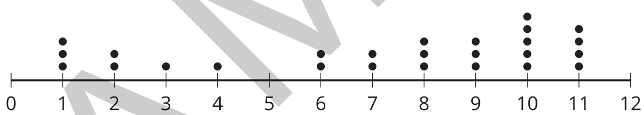
distribution 2a

2b



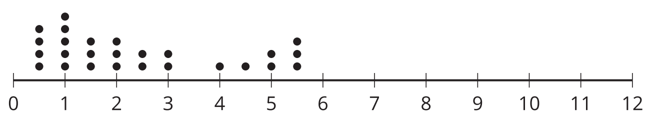
distribution 2b

3a



distribution 3a

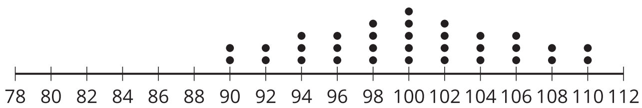
3b



distribution 3b

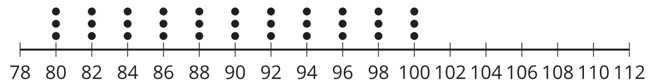
Sec D

4a



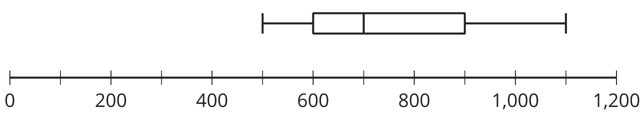
distribution 4a

4b



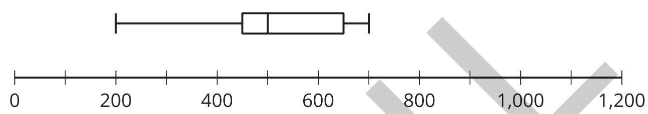
distribution 4b

5a



distribution 5a

5b



distribution 5b

6a

A political podcast has mostly reviews that either love the podcast or hate it.

6b

A cooking podcast has reviews that neither hate nor love the podcast.

7a

Stress testing concrete from site A has all 12 samples break at 450 pounds per square inch (psi).

7b

Stress testing concrete from site B has samples break every 10 psi starting at 450 psi until the last core is broken at 560 psi.

7c

Stress testing concrete from site C has 6 samples break at 430 psi and the other 6 break at 460 psi.

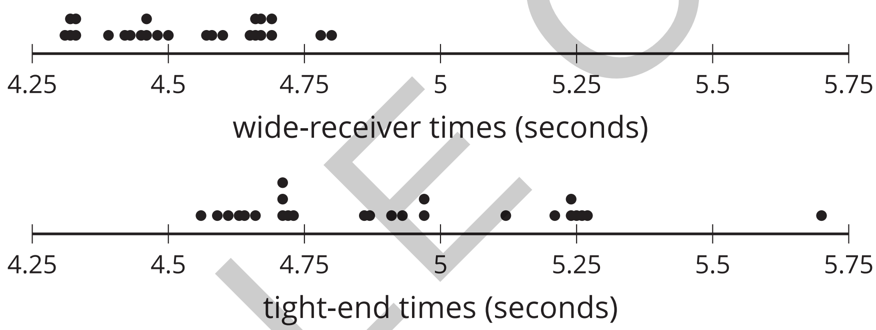
Lesson 15 Summary

To compare data sets, it is helpful to look at the measures of center and measures of variability. The shape of the distribution can help choose the most useful measure of center and measure of variability.

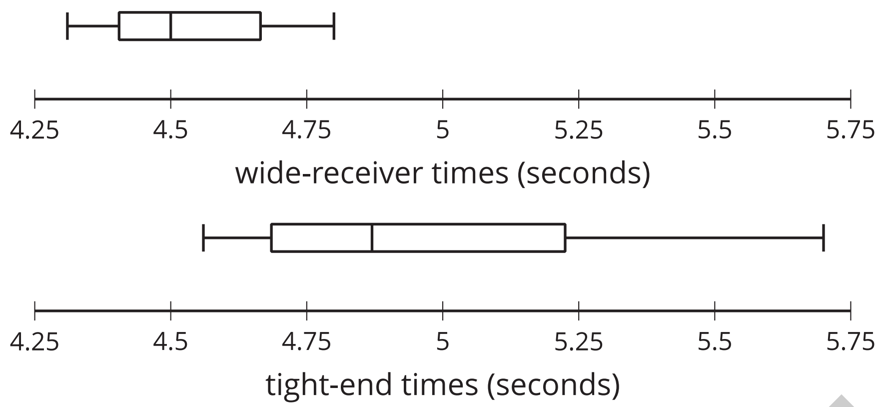
When distributions are symmetric or approximately symmetric, the mean is the preferred measure of center and should be paired with the standard deviation as the preferred measure of variability. When distributions are skewed or when outliers are present, the median is usually a better measure of center and should be paired with the interquartile range (IQR) as the preferred measure of variability.

Once the appropriate measure of center and measure of variability are selected, these measures can be compared for data sets with similar shapes.

For example, let’s compare the number of seconds it takes football players to complete a 40-yard dash at two different positions. First, we can look at a dot plot of the data to see that the tight-end times do not seem distributed symmetrically, so we should probably find the median and IQR for both sets of data to compare information.



The median and IQR could be computed from the values, but can also be determined from a box plot.

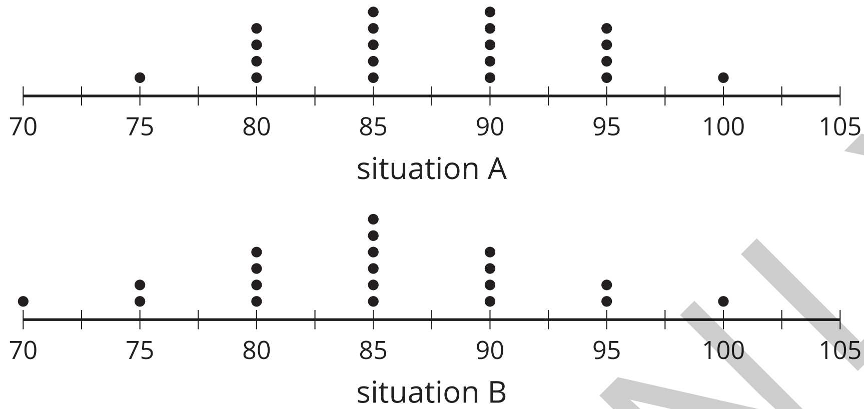


This shows that the tight-end times have a greater median (about 4.9 seconds) compared to the median of wide-receiver times (about 4.5 seconds). The IQR is also greater for the tight-end times (about 0.5 seconds) compared to the IQR for the wide-receiver times (about 0.25 seconds).

This means that the tight ends tend to be slower in the 40-yard dash when compared to the wide receivers. The tight ends also have greater variability in their times. Together, this can be taken to mean that, in general, a typical wide receiver is faster than a typical tight end is, and the wide receivers tend to have more similar times to one another than the tight ends do to one another.

Practice Problems

1 Twenty students participated in a psychology experiment that measured their heart rates in two different situations.



a. What are the appropriate measures of center and variability to use with the data? Explain your reasoning.

b. Which situation shows a greater typical heart rate?

c. Which situation shows greater variability?

2 a. Invent two situations that you think would result in distributions with similar measures of variability. Explain your reasoning.

b. Invent two situations that you think would result in distributions with different measures of variability. Explain your reasoning.

3 The data set and some summary statistics are listed.

11.5 12.3 13.5 15.6 16.7 17.2 18.4 19 19.5 21.5

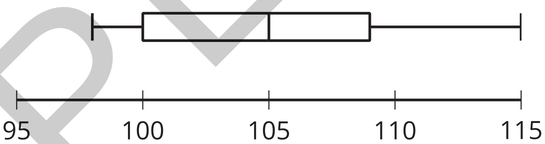
- mean: 16.52
- median: 16.95
- standard deviation: 3.11
- IQR: 5.5

a. How does adding 5 to each of the values in the data set affect the shape of the distribution?

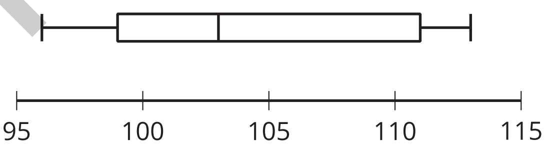
b. How does adding 5 to each of the values in the data set affect the measures of center?

c. How does adding 5 to each of the values in the data set affect the measures of variability?

4 Here are two box plots:



box plot A



box plot B

a. Which box plot has a greater median?

b. Which box plot has a greater measure of variability?

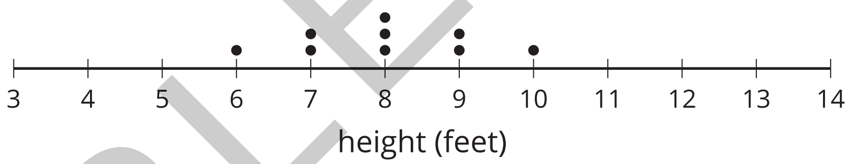
5 from Unit 1, Lesson 13

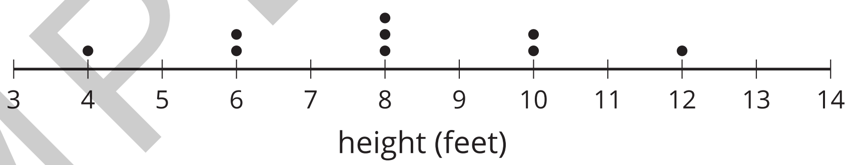
The depth of two lakes is measured at multiple spots. For the first lake, the mean depth is about 45 feet with a standard deviation of 8 feet. For the second lake, the mean depth is about 60 feet with a standard deviation of 27 feet.

Noah says the second lake is generally deeper than the first lake. Do you agree with Noah?

6 from Unit 1, Lesson 12

The dot plots display the height, rounded to the nearest foot, of maple trees from two different tree farms.



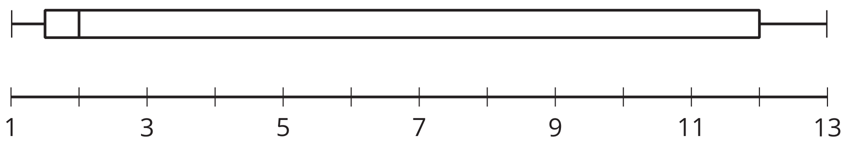


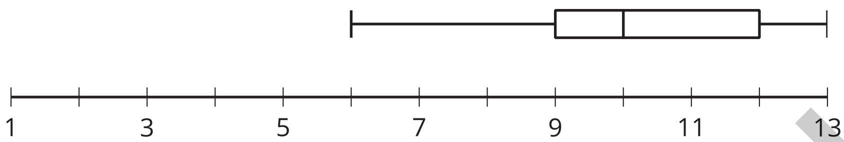
a. Compare the mean and standard deviation of the two data sets.

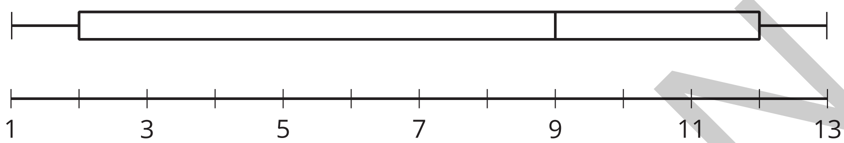
b. What does the standard deviation tell you about the trees at these farms?

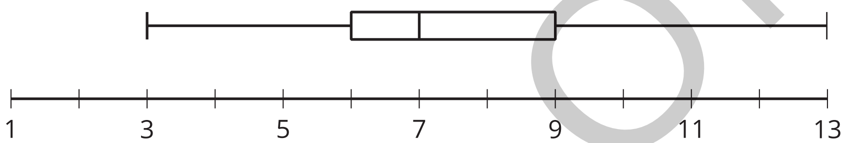
7 from Unit 1, Lesson 11

Which box plot has an IQR of 10?

A.


B.


C.


D.


Sec D

8 from Unit 1, Lesson 9

What effect does eliminating the lowest value, -6, from the data set have on the mean and median?

-6 3 3 3 3 5 6 6 8 10

Analyzing Data

Let's answer statistical questions by analyzing data, and comparing and contrasting their distribution shape and measures of center and variability.

16.1 Experimental Conditions

To test reaction time, the person running the test will hold a ruler at the 12 inch mark. The person whose reaction time is being tested will hold their thumb and forefinger open on either side of the flat side of the ruler at the 0 inch mark on the other side of the ruler. The person running the test will drop the ruler and the other person should close their fingers as soon as they notice the ruler moving to catch it. The distance that the ruler fell should be used as the data for this experiment.

With your partner, write a statistical question that can be answered by comparing data from two different conditions for the test.

16.2 Dropping the Ruler

Earlier, you and your partner agreed on a statistical question that can be answered using data collected in 2 different ruler-dropping conditions. With your partner, run the experiment to collect at least 20 results under each condition.

Analyze your 2 data sets to compare the distributions from the statistical questions. Next, create a visual display that includes:

- Your statistical question.
- The data that you collected.
- A data display.
- The measure of center and variability that you found that are appropriate for the data.
- An answer to the statistical question with any supporting mathematical work.

16.3 Heights and Handedness

Is there a connection between a student's dominant hand and their height? Use the table of information to compare the size of students with different dominant hands.

Practice Problems

1 from Unit 1, Lesson 15

Here are the statistics for the high temperatures in a city during October:

- mean of 65.3 degrees Fahrenheit
- median of 63.5 degrees Fahrenheit
- standard deviation of 9.3 degrees Fahrenheit
- IQR of 7.1 degrees Fahrenheit

Recall that the temperature C, measured in degrees Celsius, is related to the temperature F, measured in degrees Fahrenheit, by $C = \frac{5}{9}(F - 32)$.

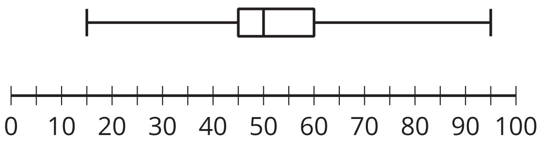
a. Describe how the value of each statistic changes when 32 is subtracted from the temperature in degrees Fahrenheit.

b. Describe how the value of each statistic further changes when the new values are multiplied by $\frac{5}{9}$.

c. Describe how to find the value of each statistic when the temperature is measured in degrees Celsius.

2 from Unit 1, Lesson 15

Here is a box plot.



Give an example of a box plot that has a greater median and a greater measure of variability, but the same minimum and maximum values.

3 from Unit 1, Lesson 14

The mean vitamin C level for 20 dogs was 7.6 milligrams per liter, with a standard deviation of 2.1 milligrams per liter.

One dog's vitamin C level was not in the normal range. It was 0.9 milligrams per liter, which is a very low level of vitamin C.

a. If the value 0.9 is eliminated from the data set, does the mean increase or decrease?

b. If the value 0.9 is eliminated from the data set, does the standard deviation increase or decrease?

Sec E

4 from Unit 1, Lesson 14

The data set represents the number of hours that fifteen students walked during a two-week period.

6 6 7 8 8 8 9 10 10 12 13 14 15 16 30

The median is 10 hours, Q1 is 8, Q3 is 14, and the IQR is 6 hours. Are there any outliers in the data? Explain or show your reasoning.

5 from Unit 1, Lesson 14

Here are some summary statistics about the number of accounts that follow some bands on social media.

- mean: 15,976 followers
- median: 16,432 followers
- standard deviation: 3,279 followers
- Q1: 13,796
- Q3: 19,070
- IQR: 5,274 followers

a. Give an example of a number of followers that a very popular band might have that would be considered an outlier for this data. Explain or show your reasoning.

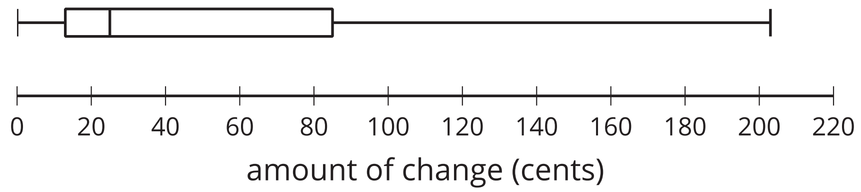
b. Give an example of a number of followers that a relatively unknown band might have that would be considered an outlier for this data. Explain or show your reasoning.

6 from Unit 1, Lesson 13

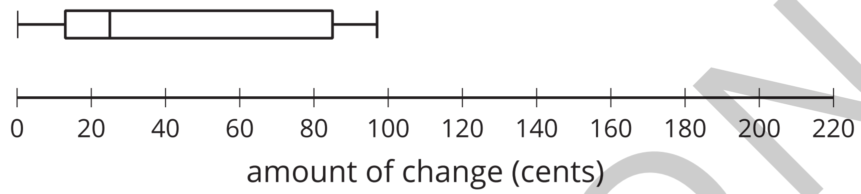
The weights of one population of brown bears have a mean of 428 pounds and standard deviation of 36 pounds. The weights of another population of brown bears have a mean of 397 pounds and standard deviation of 25 pounds. Andre says the two populations are similar. Do you agree? Explain your reasoning.

7 from Unit 1, Lesson 10

The box plot represents the distribution of the amount of change, in cents, that 50 people were carrying when surveyed.



The box plot represents the distribution of the same data set, but with the maximum, 203, removed.



The median is 25 cents for both plots. After examining the data, the value 203 is removed because it was an error in recording.

a. Explain why the median remains the same when 203 cents was removed from the data set.

b. When 203 cents is removed from the data set, does the mean remain the same? Explain your reasoning.

Sec E

Learning Targets

Lesson 1 Getting to Know You

- I can tell statistical questions from non-statistical questions and can explain the difference.
- I can tell the difference between numerical and categorical data.

Lesson 2 Data Representations

- I can find the five-number summary for data.
- I can use a dot plot, histogram, or box plot to represent data.

Lesson 3 A Gallery of Data

- I can graphically represent the data that I collected and critique the representations of others' data.

Lesson 4 The Shape of Distributions

- I can use a graphical representation of data to suggest a situation that produced the data pictured.
- I can use the terms "symmetric," "skewed," "uniform," "bimodal," and "bell-shaped" to describe the shape of a distribution.

Lesson 5 Calculating Measures of Center and Variability

- I can calculate mean absolute deviation, interquartile range, mean, and median for a set of data.

Lesson 6 Mystery Computations

- I can determine basic relationships between cell values in a spreadsheet by changing the values and noticing what happens in another cell.

Lesson 7 Spreadsheet Computations

- I can use a spreadsheet as a calculator to find solutions to word problems.

Lesson 8 Spreadsheet Shortcuts

- I can use shortcuts to fill in cells on a spreadsheet.

Lesson 9 Technological Graphing

- I can use technology to create graphic representations of data and calculate statistics.

Lesson 10 The Effect of Extremes

- I can describe how an extreme value will affect the mean and median.

- I can use the shape of a distribution to compare the mean and median.

Lesson 11 Comparing and Contrasting Data Distributions

- I can arrange data sets in order of variability, given graphic representations.

Lesson 12 Standard Deviation

- I can describe standard deviation as a measure of variability.
- I can use technology to compute standard deviation.

Lesson 13 More Standard Deviation

- I can use standard deviation to say something about a situation.

Lesson 14 Outliers

- I can find values that are outliers, investigate their source, and figure out what to do with them.
- I can tell how an outlier affects mean, median, IQR, or standard deviation.

Lesson 15 Comparing Data Sets

- I can compare and contrast situations using measures of center and measures of variability.

Lesson 16 Analyzing Data

- I can collect data from an experiment and compare the results using measures of center and measures of variability.

Illustrative Mathematics® LEARN MATH FOR LIFE | Kendall Hunt

ALGEBRA 1

Teacher Guide

UNIT 1

Book 1
Certified by Illustrative Mathematics®

ISBN 9798385123070

HS_vII

20240203

Table of Contents

UNIT 1: ONE-VARIABLE STATISTICS

UNIT 2: LINEAR EQUATIONS AND SYSTEMS

Unit 1: One-Variable Statistics

Unit Narrative

In this unit, students collect, display, and analyze data using statistics such as mean, median, interquartile range, and standard deviation.

In grades 6–8, students used histograms, dot plots, and box plots as a way to summarize data and worked with basic measures of center (mean and median) as well as measures of variability (mean absolute deviation and interquartile range). These concepts are revisited in the first two sections of this unit, but with a focus on interpretation and what they reveal about the data in addition to the mechanics of constructing the data displays.

The optional third section is available to familiarize students with spreadsheets and technology that will be used to calculate statistics such as mean, median, quartiles, and standard deviation as well as create data displays.

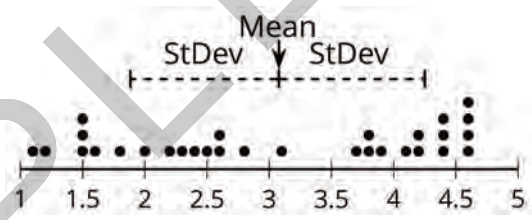
The fourth section introduces additional ways to interpret data using standard deviation and outliers. They finish the unit by using these tools to compare related data sets using measures of center and measures of variability.

The last lesson gives students a chance to practice their skills by posing a statistical question, designing an experiment, collecting data, and analyzing their data.

Because the first half of the unit mostly revisits material from middle school, a Mid-unit Assessment is not included in this unit. The *Cool-downs* and *Checkpoints* can be used to monitor student understanding.

In this unit, only the population standard deviation is used. Sample standard deviation is introduced in a later course.

Geogebra's spreadsheets are chosen for their versatility for the on-level mathematics in this course. While other spreadsheet programs have additional functionality and uses, they are limited in other ways. That said, please adapt the materials to the needs of your students.



Materials To Gather

- Chart paper
- Internet-enabled device
- Math Community Chart
- Rulers
- Spreadsheet technology
- Statistical technology
- Sticky notes
- Tools for creating a visual display
  Any way for students to create work that can be easily displayed to the class. Examples: Chart paper and markers, whiteboard space and markers, shared online drawing tool, access to a document camera.

Materials Needed

Lesson	Materials to Gather	Materials to Copy
Lesson 1	• Chart paper: Activity 1 • Math Community Chart: Activity 1 • Sticky notes: Activity 1	• 6–12 Blank Math Community Chart (1 copy for every 30 students): Activity 1
Lesson 2		
Lesson 3	• Math Community Chart: Activity 1 • Sticky notes: Activity 1 • Tools for creating a visual display: Activity 2	
Lesson 4		• Matching Distributions Cards (1 copy for every 2 students): Activity 2
Lesson 5		• Heartbeats Part 1 Handout (1 copy for every 2 students): Activity 2 • Algebra 1 Unit 1 Useful Terms and Displays (1 copy for every 30 students): Activity 3
Lesson 6	• Internet-enabled device: Activity 2, Activity 3 • Spreadsheet technology: Activity 2, Activity 3	

Lesson 7	• Internet-enabled device: Activity 2, Activity 3, Activity 4 • Spreadsheet technology: Activity 2, Activity 3, Activity 4	
Lesson 8	• Internet-enabled device: Activity 2, Activity 3 • Spreadsheet technology: Activity 2, Activity 3	
Lesson 9	• Math Community Chart: Activity 1 • Statistical technology: Activity 1, Activity 2	
Lesson 10	• Internet-enabled device: Activity 2 • Statistical technology: Activity 2	
Lesson 11	• Math Community Chart: Lesson • Math Community Chart: Activity 1 • Sticky notes: Activity 1	• Describing Data Distributions Cards (1 copy for every 2 students): Activity 2
Lesson 12	• Math Community Chart: Activity 1 • Statistical technology: Activity 2, Activity 3	• Algebra 1 Unit 1 Useful Terms and Displays (1 copy for every 30 students): Activity 3
Lesson 13	• Chart paper: Activity 1 • Math Community Chart: Activity 1	• African and Asian Elephants Cards (1 copy for every 2 students): Activity 2

Lesson 14	• Statistical technology: Lesson • Statistical technology: Activity 2	• Algebra 1 Unit 1 Useful Terms and Displays (1 copy for every 30 students): Activity 1
Lesson 15		
Lesson 16	• Rulers: Activity 1	• Heights and Handedness Handout (1 copy for every 2 students): Activity 3

Check Your Readiness

1

Standards

Addressing **6.SP.A.2, 6.SP.B.5.a**

Narrative

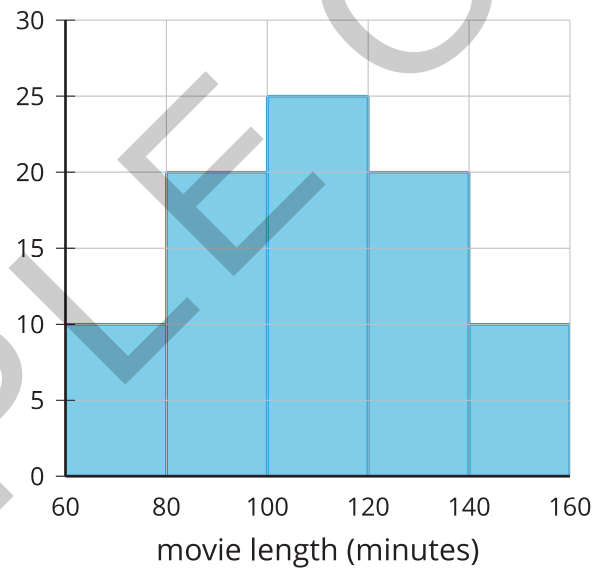
The content assessed in this problem is first encountered in Lesson 2: Data Representations.

This item will help assess how familiar students are with the information displayed in a histogram. Students need to be able to understand data summarized in a histogram to be successful when creating histograms in Lesson 2.

If most students do well with this item, plan to move quickly through Lesson 2, Activity 2, in which students interpret histograms.

Student Task Statement

Andre collected data on the length, in minutes, of some films. This is a histogram summarizing his data.



a. How many films are in Andre's data set?

b. Describe the shape of the distribution.

Solution

a. 85 films

b. Sample response: The distribution is approximately symmetric and bell-shaped.

2

Standards

Addressing **6.SP.B.5.c**

Narrative

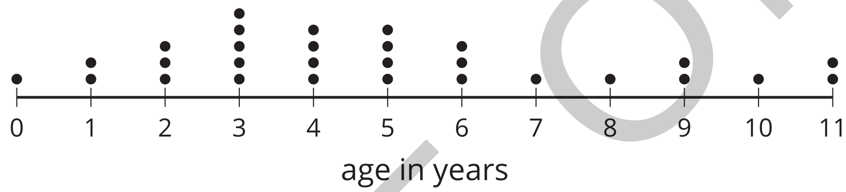
The content assessed in this problem is first encountered in Lesson 5: Calculating Measures of Center and Variability.

This item assesses whether students understand the difference between mean and median and can compute each from information given in a dot plot.

If most students struggle with this item, plan to spend more time with the *Warm-up* of Lesson 5. If additional aid is needed, consider revisiting the data from Lesson 2, and compute the measures of center.

Student Task Statement

This dot plot shows the age, in years, of 29 dogs at a dog park.



a. What is the mean age of dogs at the dog park?

b. Explain how you found the value of the mean.

c. What is the median age of dogs at the dog park?

d. Explain how you found the value of the median.

Solution

a. About 4.90 years old

b. Sample response: I added all of the values for the data together and divided by 29, the number of dogs at the dog park.

c. 4 years old

d. Sample response: I found the number in the middle of the data by crossing out pairs of high and low values until only 1 value in the middle was left.

3

Standards

Addressing **6.SP.B.5.c**

Narrative

The content assessed in this problem is first encountered in Lesson 2: Data Representations.

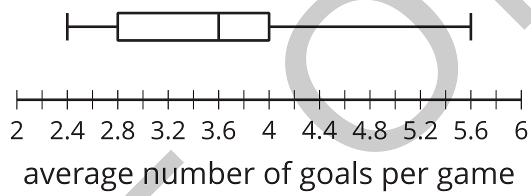
This unit will explore measures of spread, including IQR (interquartile range), to which students were introduced in grade 6. Students need to understand the information that can be gained from a box plot to be successful when they begin Lesson 2.

Students who select choice B are likely mistaking the range for the IQR. Students who select choice C are likely mistaking the median for the IQR. Students who select choice D are likely mistaking the mean for the median. It may be possible that the mean is 3.6 goals per game, but that cannot be determined from the box plot alone. Students who select choice F are likely mistaking Q1 for the minimum.

If most students struggle with this item, use Lesson 2, Activity 3 to familiarize students with the five-number summary and its connection to box plots.

Student Task Statement

The average goals scored per game are calculated for 20 soccer tournaments. The 20 averages are used to create this box plot.



Select **all** statements that must be true.

A. The interquartile range (IQR) is 1.2 goals per game.

B. The interquartile range (IQR) is 3.2 goals per game.

C. The interquartile range (IQR) is 3.6 goals per game.

D. The mean is 3.6 goals per game.

E. The median is 3.6 goals per game.

F. The minimum is 2.8 goals per game.

G. The maximum is 5.6 goals per game.

Solution

A, E, G

4

Standards

Addressing **6.SP.B.4, HSS-ID.A.1**

Narrative

The content assessed in this problem is first encountered in Lesson 2: Data Representations.

A key goal of this unit is for students to be able to represent distributions of data graphically. This problem assesses whether students can create a dot plot, box plot, and histogram from a list of data.

If most students struggle with this item, use optional Lesson 2 to revisit how to draw dot plots, box plots, and histograms.

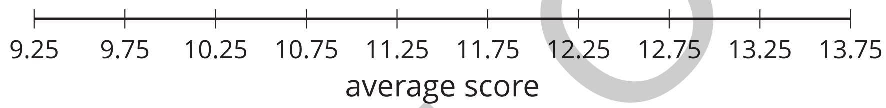
Student Task Statement

Four judges score dancers on their performance. The averages of the four judges' scores for 19 dancers are listed.

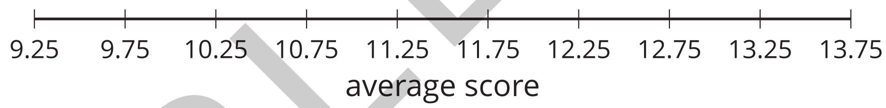
9.25 9.25 9.75 10.25 10.75 10.75 11.25 11.75 11.75 11.75 12.25 12.25 12.25

12.25 12.75 12.75 13.25 13.75 13.75

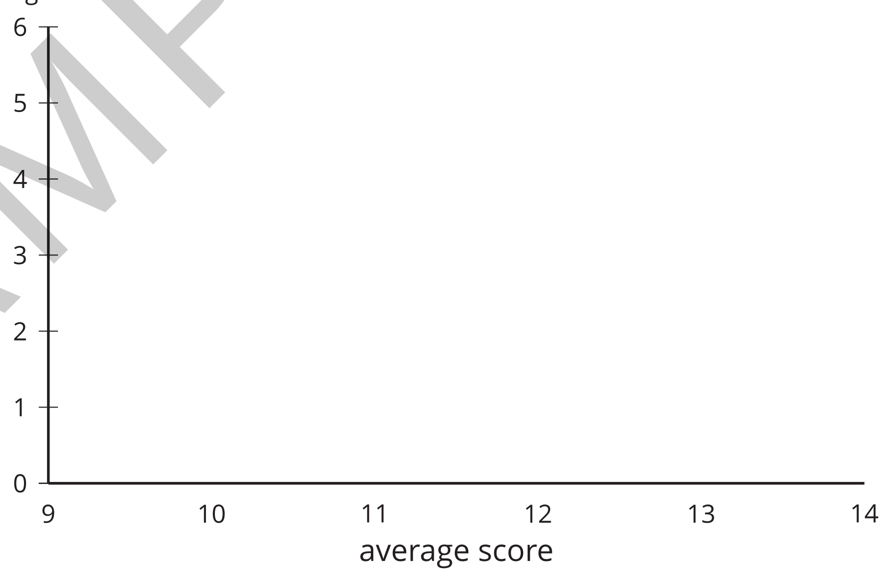
a. Create a dot plot that represents the distribution of the data.



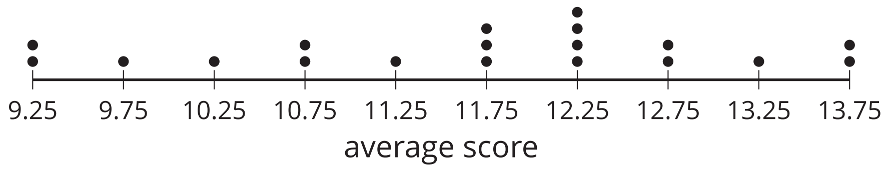
b. Create a box plot that summarizes the data.



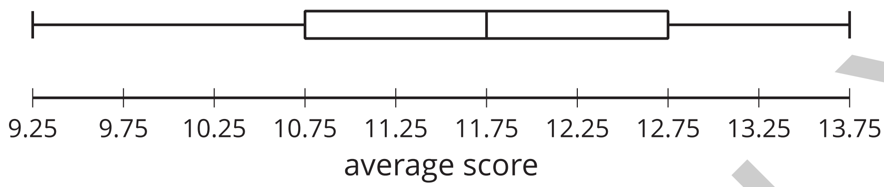
c. Create a histogram that summarizes the data.



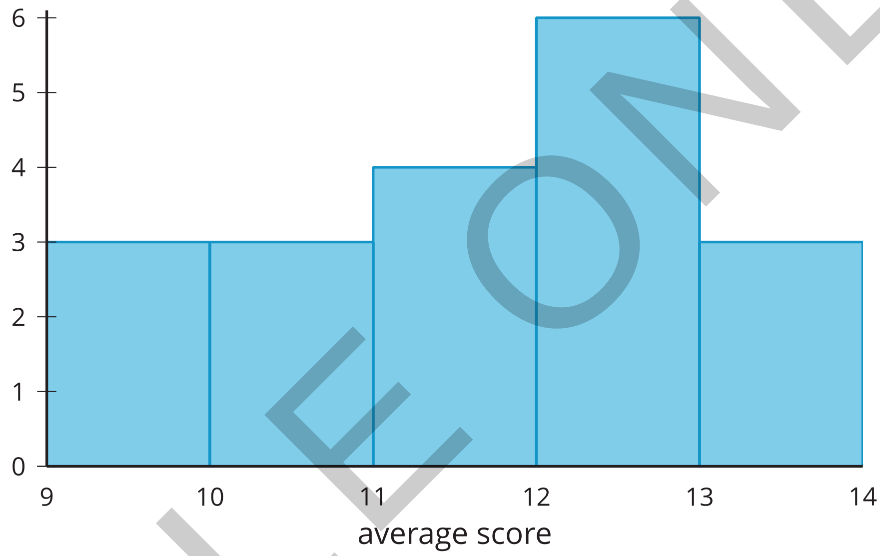
Solution



a.



b.



c.

5

Standards

Addressing **6.RP.A.3.c**

Narrative

The content assessed in this problem is first encountered in Lesson 5: Calculating Measures of Center and Variability.

In order to be able to work with quartiles and relative frequencies, students need to be comfortable working with percentages.

If most students struggle with this item, spend sufficient time in Lesson 5 working with percentages to understand quartiles.

Student Task Statement

There are 12 students in a class who are performing at the school concert. This represents 40% of the class. How many students are in the class?

Solution

30

6

Standards

Addressing **6.SP.A.3, 6.SP.B.5.c**

Narrative

The content assessed in this problem is first encountered in Lesson 5: Calculating Measures of Center and Variability.

A new measure of spread, standard deviation, is introduced in this unit. This concept will build upon students' understanding of mean absolute deviation (MAD). This problem assesses their ability to calculate and interpret MAD.

If most students struggle with this item, use optional Lesson 5, Activity 3 to revisit how to calculate mean and MAD.

Student Task Statement

The daily high temperatures, in degrees Celsius, for two weeks were recorded. The first week had a mean high temperature of 8 degrees Celsius and a mean absolute deviation (MAD) of 1 degree Celsius. Here is a list of the high temperatures, in degrees Celsius, for the second week.

5 9 7 0 3 5 5

a. What is the mean temperature for the second week? (Round to the nearest hundredths place.)
b. What is the mean absolute deviation of the second week? (Round to the nearest hundredths place.)
c. Which week had greater variability in high temperature? Explain your reasoning.

Solution

a. 4.86 degrees Celsius
b. 1.92 degrees Celsius
c. The second week had greater variability in high temperature because it had a larger MAD.

7

Standards

Addressing **6.SP.B.5**

Narrative

The content assessed in this problem is first encountered in Lesson 5: Calculating Measures of Center and Variability.

Students need to understand measures of center (mean and median) and measures of variability (interquartile range and mean absolute deviation) to be successful.

If most students struggle with this item, spend additional time with optional Lesson 5 to discuss the meaning of measures of center and measures of variability in context.

Student Task Statement

For a science project, a student plants seeds of 2 different varieties of corn plants. The table displays the heights, in inches, of each of the corn plants after 8 weeks.

variety A	46	46	47	48	49	49	50	6	48	49
variety B	48	46	45	47	46	48	49	47	48	48

a. How many plants of each variety were planted?

b. The mean height of the corn plants of variety A is 43.8 inches. Interpret the value of the mean in the situation.

c. The median height of the corn plants of variety A is 48 inches. Interpret the value of the median in the situation.

d. For the corn plants in variety A, describe the overall pattern of corn plant heights and any striking deviations from the overall pattern of corn plant heights.

e. For the corn plants in variety B, describe the overall pattern of corn plant heights and any striking deviations from the overall pattern of corn plant heights.

f. The interquartile range of the corn plant heights of variety A is 3 inches. The interquartile range of the corn plant heights of variety B is 2 inches. What does the interquartile range tell you about the data for each variety of corn?

g. Which variety of corn plants shows a greater spread of its distribution of heights? Explain your reasoning.

Solution

a. 10

b. Sample response: This means that the average height of the corn plants is 43.8 inches. If all ten plants were put in a line and then the line was cut into 10 equal sections, each section would be 43.8 inches long.

c. Sample response: The median of 48 means that half of the plants are less than or equal to 48 inches tall and half are greater than or equal to 48 inches tall.

d. Sample response: The heights are mostly between 46 and 50 inches tall with the exception of the 6-inch tall plant.

e. Sample response: The heights are all between 45 and 49 inches.

f. Sample response: The interquartile range describes how different the heights for the middle 50% of the plants can be for each variety of corn plant. They describe the variability of the middle half of the data.

g. Variety A has a greater spread in its distribution of heights because the interquartile range is greater than the heights for variety B.

8

Standards

Addressing **6.SP.B.5.b**

Narrative

The content assessed in this problem is first encountered in Lesson 5: Calculating Measures of Center and Variability.

If most students struggle with this item, use optional Lesson 5, Activity 3 to interpret MAD in context.

Student Task Statement

Chicken eggs can be categorized as large if they weigh at least 2 ounces. Clare weighs 48 large eggs and finds that they have a mean weight of 2.1 ounces and a mean absolute deviation of 0.08 ounces. Interpret 0.08 ounces in this situation.

Solution

Sample response: A MAD of 0.08 ounces means that, on average, the 48 eggs are within 0.08 ounces of the mean or are between 2.02 and 2.18 ounces.

End-of-Unit Assessment

1

Standards

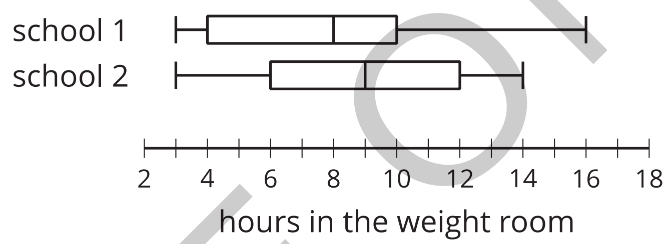
Addressing **HSS-ID.A.2**

Narrative

Students use a box plot to compare the median and interquartile range for two populations.

Student Task Statement

The two box plots summarize the number of hours spent in the weight room for all the players on the football team for two different high schools. Which of the statements must be true about the distribution of data represented in the boxplots?



A. Players at school 1 typically spent more time in the weight room than players at school 2.

B. The middle half of the data for school 1 has more variability than the middle half of the data for school 2.

C. The median hours spent in the weight room for school 1 is less than the median for school 2 and the interquartile ranges for both schools are equal.

D. The total number of hours spent in the weight room for players at school 2 is greater than the total number of hours for players at school 1.

Solution

C

2

Standards

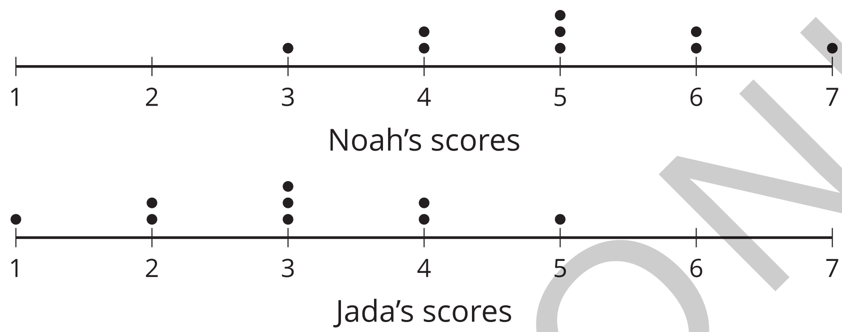
Addressing **HSS-ID.A.2**

Narrative

Choice A is incorrect because Noah’s distribution is the same as Jada’s. Although the data points are shifted further to the right on Noah’s plot, the variability is the same. Choice D is incorrect because even though the mean for Noah’s scores was higher than the mean for Jada’s scores, the data overlap and some of Jada’s scores are higher than some of Noah’s scores.

Student Task Statement

The dot plots show 9 scores on a 10-question trivia game for two students. Select **all** the statements that must be true.



A. Noah’s scores have greater variability than Jada’s scores.

B. The standard deviation of Noah’s scores is equal to the standard deviation of Jada’s scores.

C. The mean of Noah’s scores is greater than the mean of Jada’s scores.

D. Noah scored better than Jada on every trivia game.

E. Using only Noah’s scores, the mean is equal to the median.

Solution

B, C, E

3

Standards

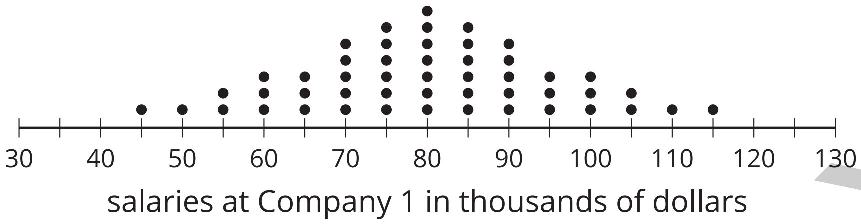
Addressing **HSS-ID.A.3**

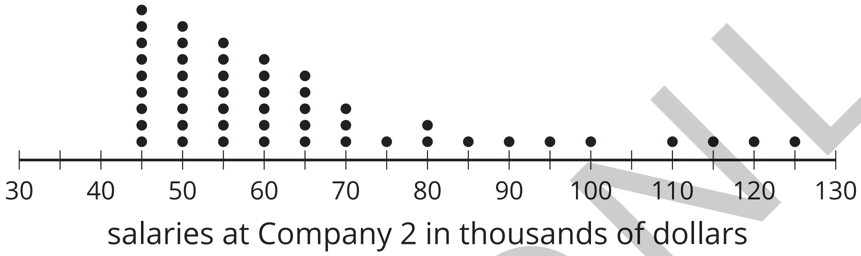
Narrative

Students who select choice A may not realize that the data for Company 2 are skewed and that the mean and median are not the same like they are for Company 1. Students who select choice C might be using the center of the range rather than the actual median.

Student Task Statement

The dot plots show the salaries for the employees at two small companies before a new company head is hired at each company.





Select **all** the statements that must be true.

A. The mean and median values of salaries shown in the dot plot for Company 2 are the same.

B. The standard deviation of salaries shown in the dot plot for Company 2 is greater than the standard deviation of salaries at Company 1.

C. The median values of salaries shown in the dot plots for the two companies are about the same.

D. The median value of salaries at Company 1 will remain unchanged after a company head is hired to have a salary of 500 thousand dollars.

E. The new company head with a salary of 500 thousand dollars will affect the mean value for salaries at Company 2 more than the median value for salaries at Company 2.

Solution

B, D, E

4

Standards

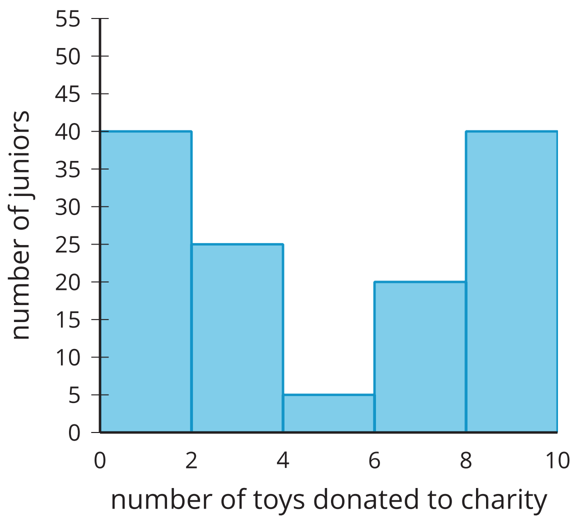
Addressing **HSS-ID.A.2**

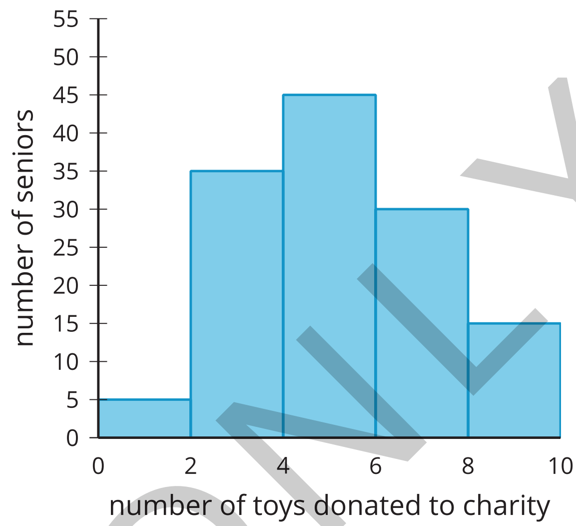
Narrative

Students should make the connection between standard deviation and variability and be able to interpret visual representations of variability in distributions.

Student Task Statement

A school has a toy drive for a holiday in which students bring in toys to be donated to charity. The number of toys donated by juniors and seniors are summarized in the histograms.





Is the standard deviation of the number of toys donated greater for seniors or juniors?

Solution

juniors

5

Standards

Addressing **HSS-ID.A.1, HSS-ID.A.2**

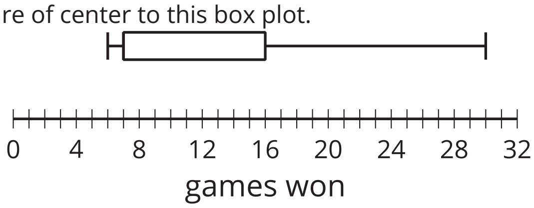
Student Task Statement

The chess club at a school has 15 members. The number of games won in tournament play this season by each member is listed.

6 6 6 7 10 11 12 13 14 14 15 16 18 18 30

a. What measure of center is most appropriate to use to describe a typical value for the data in this distribution? Explain your reasoning.

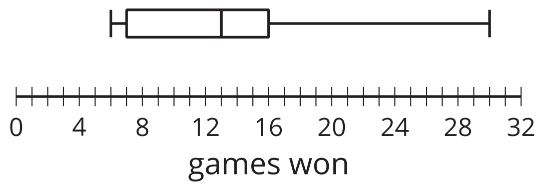
b. Add the appropriate measure of center to this box plot.



c. What measure is most appropriate for describing variability in this data distribution?

Solution

a. Median, since the data appear to be skewed to the right.



b.

c. Interquartile range.

Minimal Tier 1 response:

- Work is complete and correct.
- Sample:

a. See box plot.

b. Median, because the data are more spread out on one side.

c. Interquartile range.

Tier 2 response:

- Work shows good conceptual understanding and mastery, with either minor errors or correct work with insufficient explanation or justification.
- Sample errors: One or two of the 5 values in the box plot are incorrect; missing or incorrect answers to Parts b and/or c; answer to Part b is correct but work does not refer in some way to the shape of the distribution.

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.
- Sample errors: Two or more of the 5 values in the box plot are incorrect; two or more Tier 2 error types.

6

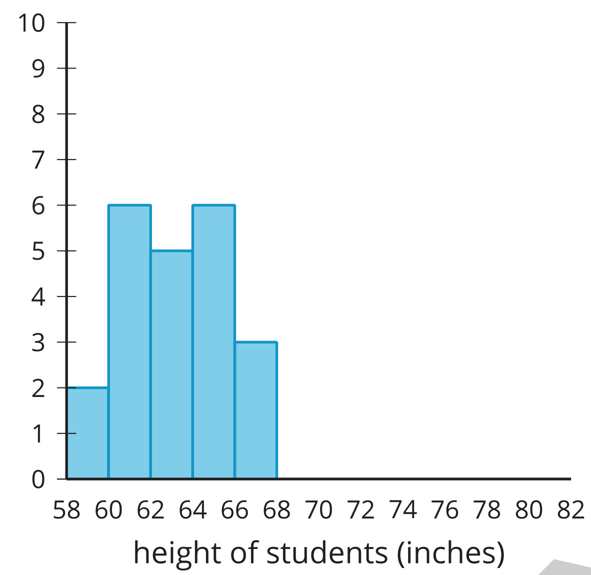
Standards

Addressing **HSS-ID.A.1, HSS-ID.A.3**

Student Task Statement

Diego arranges the students in his math class from shortest to tallest and measures the height in inches of each student in the class. The heights of the 22 shortest students are summarized in the histogram. The 8 tallest students have their heights recorded here.

73	73	73	75	75	77	79	81

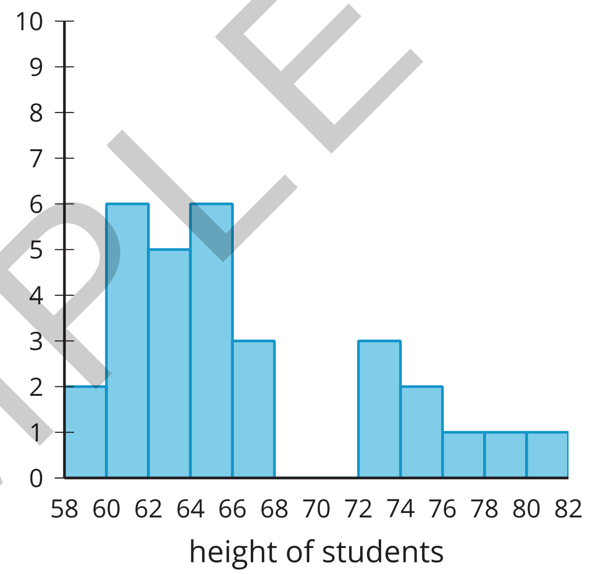


a. Complete the histogram using the data for the 8 tallest students in the class.

b. Use the shape of the distribution to compare the mean and median. Are the mean and median equal? If not, which is greater? Explain your reasoning.

Solution

a.



b. The mean is greater than the median. The distribution is skewed to the right, so the values greater than the typical values will pull the mean up to a higher value and not affect the median as much.

Minimal Tier 1 response:

- Work is complete and correct.
- Acceptable errors: One or two data points are omitted or incorrect, provided the work still shows understanding that repeated data points are stacked vertically.

- Sample:
  a. See graph.
  b. The mean, because the distribution is skewed to the right.

Tier 2 response:

- Work shows good conceptual understanding and mastery, with either minor errors or correct work with insufficient explanation or justification.
- Sample errors: Correct histogram with incorrect or poorly justified answer to Part b; reasonable answer to Part b based on an incorrect histogram; histogram has three or more incorrect or omitted data points.

Tier 3 response:

- Significant errors in work demonstrate lack of conceptual understanding or mastery.
- Sample errors: Two or more Tier 3 error types.

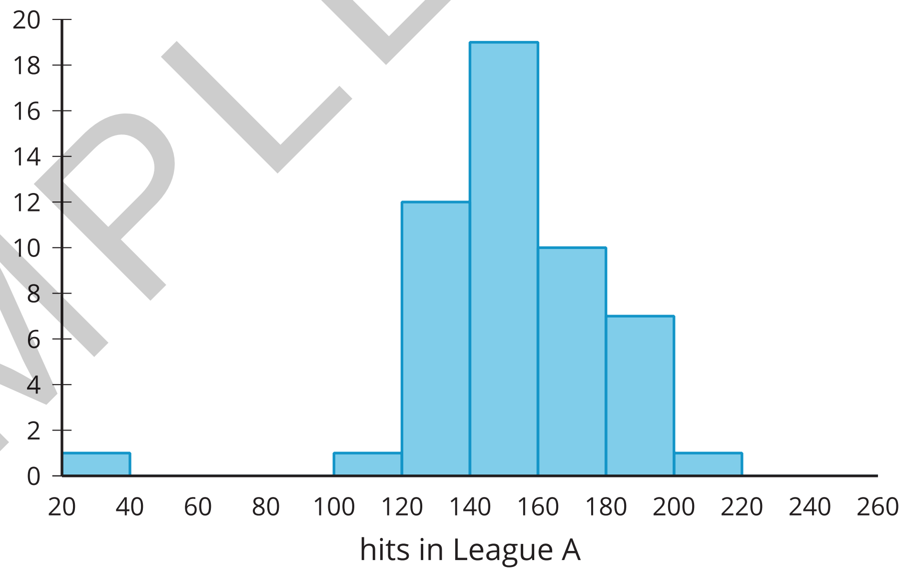
7

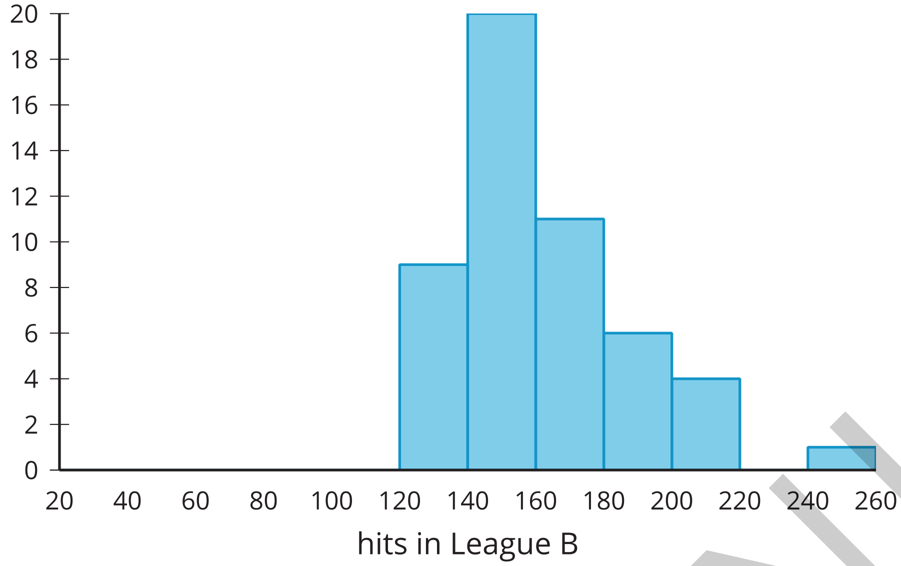
Standards

Addressing **HSS-ID.A.2, HSS-ID.A.3**

Student Task Statement

The histograms and summary statistics summarize the data for the number of hits in the season by baseball players in two leagues.





Some summary statistics for the number of hits by players in each league.

	mean	median	standard deviation	minimum	Q1	Q3	maximum
League A	151.12	148	26.83	29	136	167	207
League B	163.25	157	24.93	136	145	178	256

a. Each data set contains one outlier. Based on this information, what is the most appropriate measure of center? What is the most appropriate measure of variability?

b. Compare the number of hits by players in the two leagues using the measure of center and measure of variability that you selected.

c. The outlier for League A is 29. After looking back at the original data, it is discovered that this should be 129. Choose which values will be more likely to change significantly and explain your reasoning for each:
   - Mean or median
   - Standard deviation or interquartile range

Solution

Sample response:

a. Median and interquartile range

b. League B has a greater median number of hits (157 hits compared to 148 hits for League A) and a slightly greater interquartile range (33 hits compared to 31 hits for League A).

c. Mean and standard deviation. Sample reasoning: Because these measures use the value of each number in the data set, such a great change will affect the mean and standard deviation a lot. The median and IQR will likely be unchanged because the updated value is still in the lowest quarter of the data.

Minimal Tier 1 response:

- Work is complete and correct, with complete explanation or justification.

- Sample:

a. Median and interquartile range
b. League B has a greater median number of hits and greater interquartile range.
c. Mean and standard deviation because these measures use the value of each number in the data set.

Tier 2 response:

- Work shows good conceptual understanding and mastery, with either minor errors or correct work with insufficient explanation or justification.
- Sample errors: One of Parts b or c is correct but not justified or incorrectly justified, one of Parts b or c is incorrect but explanation demonstrates a general understanding of the concept, work for Part b has a minor mistake in the formula for determining outliers (with work shown), any incorrect answer to Part a.

Tier 3 response:

- Work shows a developing but incomplete conceptual understanding, with significant errors.
- Sample errors: One of Parts b or c is incorrect and not well justified, two Tier 2 error types, work shows a general lack of understanding of outliers.

Tier 4 response:

- Work includes major errors or omissions that demonstrate a lack of conceptual understanding and mastery.
- Sample errors: Three or more Tier 2 error types.

Section A: Getting to Know You

Goals

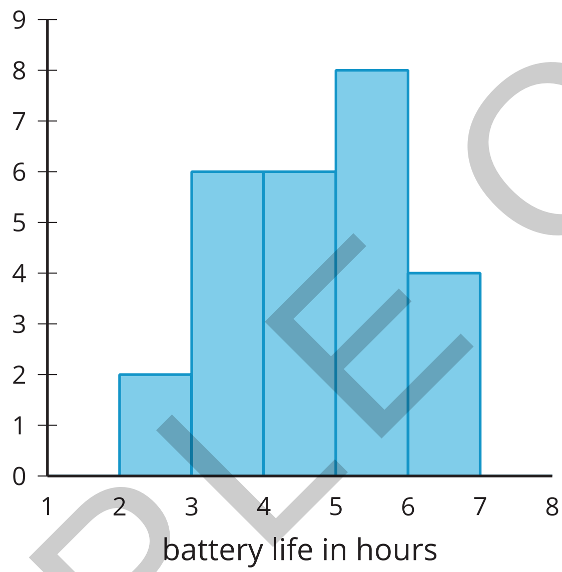
- Create and interpret data displays such as dot plots, histograms, and box plots.

Section Narrative

In this first section of Algebra 1, students begin to build community by collecting data from their classmates, classifying the different kinds of data they might collect, and creating displays of the information that they collected.

The optional second lesson gives students a chance to practice creating dot plots, histograms, and box plots using given data before attempting to do the same with the data that they collected. This lesson would be an opportunity to introduce any technological tools the class will use for creating these displays.

This section intentionally allows extra time for students to learn new routines and establish norms for the course.



Teacher Reflection Questions

- **Math Content and Student Thinking:** In an earlier grade, students created dot plots, box plots, and histograms. How did this experience help them approach these lessons?
- **Pedagogy:** Which curriculum resources are you using as you plan lessons? In what ways are they helping you use student thinking to drive the learning during the lessons?
- **Access and Equity:** How did you use the *Collect and Display* routines in this section to connect student language to new vocabulary or mathematical ideas?

Section A Checkpoint

1

Goals Assessed

- Create and interpret data displays such as dot plots, histograms, and box plots.

Student Task Statement

What do dot plots and box plots have in common? What is different?

Solution

Sample response:

Both dot plots and box plots can show data visually. They both easily show the maximum and minimum values. Dot plots show all of the data values while box plots do not. Box plots show the median, quartiles, and interquartile range while dot plots do not easily show those summary statistics.

Responding To Student Thinking

Points to Emphasize
If students struggle to compare and contrast dot plots and box plots, revisit the characteristics of each by asking students what information they can see in the different types of representations in the activity:

Algebra 1, Unit 1, Lesson 9, Activity 3 Making Digital Displays

Unit 1, Lesson 1

Getting to Know You

Goals

- Describe (orally and in writing) the difference between statistical and non-statistical questions.
- Describe (orally and in writing) the distinctions between numerical and categorical data.

Sec A

Learning Targets

- I can tell statistical questions from non-statistical questions and can explain the difference.
- I can tell the difference between numerical and categorical data.

Lesson Narrative

The mathematical purpose of this lesson is to understand what makes a question statistical and to classify data as numerical or categorical. **Numerical data** are responses to questions that are numbers that can be ordered in a natural way. **Categorical data** are responses to questions that fit into distinct categories.

Students learn to recognize **statistical questions** as questions that anticipate variability in the data. In this lesson, students recall the concept of variability to discuss the difference between statistical and **non-statistical questions** while they collect survey data from their classmates. (The data will be used again in later lessons, so the collected data should be kept in a spreadsheet or a folder.) Students classify questions as being statistical or non-statistical, and classify the data that they collect from statistical questions as numerical or categorical.

Students reason abstractly and quantitatively (MP2) as they classify data resulting from statistical questions as numerical or categorical because they must make sense of data in relation to the question being asked. They also refine their language from informal to more precise wording (MP6) for several vocabulary terms that will be used throughout the unit.

Math Community

This is the first exercise that focuses on the work of building a mathematical community. Students have the opportunity to think about what a mathematical community is and to share their initial thoughts about what it looks like and sounds like to do math together in a community.

Standards

Building On	**6.SP.A.1**
Building Towards	**HSS-ID.A.1**

Instructional Routines

- MLR2: Collect and Display

Required Materials

Materials To Gather

- Chart paper: Activity 1
- Math Community Chart: Activity 1
- Sticky notes: Activity 1

Materials To Copy

- 6–12 Blank Math Community Chart (1 copy for every 30 students): Activity 1

Required Preparation

Activity 1:

Make a space for students to place their sticky notes at the end of the *Warm-up*. For example, hang a sheet of chart paper on a wall near the door.

Lesson:

At the end of the "Representing Data About You and Your Classmates" activity, students will need to keep their data for use in a later lesson in which they will graphically represent the data collected in this activity. If they record the data in workbooks, it will be easy to retrieve later. If students record data some other way, be sure that your method allows them to easily retrieve the data later.

Sec A

Student Facing Learning Goals

Let's work together to collect data and explore statistical questions.

1.1 Types of Questions

Warm-up

5 mins

Activity Narrative

This *Warm-up* prompts students to compare four survey questions. It gives students a reason to use language precisely (MP6) and gives you the opportunity to hear how they use terminology and talk about characteristics of the items in comparison to one another.

Standards

Building On	**6.SP.A.1**
Building Towards	**HSS-ID.A.1**

Launch

Arrange students in groups of 2–4. Display the questions for all to see. Give students 1 minute of quiet think time and then time to share their thinking with their small group. In their small groups, tell each student to share the items in their group and the property that those items share that any other items do not.

Student Task Statement

Use the four questions to make a group of two related questions. What property do the items in the group share that the others do not?

- Question A: How many potato chips are in this bag of chips?
- Question B: What is the typical number of chips in a bag of chips?
- Question C: What type of chips are these?
- Question D: What type of chips do students in this class prefer?

Student Response

Sample responses:

- Group of A and C. There is only 1 correct answer to the question.
- Group of B and D. They require collecting and analyzing data to find a solution.
- Group of A and B. The solutions are numerical.
- Group of C and D. The solutions are categorical.

Activity Synthesis

Invite each group to share one reason why a particular set of two questions go together. Record and display the responses for all to see. After each response, ask the class if they agree or disagree. Since there is no single correct answer to the question of which questions go together, attend to students' explanations, and ensure the reasons given are correct.

During the discussion, prompt students to explain the meaning of any statistical terminology they use, such as "numerical data," "categorical data," or "average," and to clarify their reasoning as needed. Consider asking:

- "How do you know . . . ?"
- "What do you mean by . . . ?"
- "Can you say that in another way?"

Math Community

After the *Warm-up*, tell students that today is the start of planning the type of mathematical community they want to be a part of for this school year. The start of this work will take several weeks as the class gets to know one another, reflects on past classroom experiences, and shares their hopes for the year.

Display and read aloud the question "What do you think it should look like and sound like to do math together as a mathematical community?" Give students 2 minutes of quiet think time and then 1–2 minutes to share with a partner. Ask students to record their thoughts on sticky notes and then place the notes on the sheet of chart paper. Thank students for sharing their thoughts and tell them that the sticky notes will be collected into a class chart and used at the start of the next discussion.

After the lesson is complete, review the sticky notes to identify themes. Make a Math Community Chart to display in the classroom. See the blackline master Blank Math Community Chart for one way to set up this chart. Depending on resources and wall space, this may look like a chart paper hung on the wall, a regular sheet of paper to display using a document camera, or a digital version that can be projected. Add the identified themes from the students' sticky notes to the student section of the "Doing Math" column of the chart.

1.2 Representing Data About You and Your Classmates

25 mins

Activity Narrative

In this activity, students begin classifying questions as "statistical" or "non-statistical" and data as "numerical" or "categorical" while getting to know a little about their classmates.

Each group of four students is assigned three questions. One of the three they are assigned is a non-statistical question, one would generate numerical data, and one would generate categorical data. The questions could be changed to questions that are more relevant to students and would help students learn about each other, as long as each set of three questions contains one question of each of the three types described. Groups also generate a fourth question of their own that can be answered with data. First, the group comes up with four survey questions that they can ask their classmates to collect data about their four questions of interest. Then, they collect data from their classmates by asking the survey questions. Finally, they summarize their results to answer the four questions of interest and reflect on the nature of the different questions they attempted to answer.

In a later lesson, students will represent the distribution of data collected in this activity graphically. If they record the data in their workbooks, it will be easy to retrieve later. If students record data some other way, be sure that your method allows them to easily retrieve the data later.

This is the first time *Math Language Routine 2: Collect and Display* is suggested in this course. In this routine, the teacher circulates, listens, and jots down words, phrases, drawings, or writing that students use. The language collected is displayed visually for the whole class to use throughout the lesson and unit. The purpose of this routine is to capture a variety of students' words and phrases—including especially everyday or social language and non-English—in a display that students can refer to, build on, or make connections with during future discussions, and to increase students' awareness of language used in mathematics conversations.

Access for English Language Learners

This activity uses the *Collect and Display* math language routine to advance conversing and reading as students clarify, build on, or make connections to mathematical language.

Standards

Building Towards **HSS-ID.A.1**

Instructional Routines

- MLR2: Collect and Display

Launch

Arrange students in groups of 4. Assign each group one of the following sets of three questions.
Set A

1. On average, how many letters are in the preferred names of students in this class?
2. Which does the class prefer: reading a book or watching the movie version of that book?
3. How many periods (or blocks) have there been today before this math class?

Set B

1. On average, how many speed limit signs do students in our class see on the way to school?
2. Would the class rather have a snow day or a field-trip day?
3. In what year was the 13th Amendment ratified?

Set C

1. About how long did it take students in this class to get to school this morning?
2. Which combination does the class prefer: peanut butter and banana or strawberry and banana?
3. What is the lightest element listed on the periodic table?

Set D

1. On average, how many movies in a theater did each student in the class watch this summer?

2. Does the class prefer to write on paper with lines or without lines?
3. How many seats are in the classroom?

Give groups 2 minutes to complete the first 2 questions. Then provide 10-15 minutes for students to meet their classmates and collect data. If needed, remind students to move on to meet different students and collect additional data. Consider using a timer set to 45 seconds to keep conversations moving. After students have collected their data, tell them to return to their original groups to finish analyzing the data. Give students time to discuss the questions in their groups, and then pause for a whole-class discussion.

Sec A

Use *Collect and Display* to create a shared reference that captures students' developing mathematical language. Collect the language that students use to talk about the similarities and differences between the types of data collected. Display words, phrases, and sentences, such as "These are all numbers." or "This has only one answer."

Access for English Language Learners

Representation: Access for Perception. Read all questions aloud. Students who both listen to and read the information will benefit from extra processing time.
Supports accessibility for: Language

Student Task Statement

Your teacher will assign you a set of 3 questions.

1. Write an additional question of interest that requires data collected from the class to answer.
2. For each of the 4 questions of interest, write a survey question that will help you collect data from the class that can be analyzed to answer the question of interest. Ask the 4 survey questions to 15 classmates, and record their responses to collect data. Then return your group.
3. Summarize the data for each question in a sentence or two, and share the results with your group.
4. With your group, decide what the responses for the questions numbered 1 have in common. Then do the same for questions numbered 2 and 3.
5. Does the question you wrote fit best with the questions numbered 1, 2, or 3? Explain your reasoning.

responder's name	question 1 response	question 2 response	question 3 response	my question response

Student Response

Sample responses:

1. What is the most popular favorite color for students in this class?
2. How many letters are in your preferred name? Would you prefer reading a book or watching the movie version of that book? How many periods came before this class today? What is your favorite color?
3. The students in this class tend to have about five or six letters in their preferred name. Most students prefer reading the book. There have been two periods before this class. Red is the most popular favorite color for the class.
4. Questions that are numbered 1 are answered with numbers and may have different answers from different people. Questions that are numbered 2 are answered with words or phrases and may have different answers from different people. Questions that are numbered 3 have a single right answer (although most people did not know the answers).
5. The question "What is the most popular favorite color for students in this class?" fits best with questions that are numbered 2 because there are a variety of answers that can be put into categories, but are not connected to numbers.

Building on Student Thinking

Students may confuse statistical questions (or "questions of interest") with survey questions. Explain that the set of three questions are statistical questions that can be answered using the survey questions. For example, students may think they should ask each of their classmates for the average number of speed limit signs the class sees. However, students can ask each classmate "How many speed limit signs do you see on the way to school?" and collectively use the answers to this survey question to answer the statistical question about the average number of speed limit signs their classmates see on their way to school.

Sec A

Are You Ready for More?

1. Find a news article that uses numerical data to discuss a statistical question.
2. Find a news article that uses categorical data to discuss a statistical question.

Extension Student Response

Answers vary.

Activity Synthesis

Display the numbered questions from all of the sets for all to see.

Direct students' attention to the reference created using *Collect and Display*. Ask students to share their responses. Invite students to borrow language from the display as needed, and update the reference to include additional phrases as they respond. (For example, "All of these questions require numerical data to find an answer.")

Then tell students that we call data collected to answer questions of interest like those numbered 1 numerical data, and that we call data collected by questions like those numbered 2 categorical data. Questions like those numbered 3 are non-statistical questions because there will be no variability in the responses. Questions like those numbered 1 and 2 are called statistical questions because they require the collection of data and there is anticipated variability in the responses.

Ask students to sort the collected language from the display into either numerical or categorical data as well as either statistical or non-statistical questions. For example, "There is only 1 answer" would be categorized as "non-statistical."

Access for Students with Disabilities

Representation: Develop Language and Symbols. Create a display of important terms and vocabulary. Invite students to suggest language or diagrams to include that will support their understanding of numerical data, categorical data, statistical questions, and non-statistical questions.
Supports accessibility for: Conceptual Processing, Language

Lesson Synthesis

To promote student understanding of the differences between statistical and non-statistical questions and classifying data as numerical or categorical, ask:

- "What makes a question statistical?" (There is variability in the data collected.)
- "What is an example of a non-statistical question?" (What value for x makes the equation $x + 5 = 7$ true?)

- "What is an example of a statistical question that we have not used in class?" (On average, how many people eat breakfast every day?)
- "What type of data are collected to answer the statistical question, 'Would the class rather have pizza or donuts?'" (Categorical)
- "What is an example of a statistical question that results in numerical data?" (What is the typical surface area of styrofoam pellets?)

To help prepare students for the next lesson, ask:

- "What are some different ways to represent data graphically?" (Bar graphs, dot plots, box plots, pie charts, and histograms.)

Sec A

1.3 Categorizing Questions

Cool-down

5 mins

Standards

Building Towards **HSS-ID.A.1**

Student Task Statement

Categorize each of these questions as one of these types, then explain your reasoning for putting the question in that category.

- Statistical question requiring numerical data to answer it
- Statistical question requiring categorical data to answer it
- Non-statistical question

1. On average, how many books does each person in the United States read each year?
2. How many acts are in the play *Romeo and Juliet*?
3. Which book was read most by students in the class this summer?
4. How many books are in the classroom right now?

Student Response

1. Statistical question requiring numerical data to answer it. The data will be numbers and will have some variability.
2. Non-statistical question since there is one right answer to the question.
3. Statistical question requiring categorical data to answer it. The data will be words or phrases and will have some variability.
4. Non-statistical question since there is one right answer to the question.

Responding To Student Thinking

Points to Emphasize
If students struggle with what qualifies as a statistical question, focus on the distinction again when students use the data that they collected in this lesson:

Lesson 1 Summary

Sec A

Statistics is about using data to solve problems or make decisions. There are two types of data:

- **Numerical data** are expressed using numbers that can be put in order. For example, the question “How tall are the students in this class?” would involve measuring the height of each student, resulting in numerical data.
- **Categorical data** are expressed using characteristics. For example, the question “What brand of phones do people use?” would involve surveying several people, with their answers resulting in categorical data.

The question that you ask determines the type of data that you collect and whether or not there is *variability* in the data collected. In earlier grades, you learned that there is variability in a data set if not all of the values in the data set are the same. These are examples of **statistical questions** because they are answered by collecting data that have variability:

- “What is the average class size at this school?” would produce numerical data with some variability.
- “What are the favorite colors of students in this class?” would produce categorical data with some variability.

These are examples of **non-statistical questions** because they are answered by collecting data that does not vary:

- “How many students are on the roster for this class?” has only one possible answer. There is only one value in the data set, so there is no variability.
- “What color is this marker?” has only one answer. There is only one value in the data set, so there is no variability.

Glossary

- categorical data
- non-statistical question
- numerical data
- statistical question

Practice Problems

1 Student Task Statement

Write a question of interest for which you would expect to collect numerical data.

Solution

Sample response: How many steps do I typically walk in a day?

2 Student Task Statement

Write a question of interest for which you would expect to collect categorical data.

Solution

Sample response: What are the most popular sports teams among the ninth graders at my school?

3 Student Task Statement

Select **all** the statistical questions.

A. What is the typical amount of rainfall for the month of June in the Galapagos Islands?

B. How much did it rain yesterday at the Mexico City International Airport?

C. Why do you like to listen to music?

D. How many songs does the class usually listen to each day?

E. How many songs did you listen to today?

F. What is the capital of Canada?

G. How long does it typically take for 2nd graders to walk a lap around the track?

Solution

A, D, G

Unit 1, Lesson 2

Data Representations

Goals

- Create a dot plot, histogram, and box plot to represent numerical data.
- Identify (in writing) the five-number summary that describes given statistical data.
- Interpret a box plot that represents a data set.

Learning Targets

- I can find the five-number summary for data.
- I can use a dot plot, histogram, or box plot to represent data.

Lesson Narrative

This lesson is optional because it revisits content that is below grade-level. If the pre-unit diagnostic assessment indicates that students know the representations well, this lesson may be safely skipped.

This lesson serves as a brief review of the meaning of dot plots, histograms, and box plots and how they are created.

Students calculate values for the five-number summary and use those values to create dot plots. They also compare different representations of the same data using histograms with different intervals, dot plots, and box plots.

When students identify the information displayed by different graphical representations, they are building knowledge about when to use appropriate tools (MP5), which will help them make choices about how to represent data. Students make use of structure (MP7) to connect visual representations of data sets and to reason abstractly and quantitatively (MP2) by interpreting values in the given contexts.

Standards

Building On	**6.SP.B.4**
Addressing	**HSS-ID.A.1**
Building Towards	**HSS-ID.A.1, HSS-ID.A.2**

Instructional Routines

- MLR1: Stronger and Clearer Each Time
- Notice and Wonder

Student Facing Learning Goals

Let's represent and analyze data using dot plots, histograms, and box plots.

2.1 Notice and Wonder: Battery Life

Warm-up

5 mins

Activity Narrative

The purpose of this *Warm-up* is to elicit the idea that the same data can be displayed in different ways, which will be useful when students create different data displays in a later activity. While students may notice and wonder many things about these images, the comparison of the three representations and interpreting the information in each representation are the important discussion points.

This is the first "Notice and Wonder" activity in the course. Students are shown three statistical displays representing the same data set and are asked "What do you notice? What do you wonder?"

Students are given time to write down what they notice and wonder about the displays and then time to share their thoughts. Their responses are recorded for all to see. Often, the goal is to elicit observations and curiosities about a mathematical idea that students are about to explore. Pondering the two open questions allows students to build interest about and gain entry into an upcoming task.

This prompt gives students opportunities to see and make use of structure (MP7). Specifically, they might use the structure of the three representations, particularly the structure of the horizontal number line, to find mathematically important similarities in how the same set of data is represented.

Standards

Building On	**6.SP.B.4**
Building Towards	**HSS-ID.A.1**

Instructional Routines

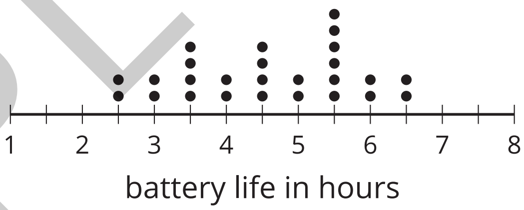
- Notice and Wonder

Launch

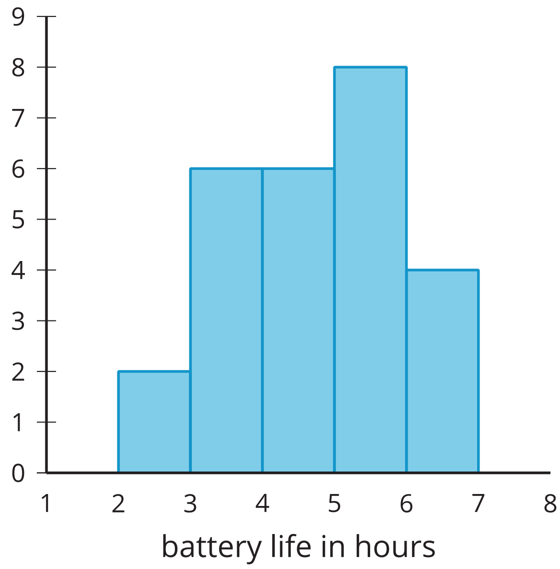
Arrange students in groups of 2. Display the images for all to see. Ask students to think of at least one thing they notice and at least one thing they wonder. Give students 1 minute of quiet think time, and then 1 minute to discuss with their partner the things they notice and wonder.

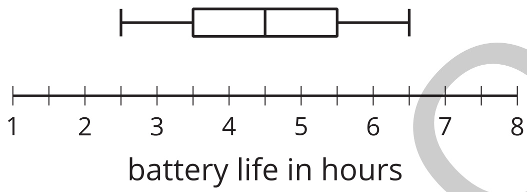
Student Task Statement

The dot plot, histogram, and box plot summarize the hours of battery life for 26 cell phones that are constantly streaming video. What do you notice? What do you wonder?



Sec A





Student Response

Things students may notice:

- The box plot has a vertical line of symmetry but the histogram does not.
- 5.5 hours is the most common battery life.
- 2.5 hours is not very long for a cell phone to last.

Things students may wonder:

- Is the 3 hours of battery life included in the second bar in the histogram or the first bar?
- Why does the histogram have intervals of 1 instead of 2?
- How do you create a box plot from data?

Activity Synthesis

Ask students to share the things they noticed and wondered. Record and display their responses without editing or commentary for all to see. If possible, record the relevant reasoning on or near the images. Next, ask students, "Is there anything on this list that you are wondering about now?" Encourage students to observe what is on display and to respectfully ask for clarification, point out contradicting information, or voice any disagreement.

The goal is to help students recall different ways to represent distributions of data. Highlight the similarities between the dot plot and the histogram. Tell students that the tallest bar in the histogram is created from the two data values at 5 and the six data values at 5.5 in the dot plot, and that the final bar is created from the two data values at 6 and the two data values at 6.5 in the dot plot. If time permits, discuss questions such as

- “Which representation(s) shows all the data values?” (The dot plot shows all the data values.)
- “How do you create a box plot?” (You calculate the values for the five-number summary and then graph them on a number line. The first quartile, the median, and the third quartile are used for the box, and the minimum value and maximum value are used for the whiskers.)

2.2 Tomato Plants: Histogram

Optional

15 mins

Activity Narrative

The mathematical purpose of this activity is to represent and analyze data with histograms. Students will create two different histograms from the same data set by organizing data into different intervals.

This is the first time Math Language Routine 1: Stronger and Clearer Each Time is suggested in this course. In this routine, students are given a thought-provoking question or prompt and asked to create a first draft response in writing. It is not necessary that students finish this draft before moving to the structured partner meetings step. Students then meet with 2–3 partners to share and refine their response through conversation. While meeting, listeners ask questions such as “What did you mean by . . . ?” and “Can you say that another way?” Finally, students write a second draft of their response reflecting ideas from partners, and improvements on their initial ideas. Students should be encouraged to incorporate any good ideas and words they get from their partners to make their second draft stronger and clearer.

Access for English Language Learners

This activity uses the *Stronger and Clearer Each Time* math language routine to advance writing, speaking, and listening as students refine mathematical language and ideas.

Standards

Addressing HSS-ID.A.1

Building Towards HSS-ID.A.2

Instructional Routines

- MLR1: Stronger and Clearer Each Time

Launch

Arrange students in groups of 2.

Access for English Language Learners

Action and Expression: Internalize Executive Functions. To support organization, provide students with grid or graph paper to organize their two histograms with different interval widths.
Supports accessibility for: Language, Organization

Student Task Statement

A histogram can be used to represent the distribution of numerical data.

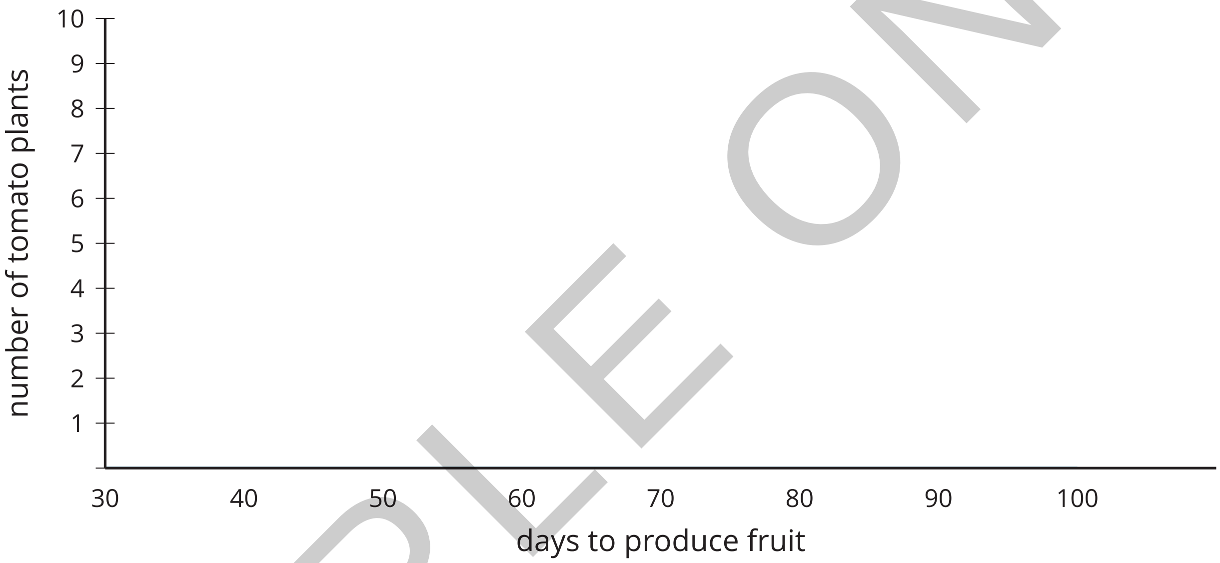
1. The data represent the number of days it takes for different tomato plants to produce tomatoes. Use the information to complete the frequency table.

Sec A

47	52	53	55	57	60	61	62	63	65	65	65	65
68	70	72	72	75	75	75	76	77	78	80	81	82
85	88	89	90									

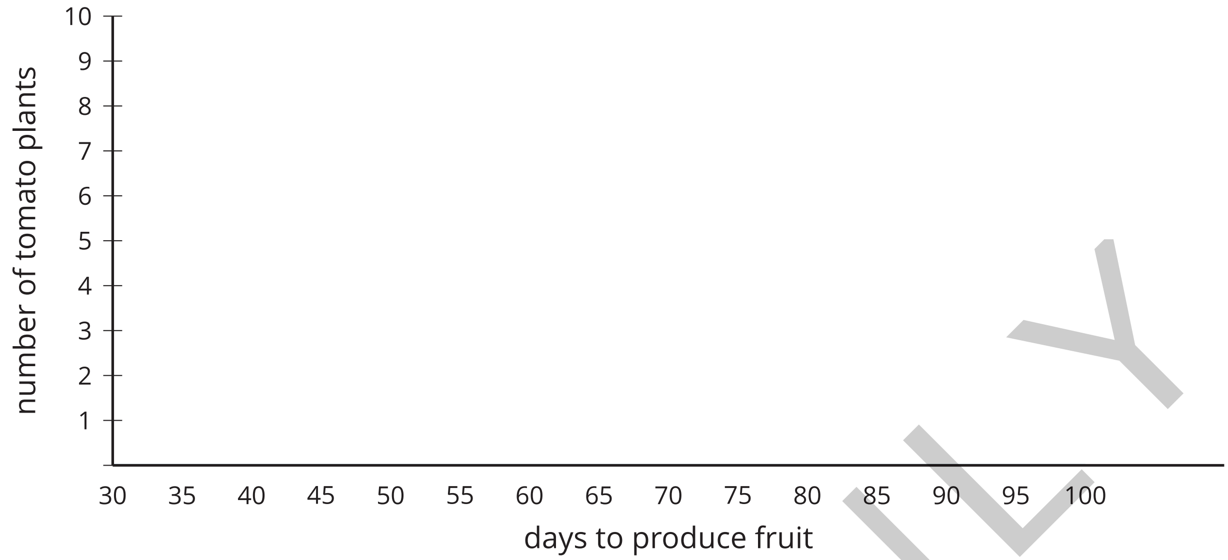
days to produce fruit	frequency
40–50	
50–60	
60–70	
70–80	
80–90	
90–100	

2. Use the set of axes and the information in your table to create a histogram.



3. The histogram you created has intervals of width 10 (like 40–50 and 50–60). Use the set of axes and data to create another histogram with an interval of width 5. How does this histogram differ from the other one?

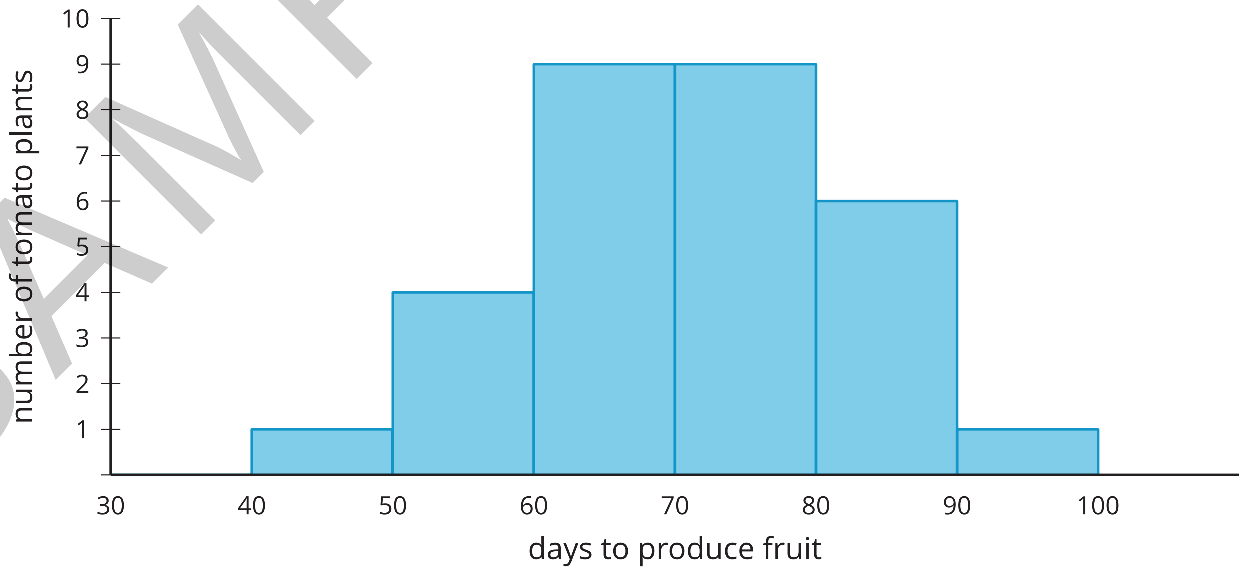
KH | iM Illustrative® Mathematics LEARN MATH FOR LIFE



Student Response

1.

days to produce fruit	frequency
40–50	1
50–60	4
60–70	9
70–80	9
80–90	6
90–100	1



2.

3. Sample response: This histogram has more bars with shorter heights. It shows a little more detail of the distribution shape than the histogram with larger intervals.

Building on Student Thinking

Students may struggle to know how to place numbers that lie on the boundary between intervals. For example, students may not know if a value like 60 should be included in the interval 50–60 or 60–70. Explain to students that the lower boundary value is included in the interval, and the upper boundary value is not. For example, the interval 60–70 includes all the values that are greater than or equal to 60 and less than 70.

Are You Ready for More?

It often takes some playing around with the interval widths to figure out which gives the best sense of the shape of the distribution.

1. What might be a problem with using interval widths that are too large?
2. What might be a problem with using interval widths that are too small?
3. What other considerations might go into choosing the width of an interval?

Extension Student Response

1. Sample response: If the interval widths are too large we will end up with very few intervals and important features within an interval could be lost. At the most extreme, we would end up with a single interval including everything, which would not be very helpful.
2. Sample response: If the interval widths are too small we end up with a histogram that is essentially the same as a dot plot. In some cases this could be useful, but in situations where each data value occurs only once or twice, every bar will have the same height, not allowing us to get a good sense of the shape of the distribution.
3. Sample response: There may be natural interval widths based on the context—such as 10 years (decades), common age ranges, or 100-yard widths. There are also common interval widths that people expect, such as 1, 2, 5, 10, 50, 100, and so on. It usually does not make sense to have fewer than 5 bars or more than 50 bars in a histogram.

Activity Synthesis

The purpose of this discussion is to make sure that students know how to create and begin to interpret histograms.

Use *Stronger and Clearer Each Time* to give students an opportunity to revise and refine their responses when explaining the differences between the two histograms. In this structured-pairing strategy, students bring their first draft response into conversations with 2–3 different partners. They take turns being the speaker and the listener. As the speaker, students share their initial ideas and read their first draft. As the listener, students ask questions and give feedback that will help their partners clarify and strengthen their ideas and writing.

Consider displaying these prompts for feedback:

- "What do you mean when you say . . . ?"
- "Can you describe that another way?"
- "How do you know . . . ."

Close the partner conversations, and give students 3–5 minutes to revise their first draft. Encourage students to incorporate any good ideas and words that they get from their partners to make their next draft stronger and clearer.

Here is an example of a second draft: "The second histogram shows a little more detail than the first histogram."

As time allows, invite students to compare their first and final drafts. Select 2–3 students to share how their drafts changed and why they made the changes they did.

After *Stronger and Clearer Each Time*, consider asking these questions for discussion about creating and interpreting these histograms.

- "Where did you put the 60? In the 50–60 interval or 60–70 interval?" (Tell students that we use the convention of including the 60 in the 60–70 interval. The interval 60–70 means all the values greater than or equal to 60, but less than 70. The interval 50–60 means all values greater than or equal to 50 but less than 60.)
- "What information is easily seen in a histogram?" (The shape of the distribution as well as estimates for the measure of center and measure of variability.)
- "According to each histogram, what appears to be the typical number of days it takes a tomato plant to produce tomatoes?" (Using the first histogram, it appears that the typical number of days is somewhere between 60 and 80. Looking at the second histogram, it appears that the typical number of days could be between 75 and 80.)
- "What information is not seen in a histogram?" (You are not able to see the actual values. You know only the number of values within an interval rather than the values themselves.)

2.3 Tomato Plants: Box Plot

Optional

10 mins

Activity Narrative

The mathematical purpose of this activity is to represent the distribution of data on the real number line with a box plot and to help students think informally about the median as a measure of center. Students calculate the values for the five-number summary and create a box plot. The median, quartiles, and extreme values split the data set into four intervals, with approximately the same number of data values in each. Students interpret these values in the given context (MP2). Although these intervals are often called "quartiles," the term "quarters" is used in these materials to avoid confusion with the quartile values Q1 and Q3.

Standards

Addressing **HSS-ID.A.1**

Launch

Keep students in groups of 2. Give students 5 minutes to answer the questions. Ask them to compare their answer with their partner after each question.

Access for Students with Disabilities

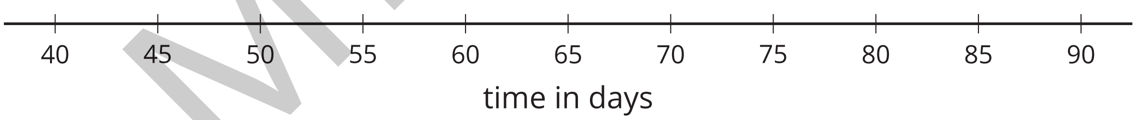
Action and Expression: Internalize Executive Functions. To support development of organizational skills in problem solving, chunk this task into more manageable parts. For example, instruct students to refer to their sequential data, divide the data into quarters, and then find the median, Q1, and Q3.
Supports accessibility for: Memory, Organization

Student Task Statement

A box plot can also be used to represent the distribution of numerical data.

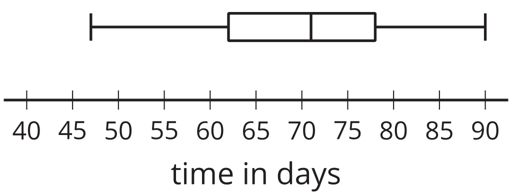
minimum	Q1	median	Q3	maximum

1. Using the same data as in the previous activity for tomato plants, find the median, and add it to the table. What does the median represent for these data?
2. Find the median of the least 15 values to split the data into the first and second quarters. This value is called the first quartile. Add this value to the table under Q1. What does this value mean in this situation?
3. Find the value (the third quartile) that splits the data into the third and fourth quarters, and add it to the table under Q3. Add the minimum and maximum values to the table.
4. Use the **five-number summary** to create a box plot that represents the number of days it takes for these tomato plants to produce tomatoes.



Student Response

1. Median: 71 days. Half (or 15) of the tomato plants produced tomatoes in less than or equal to 71 days, and half of them took at least 71 days to produce fruit.
2. Q1: 62 days. About one-fourth of the tomato plants took 62 days or less to produce tomatoes, and about 75% of the tomato plants took 62 days or more to produce tomatoes.
3. Q3: 78 days. Minimum: 47 days. Maximum: 90 days.



4.

Building on Student Thinking

For students who have difficulty calculating the median, remind them that the median is the middle of a sequential data set. For students who have difficulty finding Q1 and Q3, ask them how many groups we should have if we are splitting the data into "quarters." The data should be divided into four equal groups. The median of the lower half of the values is Q1, and the median of the upper half of the values is Q3.

Activity Synthesis

The goal is to make sure that students understand the five-number summary and to help them think informally about the median as a measure of center. Here are some questions for discussion.

- "What information is easily seen in the box plot?" (The minimum value, quartiles including the median, and the maximum value. This also highlights the interquartile range and the range.)
- "According to the box plot, what is the typical number of days it takes a tomato plant to produce fruit?" (The typical number of days is 71 because the median of the data is 71 days.)

Lesson Synthesis

In this lesson students viewed data represented by dot plots, histograms, and box plots.

- "What are the strengths of each of the representations?" (A dot plot lets you see all of the data and how they are distributed. A histogram summarizes the data using intervals, resulting in fewer columns than in a dot plot. A box plot displays the five-number summary graphically.)
- "What are the weaknesses of each of the representations?" (A dot plot has many columns of dots that can make it difficult to determine patterns graphically. Both the histogram and the box plot do not display each individual value in the data set, which means that the mean cannot be calculated directly from either representation.)
- "How do you find the 'typical' value for a data set?" (You can calculate the mean or median, or estimate the mean or median using a graphical representation.)

2.4 Reasoning About Representations

Cool-down

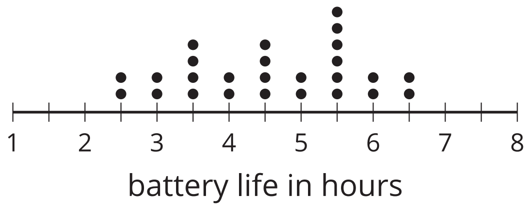
5 mins

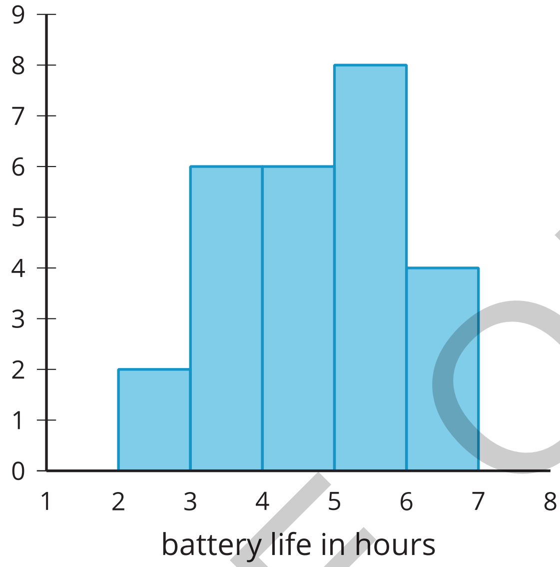
Standards

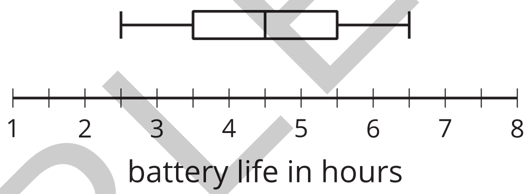
Addressing	HSS-ID.A.1
Building Towards	HSS-ID.A.2

Student Task Statement

The dot plot, histogram, and box plot represent the distribution of the same data in 3 different ways.







1. What information can be seen most easily in the dot plot?
2. What information can be seen most easily in the histogram?
3. What information can be seen most easily in the box plot?

Student Response

Sample response:

1. The actual values, the shape of the distribution, and the most common value are easily seen in the dot plot.
2. The shape of the distribution and the most common interval of data are easily seen in the histogram.
3. The five-number summary (minimum, first quartile, median, third quartile, and maximum) are easily seen in the box plot.

Responding To Student Thinking

More Chances

Students will have more opportunities to understand the mathematical ideas addressed here. There is no need to slow down or add additional work to the next lessons.

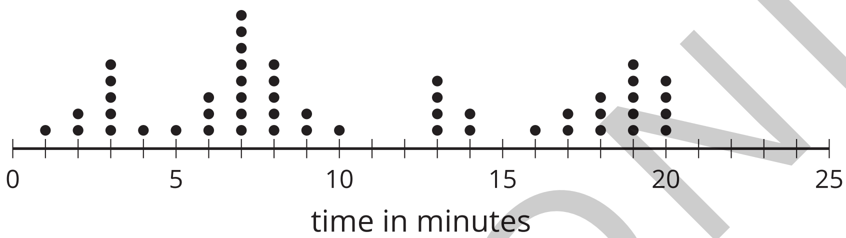
Lesson 2 Summary

The table shows a list of the number of minutes people could intensely focus on a task before needing a break. Fifty people of different ages are represented.

19 7 1 16 20 2 7 19 9 13 3 9 18 13 20 8 3 14 13 2 8 5 17 7

18 17 8 8 7 6 2 20 7 7 10 7 6 19 3 18 8 19 7 13 20 14 6 3

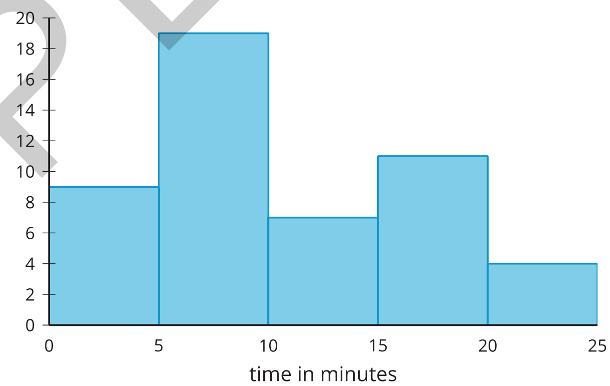
19 4



In a situation like this, it is helpful to represent the data graphically to better notice any patterns or other interesting features in the data. A dot plot can be used to see the shape and **distribution** of the data.

There were quite a few people that lost focus at around 3, 7, 13, and 19 minutes, and nobody lost focus at 11, 12, or 15 minutes. Dot plots are useful when the data set is not too large and shows all of the individual values in the data set. In this example, a dot plot can easily show all of the data. If the data set is very large (more than 100 values, for example), or if there are many different values that are not exactly the same, it may be hard to see all of the dots on a dot plot.

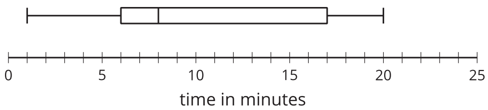
A histogram is another representation that shows the shape and distribution of the same data.



Most people lost focus between 5 and 10 minutes or between 15 and 20 minutes, while only 4 of the 50 people got distracted between 20 and 25 minutes. When creating histograms, each interval includes the number at the lower end of the interval but not the number at the upper end.

For example, the tallest bar displays values that are greater than or equal to 5 minutes but less than 10 minutes. In a histogram, values that are in an interval are grouped together. Although the individual values get lost with the grouping, a histogram can still show the shape of the distribution.

Here is a box plot that represents the same data.



Box plots are created using a **five-number summary**. For a set of data, the five-number summary consists of these five statistics: the minimum value, the first quartile, the median, the third quartile, and the maximum value. These values split the data into four sections, each representing approximately one-fourth of the data. The median of this data is indicated at 8 minutes, and about 25% of the data fall in the short second quarter of the data between 6 and 8 minutes. Similarly, approximately one-fourth of the data are between 8 and 17 minutes. Like the histogram, the box plot does not show individual data values, but other features such as quartiles, range, and median are seen more easily. Dot plots, histograms, and box plots provide three different ways to look at the shape and distribution while highlighting different aspects of the data.

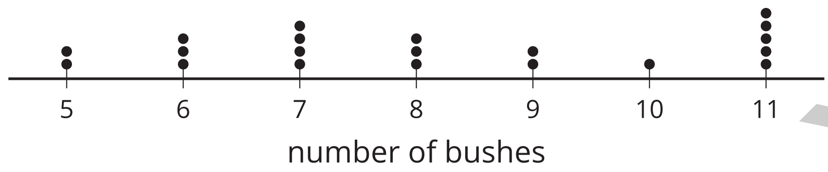
Glossary

- distribution
- five-number summary

Practice Problems

1 Student Task Statement

The dot plot displays the number of bushes in the yards for houses in a neighborhood. What is the median?



Solution

8 bushes

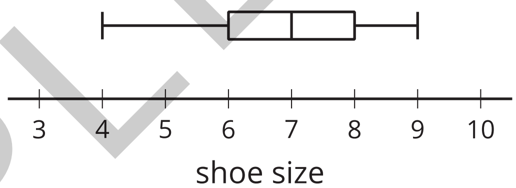
2 Student Task Statement

The data set represents the shoe sizes of 19 students in a fifth grade physical education class.

4 5 5 5 6 6 6 6 7 7 7 7 7.5 7.5 8 8 8.5 8.5 9

Create a box plot to represent the distribution of the data.

Solution



3 Student Task Statement

The data set represents the number of pages in the last book read by each of 20 students over the summer.

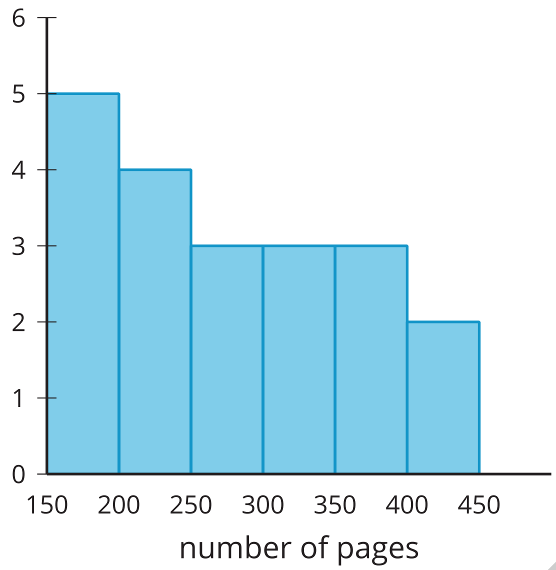
163 170 171 173 175 205 220 220 220 253 267 281 305 305 305 355

371 388 402 431

Create a histogram to represent the distribution of the data.

Solution

Sample response:



4 from Unit 1, Lesson 1

Student Task Statement

Each set of data was collected from surveys to answer statistical questions. Select **all** of the data sets that represent numerical data.

A. {1, 1.2, 1.4, 1.4, 1.5, 1.6, 1.8, 1.9, 2, 2, 2.1, 2.5}

B. {Red, Red, Yellow, Yellow, Blue, Blue, Blue}

C. {45, 60, 60, 70, 75, 80, 85, 90, 90, 100, 100, 100}

D. {-7, -5, -3, -1, -1, -1, 0}

E. {98.2, 98.4, 98.4, 98.6, 98.6, 98.6, 98.6, 98.7, 98.8, 98.8}

F. {Yes, Yes, Yes, Yes, Maybe, Maybe, No, No, No}

G. {A, A, A, B, B, B, C, C, C}

Solution

A, C, D, E

5 from Unit 1, Lesson 1

Student Task Statement

Is "What is the typical distance a moped can be driven on a single tank of gas?" a statistical question? Explain your reasoning.

Solution

Sample response: Yes, it is a statistical question, because the data required to answer it would vary. The distance the moped traveled would likely depend on speed or the number of hills.

Unit 1, Lesson 3

A Gallery of Data

Goals

- Create and critique graphical representations of student-collected data.

Learning Targets

- I can graphically represent the data that I collected and critique the representations of others' data.

Lesson Narrative

Students represent and interpret data that they collected from their classmates in a previous lesson. They reason abstractly and quantitatively (MP2) by creating a display from data and interpret the meaning of the displays that others have created. Students also make use of structure (MP7) to recognize differences in distributions with the same shape, but different centers.

Math Community

In this lesson, students review the themes that arose when they shared their initial thoughts in Exercise 1 about what they think it should look like and sound like to do math together as a community. Students then have a chance to both affirm and add to the ideas that were generated.

Standards

Addressing	**HSS-ID.A.1**
Building Towards	**HSS-ID.A.2**

Instructional Routines

- MLR7: Compare and Connect
- Notice and Wonder

Required Materials

Materials To Gather

- Math Community Chart: Activity 1
- Sticky notes: Activity 1
- Tools for creating a visual display: Activity 2

Required Preparation

Activity 2:

Students will need the numerical data that they collected from a statistical question in a previous lesson. Students will need tools to display that data—on a dot plot, histogram, and a box plot—to the whole class.

Student Facing Learning Goals

Let's make, compare, and interpret data displays.

3.1

Notice and Wonder: Dot Plots

Warm-up

5 mins

Sec A

Activity Narrative

The purpose of this *Warm-up* is to elicit the idea that distributions can be discussed in terms of shape, which will be useful when students describe data displays in a later activity. While students may notice and wonder many things about these images, the shape of the images and the values on the horizontal axis of each are the important discussion points. This prompt gives students opportunities to see and make use of structure (MP7). The specific structure they might notice is that data sets with different values can have distributions with the same shape if all of the values in the data set are increased or decreased by the same value.

Standards

Addressing	**HSS-ID.A.1**
Building Towards	**HSS-ID.A.2**

Instructional Routines

- Notice and Wonder

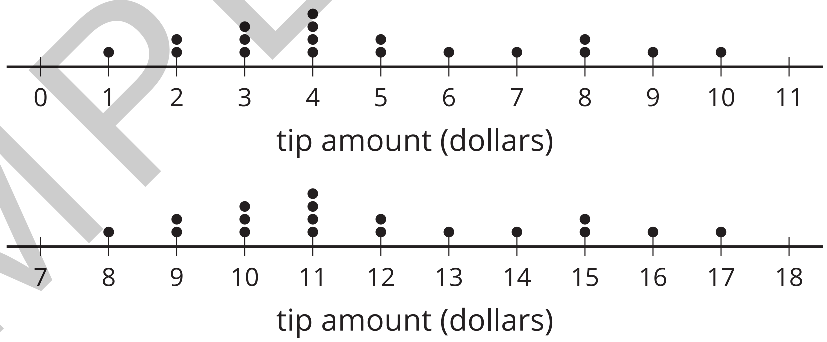
Launch

Display the dot plots for all to see. Ask students to think of at least one thing they notice and at least one thing they wonder. Give students 1 minute of quiet think time, and then 1 minute to discuss the things they notice with their partner, followed by a whole-class discussion.

Student Task Statement

The dot plots represent the distribution of the amount of tips, in dollars, left at 2 different restaurants on the same night.

What do you notice? What do you wonder?



Student Response

Things students may notice:

- The shape of the distribution is the same.
- The second dot plot is the same data shifted up $7.

Things students may wonder:

- What would a box plot of the data look like?
- What would a dinner cost that had a $17 tip?

Activity Synthesis

Ask students to share the things they noticed and wondered. Record and display their responses for all to see. If possible, record the relevant reasoning on or near the image. After all responses have been recorded without commentary or editing, ask students, "Is there anything on this list that you are wondering about now?" Encourage students to respectfully disagree, ask for clarification, or point out contradicting information. If the shape of the distribution and the values on the horizontal axis of each dot plot do not come up during the conversation, use these questions to prompt students to discuss those ideas.

- "What do you notice about the shape of each distribution?" (The data are distributed in exactly the same way in each dot plot.)
- "What is the most frequent value in each dot plot?" ($4 and $11)
- "What is the value of the highest tip in each dot plot?" ($10 and $17)
- "What is the value of the lowest tip in each dot plot?" ($1 and $8)
- "What happens if $7 is added to each of the tips in the first dot plot?" (You get the data distribution in the second dot plot.)

Math Community

After the *Warm-up*, display the class Math Community Chart for all to see and explain that the listed "Doing Math" actions come from the sticky notes students wrote in the first exercise. Give students 1 minute to review the chart. Then invite students to identify something on the chart they agree with and hope for the class or something they feel is missing from the chart and would like to add. Record any additions on the chart. Tell students that the chart will continue to grow and that they can suggest other additions that they think of throughout today's lesson during the *Cool-down*.

3.2 Data Displays

25 mins

Activity Narrative

In this lesson, students create and display a dot plot and a box plot using numerical data that they collected from a survey question in a previous lesson. The focus of the lesson is on creating graphical displays of data, so appropriate data sets take precedence over the class's actual data. Here are sample data sets for each of the four questions from the previous lesson.

- On average, how many letters are in the preferred names for students in this class? {4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 9, 10}
- On average, how many speed limit signs do students in our class see on the way to school? {2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 7, 7}
- About how long did it take students in this class to get to school this morning? {5, 5, 5, 5, 10, 10, 12, 15, 15, 15, 15, 25, 25, 25, 25, 30, 35, 40, 45, 55}
- On average, how many movies in the theater did each student in the class watch this summer? {0, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 4, 6, 6, 7, 7, 8, 10, 12}

- After groups have had a chance to complete their displays, pause the class to discuss how to do a gallery walk. To help students understand the types of responses expected, provide an example display. Ask the class to refer to the example and to answer the questions in the task.

This is the first time Math Language Routine 7: Compare and Connect is suggested in this course. In this routine, students are given a problem that can be approached using multiple strategies or representations, and they record their methods for all to see. They then compare and identify correspondences across strategies by means of a teacher-led gallery walk with commentary or teacher think-aloud (such as "I notice . . . .” or I wonder . . . ."). As they compare their own strategy to that of others, a typical discussion prompt is “What is the same and what is different?”. The purpose of this routine is to allow students to make sense of mathematical strategies and, through constructive conversations, develop awareness of the language used as they compare, contrast, and connect other ways of thinking to their own.

Access for English Language Learners

This activity uses the *Compare and Connect* math language routine to advance representing and conversing as students use mathematically precise language in discussion.

Standards

Addressing **HSS-ID.A.1**

Instructional Routines

- MLR7: Compare and Connect

Launch

Arrange students in groups of 2–4, and assign each group one of the statistical questions for which students have already collected numerical data. Provide each group with tools for creating a graphical display. Ask students to pause after the second question for the gallery walk.

Access for Students with Disabilities

Action and Expression: Develop Expression and Communication. Invite students to talk about their ideas with a partner before writing them down. Display sentence frames to support students when they explain their ideas. For example, “Each student in this class . . . .”
Supports accessibility for: Language, Organization

Student Task Statement

Your teacher will assign your group a statistical question. As a group:

1. Create a dot plot and a box plot to display the distribution of the data.
2. Write 3 comments that interpret the data. Pause here for instructions on visiting the other displays.
3. Visit each display and leave a note or question about the information in the display.

Student Response

Sample response: Each student in this class watched about 4 movies this summer. At least one person watched 12 movies and one student saw 0. Most of the data are clustered around 1 to 3 movies with a longer tail to the right.

Building on Student Thinking

Students may neglect to include titles for axes and may forget the importance of building their plots on a number line with equally spaced intervals. Monitor for groups that do not recall the details of making the different types of displays, and suggest that they refer back to their work in the previous lesson.

Are You Ready for More?

Choose one of the more interesting questions that you or a classmate asked, and collect data from a larger group—such as more students from the school. Create a data display, and compare results from the data collected in class.

Extension Student Response

Answers vary.

Activity Synthesis

The purpose of the discussion is to elicit evidence of the students' thinking about the data in the displays and interpretation of the distributions in context.

Invite groups to share their displays. After all strategies have been presented, use *Compare and Connect* to help students compare, contrast, and connect the different representations. Ask,

- "What is the same and what is different about each display?"
- "How did the comments help you understand the displays?"
- "Are there any benefits or drawbacks to one representation compared to another?"
- "What are some ways you summarized the information in the display?" (The data are clustered around one value. The dot plot shows this but the box plot does not. The box plot shows the median as a typical value.)
- "If you had collected data from all of the students in the school, instead of just from your classmates, would you rather create a dot plot or a box plot? Why?" (A box plot because all I'd have to do is find the five-number summary. In a dot plot, I'd have to plot every point, and that might be hard to do with the tools that I have).
- "What is the shape of the distribution in your dot plot?" (The data are very spread apart, but mostly evenly distributed.)
- "What information is displayed by the dot plot that is not displayed by the box plot?" (The dot plot displays all the values in the data set.)
- "What information is displayed by the box plot that is not displayed by the dot plot?" (The box plot displays the quartiles and the median.)

Lesson Synthesis

In this lesson, students created two different data displays from information they collected in a previous lesson.

- "When you look at the two data displays you made, what information jumps out at you?" (The dot plot shows the shape of the data, and it is easy to see the frequency of each value. The box plot shows the median and gives me an idea of the interval that contains the middle fifty percent of the data.)
- "What are some contexts in which you have seen dot plots, box plots, or histograms outside of this class?" (I have

used dot plots in science class when we collected data from an experiment. I have used histograms when we were reading technical writing in English class.)

- "What do you understand about data displays?" (I can use data displays to show the distribution of data. Different displays allow me to notice information about the distribution of the data in different ways.)

3.3 Why Graphical Representations?

Cool-down

5 mins

Sec A

Standards

Addressing **HSS-ID.A.1**

Launch

Math Community

Before distributing the *Cool-downs*, display the Math Community Chart and the community building question "Is there anything that you would like to add to the student 'Doing Math' section of the chart?" Ask students to respond to the question after completing the *Cool-down* on the same sheet.

After collecting the *Cool-downs*, identify themes from the community building question. Use the themes to add to or revise the student section of the Math Community Chart before Exercise 3.

Student Task Statement

A large school system summarizes the number of teachers at 51 schools in the area.

20–29	5
30–39	7
40–49	5
50–59	6
60–69	6
70–79	4
80–89	2
90–99	7
100–109	8
110–119	1

minimum: 20 teachers

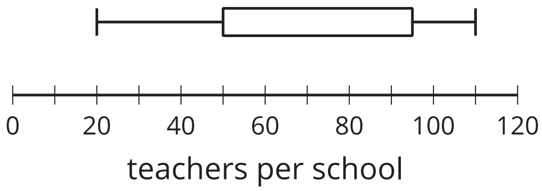
maximum: 110 teachers

median: 65 teachers

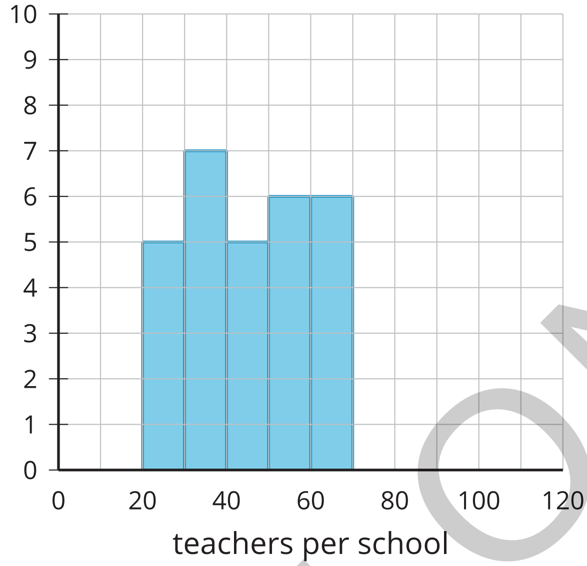
Q1: 40

Q3: 95

1. Fix the box plot to show this information.



2. Finish the histogram to show this information.

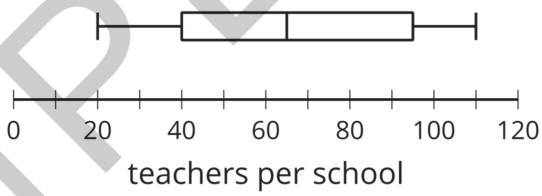


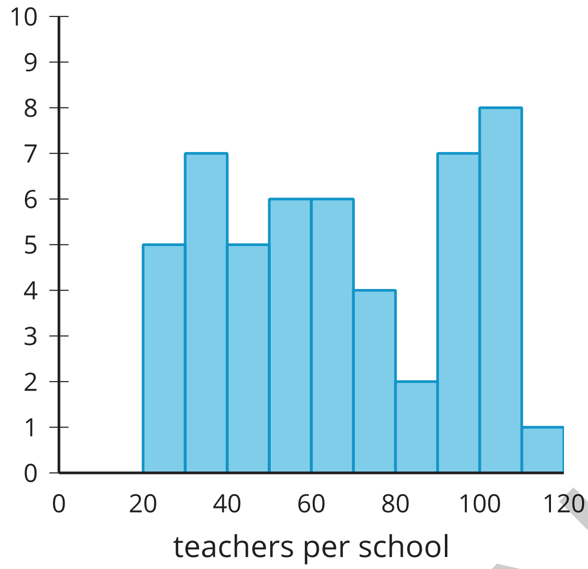
3. Which of these data displays most easily shows how many schools have at least 100 teachers per school? Explain your reasoning.

Student Response

Sample responses:

1.





2.

3. The histogram since we can add the heights of the last two bars to get the answer. It is not possible to get this information from the box plot.

Responding To Student Thinking

Points to Emphasize
If most students struggle to create histograms and box plots, select student work to revisit after the *Warm-up* referred to here. For example, focus on work with common errors, and invite students to correct those errors.

Algebra 1, Unit 1, Lesson 4, Activity 1 Which Three Go Together: Distribution Shape

Lesson 3 Summary

We can represent a distribution of data in several different forms, including lists, dot plots, histograms, and box plots. A list displays all of the values in a data set and can be organized in different ways. This list shows the pH for 30 different water samples.

5.9 7.6 7.5 8.2 7.6 8.6 8.1 7.9 6.1 6.3 6.9 7.1 8.4 6.5 7.2 6.8 7.3 8.1 5.8

7.5 7.1 8.4 8.0 7.2 7.4 6.5 6.8 7.0 7.4 7.6

Here is the same list organized in order from least to greatest.

5.8 5.9 6.1 6.3 6.5 6.5 6.8 6.8 6.9 7.0 7.1 7.1 7.2 7.2 7.3 7.4 7.4 7.5 7.5

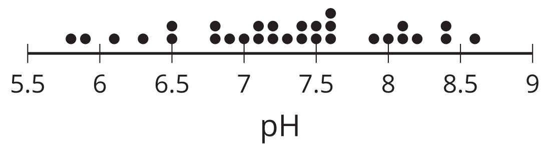
7.6 7.6 7.6 7.9 8.0 8.1 8.1 8.2 8.4 8.4 8.6

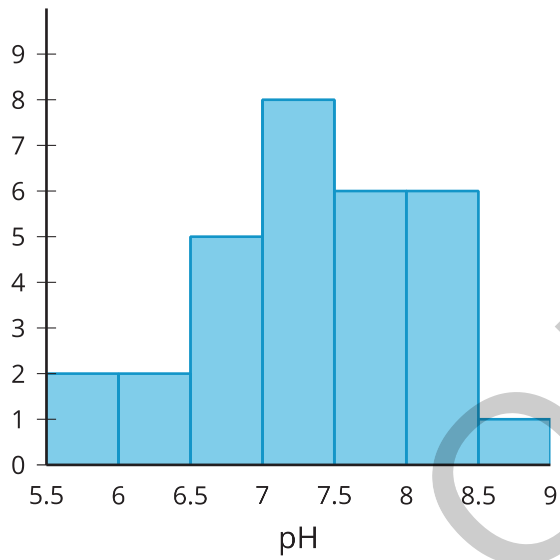
With the list organized, you can more easily:

- Interpret the data
- Calculate the values of the five-number summary
- Estimate or calculate the mean

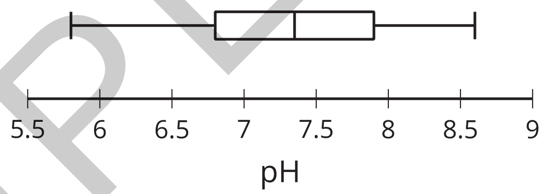
- Create a dot plot, box plot, or histogram

Here are a dot plot and histogram representing the distribution of the data in the list.





A dot plot is created by putting a dot for each value above the position on a number line. For the pH dot plot, there are 2 water samples with a pH of 6.5 and 1 water sample with a pH of 7. A histogram is made by counting the number of values from the data set in a certain interval and drawing a bar over that interval at a height that matches the count. In the pH histogram, there are 5 water samples that have a pH between 6.5 and 7 (including 6.5, but not 7). Here is a box plot representing the distribution of the same data as represented by the dot plot and histogram.

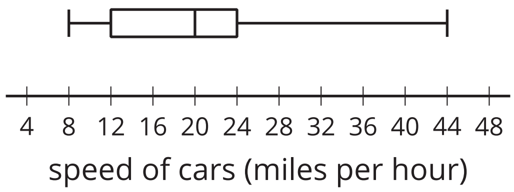


To create a box plot, you need to find the minimum, first quartile, median, third quartile, and maximum values for the data set. These 5 values are sometimes called the *five-number summary*. Drawing a vertical mark and then connecting the pieces as in the example creates the box plot. For the pH box plot, we can see that the minimum is about 5.8, the median is about 7.4, and the third quartile is around 7.9.

Practice Problems

1 Student Task Statement

The box plot represents the distribution of speeds, in miles per hour, of 100 cars as they passed through a busy intersection.



a. What is the smallest value in the data set? Interpret this value in the situation.
b. What is the largest value in the data set? Interpret this value in the situation.
c. What is the median? Interpret this value in the situation.
d. What is the first quartile (Q1)? Interpret this value in the situation.
e. What is the third quartile (Q3)? Interpret this value in the situation.

Solution

a. 8 miles per hour. The car that went through the intersection the slowest was going 8 miles per hour.
b. 44 miles per hour. The car that went through the intersection the fastest was going 44 miles per hour.
c. 20 miles per hour. Half of the cars went through the intersection at 20 miles per hour or faster, and half of them went through the intersection at 20 mph or slower.
d. 12 miles per hour. About one fourth of the cars going through the intersection were going 12 miles per hour or slower.
e. 24 miles per hour. About one fourth of the cars going through the intersection were going at least 24 miles per hour.

2 Student Task Statement

The data set represents the number of eggs produced by a small group of chickens each day for ten days.

7 7 7 7 7 8 8 8 8 9

Select **all** the values that could represent the typical number of eggs produced in a day.

A. 7.5 eggs
B. 7.6 eggs
C. 7.7 eggs
D. 8 eggs
E. 9 eggs

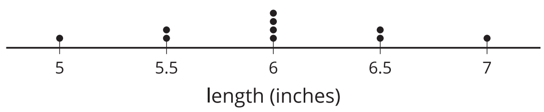
Solution

A, B

3 from Unit 1, Lesson 2

Student Task Statement

The dot plot displays the lengths of pencils (in inches) used by students in a class. What is the mean?



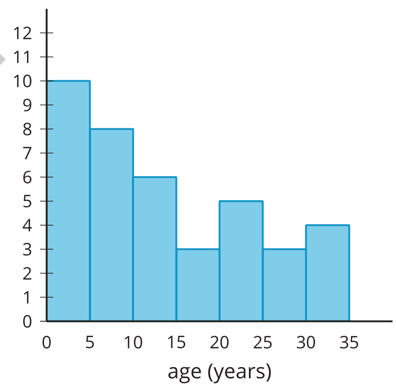
Solution

6 inches

4 from Unit 1, Lesson 2

Student Task Statement

The histogram represents ages of 39 people at a store that sells children's clothes. Which interval contains the median?



A. The interval from 0 to 5 years.

B. The interval from 5 to 10 years.

C. The interval from 10 to 15 years.

D. The interval from 15 to 20 years.

Solution

C

5 from Unit 1, Lesson 2

Student Task Statement

The data set represents the responses, in degrees Fahrenheit, collected to answer the question "How hot is the sidewalk during the school day?".

92 95 95 95 98 100 100 100 103 105 105 111 112 115 115 116 117

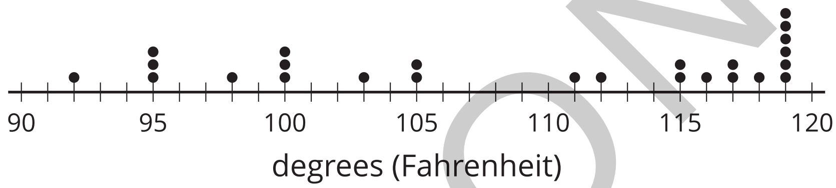
117 118 119 119 119 119 119 119

a. Create a dot plot to represent the distribution of the data.

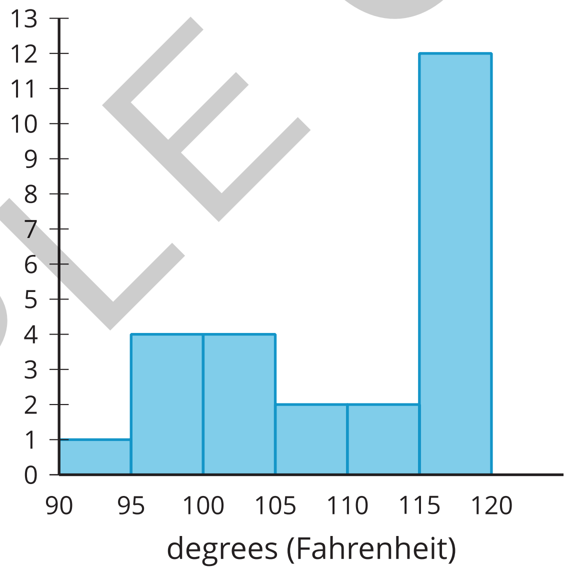
b. Create a histogram to represent the distribution of the data.

c. Which display gives you a better overall understanding of the data? Explain your reasoning.

Solution



a.



b.

Sample response:

c. Sample response: The histogram gives a better feel for the data because it displays that the temperature of the sidewalk was between 115 and 120 degrees Fahrenheit 12 times and that the temperature was between 90 and 115 degrees Fahrenheit the rest of the time.

6 from Unit 1, Lesson 1

Student Task Statement

Is "What is the area of the floor in this classroom?" a statistical question? Explain your reasoning.

Solution

Sample response: No, it is not a statistical question, because the data required would not vary. There is only one answer.

Section B: Distribution Shapes

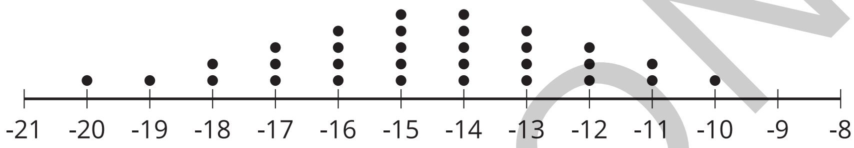
Goals

- Describe the shape of a distribution, including a measure of center and a measure of variability.

Section Narrative

Students work to describe distributions of data using terms such as “symmetric,” “skewed,” “uniform,” “bimodal,” and “bell-shaped.”

An optional lesson is included to remind students of prior learning about measures of center and measures of variation so that those terms can be included in distribution descriptions. Even if the optional lesson is skipped, consider creating a display about the measures of center and variability, like the one mentioned in a *Synthesis* of this lesson.



Teacher Reflection Questions

- **Math Content and Student Thinking:** The work on mean absolute deviation provided an opportunity to reason about variability. What are some ways students engaged with the concept of variability of different data sets?
- **Pedagogy:** Reflect on how comfortable your students are asking questions to you and to each other. What can you do to encourage students to ask questions?
- **Access and Equity:** What makes someone want to persist through a difficult math problem? In what ways are you making assumptions about which of your students are resilient, motivated, or persistent with their mathematics?

Section B Checkpoint

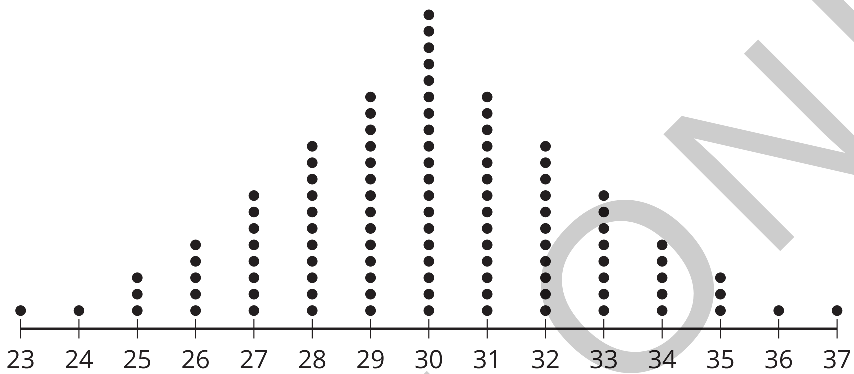
1

Goals Assessed

- Describe the shape of a distribution, including a measure of center and a measure of variability.

Student Task Statement

Describe the shape of the distribution. Include a measure of center and a measure of variability that would make sense to use, and estimate values for each.



Solution

Sample response: The distribution is bell-shaped and symmetric with a mean of 30 and a MAD of about 2.

Responding To Student Thinking

Points to Emphasize
If students struggle to describe the distribution, give students a chance to practice describing distributions as they arise over the next few lessons. For example, this practice problem could be used:

Algebra 1, Unit 1, Lesson 9, Practice Problem 3

The Shape of Distributions

Goals

- Describe (orally and in writing) the shape of a distribution, using words such as "symmetric," "skewed," "uniform," "bimodal," and "bell-shaped."
- Interpret a graphical representation to suggest a possible context for the data.

Learning Targets

- I can use a graphical representation of data to suggest a situation that produced the data pictured.
- I can use the terms "symmetric," "skewed," "uniform," "bimodal," and "bell-shaped" to describe the shape of a distribution.

Lesson Narrative

In this lesson, students describe distributions using appropriate terminology.

- In a **symmetric distribution**, the mean is equal to the median and there is a vertical line of symmetry in the center of the data display.
- In a **skewed distribution**, the mean is not usually equal to the median and one side of the distribution has more values farther from the bulk of the data than the other side has.
- A **uniform distribution** has the data values evenly distributed throughout the range of the data.
- A **bimodal distribution** has two very common data values seen in a dot plot or histogram as distinct peaks.
- A **bell-shaped distribution** has a dot plot that takes the form of a bell with most of the data clustered near the center and fewer points farther from the center.

Students practice applying this terminology as they invent reasonable contexts for a described distribution.

The *Warm-up* activity gives students a reason to begin using language precisely (MP6) and gives the teacher the opportunity to hear how students use terminology and talk about characteristics of the items in comparison to one another. In the *Card Sort*, as students trade roles, they explain their thinking and they listen to the reasoning of a partner. This provides them with opportunities both to explain their reasoning and to critique the reasoning of others (MP3).

Standards

Addressing	HSS-ID.A.1
Building Towards	HSS-ID.A.2

Instructional Routines

- Card Sort
- MLR2: Collect and Display
- MLR7: Compare and Connect
- Take Turns
- Which Three Go Together?

Required Materials

Materials To Copy

- Matching Distributions Cards (1 copy for every 2 students): Activity 2

Student Facing Learning Goals

Let's explore data and describe distributions.

4.1 Which Three Go Together: Distribution Shape

Warm-up

10 mins

Activity Narrative

This is the first *Which Three Go Together* routine in the course. In this routine, students are presented with four items or representations and asked: "Which three go together?" and "Why do they go together?"

Students are given time to identify a set of three items, explain their rationale, and refine their explanation to be more precise or find additional sets. The reasoning here prompts students to notice common mathematical attributes, look for structure (MP7), and attend to precision (MP6), which deepen their awareness of connections across representations.

This *Warm-up* prompts students to compare four distributions. It gives students a reason to use language precisely (MP6). It gives the teacher an opportunity to hear how students use terminology and talk about characteristics of the items in comparison to one another.

Before students begin, consider establishing a small, discreet hand signal that students can display when they have an answer they can support with reasoning. Signals might include a thumbs-up or a certain number of fingers that tells the number of responses they have. Using such subtle signals is a quick way to see if students have had enough time to think about the problem. It also keeps students from being distracted or rushed by hands being raised around the class.

Standards

Addressing	HSS-ID.A.1
Building Towards	HSS-ID.A.2

Instructional Routines

- Which Three Go Together?

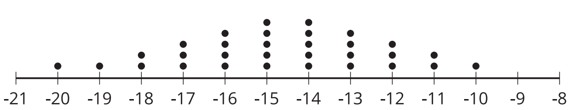
Launch

Arrange students in groups of 2–4. Display the dot plots for all to see. Give students 1 minute of quiet think time, and ask them to indicate when they have noticed three dot plots that go together and can explain why. Next, tell students to share their response with their group, and then together to find as many sets of three as they can.

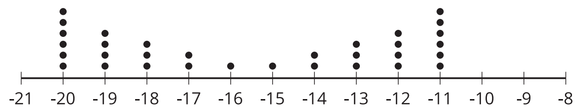
Student Task Statement

Which three go together? Why do they go together?

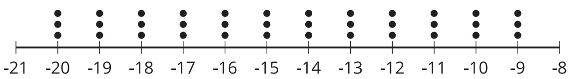
A.



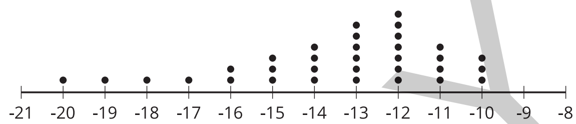
B.



C.



D.



Student Response

Sample responses:

- A, B, and C go together because they are symmetric (not skewed to one side).
- A, B, and D go together because they have different amounts of dots over different numbers (they are not uniform).
- A, C, and D go together because the median is a good description of typical values in the distribution.
- B, C, and D go together because the data are not clustered near the center.

Activity Synthesis

Invite each group to share one reason why a particular set of three goes together. Record and display the responses for all to see. After each response, ask the class if they agree or disagree. Since there is no single correct answer to the question of which three go together, attend to students' explanations and ensure that the reasons given are correct.

During the discussion, prompt students to explain the meaning of any statistical terminology they use, such as "**symmetric**," "**skewed**," "**uniform**," "**bimodal**," and "**bell-shaped**," and to clarify their reasoning as needed. Consider asking:

- "How do you know . . . ?"
- "What do you mean by . . . ?"
- "Can you say that in another way?"

15 mins

4.2 Card Sort: Matching Distributions

Activity Narrative

The mathematical purpose of this activity is to give students a chance to practice finding data displays that represent the distribution of the same data set and to practice using precise vocabulary for describing the shape of the distributions. In this partner activity, students take turns matching displays of distributions. As students trade roles, they explain their thinking and listen to the reasoning of a partner. This provides students with opportunities both to explain their reasoning and to critique the reasoning of others (MP3).

This is the first *Card Sort* routine of the course. An important aspect of this routine is to allow students time at the start to sort the cards into categories of their choosing. This step gives students the opportunity to familiarize themselves with the content of the cards without the additional pressure of organizing them in a specific fashion. It also provides insight into the aspects of each card that students attend to and the language that they use to describe their observations.

Standards

Addressing	**HSS-ID.A.1**
Building Towards	**HSS-ID.A.2**

Instructional Routines

- Card Sort
- MLR7: Compare and Connect
- Take Turns

Launch

Arrange students in groups of 2, and distribute pre-cut cards. Allow students to familiarize themselves with the representations on the cards.

1. Give students 1 minute to sort the cards into categories of their choosing.
2. Pause the class after students have sorted the cards.
3. Select groups to share their categories and how they sorted their cards.
4. Discuss as many different types of categories as time allows.

Attend to the language that students use to describe their categories and distributions, giving them opportunities to describe their distributions more precisely. Highlight the use of terms like "dot plot," "histogram," "symmetric," or "center." After a brief discussion, invite students to complete the remaining questions.

Student Task Statement

Take turns with your partner matching 2 different data displays that represent the distribution of the same set of data.

1. For each set that you find, explain to your partner how you know it's a match.
2. For each set that your partner finds, listen carefully to their explanation. If you disagree, discuss your thinking and work to reach an agreement.
3. When finished with all ten matches, describe the shape of each distribution.

Student Response

Possible student responses describing the shape of each distribution:

1. C. The shape of the distribution looks like a bell. It has a line of symmetry in the middle.
2. I. The shape of the distribution is bell-shaped. It is symmetric.
3. E. The distribution is shaped like a triangle reflected by a horizontal line at 4.5.
4. J. The distribution is bimodal. It is almost symmetric.
5. B. The shape of the distribution is a "U" where the maximum value is more frequent than the minimum value.
6. G. The distribution is shaped like a rectangle.
7. F. The shape of the distribution is uniform.
8. D. The shape of the distribution is bell-shaped, but it has a value to the extreme right.

9. H. The shape looks symmetric, except for the lowest value.
10. A. The shape looks like a fork that is symmetric around the middle value.

Building on Student Thinking

For students having trouble with the uniform distribution histograms, remind them that the lower bound for each interval is included and the upper bound is not. Ask them why this might change the last bar in each of these histograms. Some students may not know where to start to match data displays. You can tell them to look at the lowest and highest values as a starting point for finding similarities between two representations.

Activity Synthesis

Once all groups have completed the matching, discuss the following:

- "Which matches were tricky? Explain why." (The uniform distributions may be difficult.)
- "What vocabulary was useful to describe the shape of the distribution?" (symmetric, skewed, uniform, bimodal, bell-shaped)
- "Were there any matches that could be described by more than one of these vocabulary terms?" (Yes, symmetric or skewed can also be used with some of the other terms for some of the distributions.)

If necessary, ask students to revoice less formal descriptions of the shape of the distribution using formal language, including:

- Symmetric distribution
- Skewed distribution
- Uniform distribution
- Bimodal distribution
- Bell-shaped distribution

Access for English Language Learners

MLR7 Compare and Connect. Invite students to prepare a visual display that shows the strategy they used to sort the cards. Encourage students to include details that will help others interpret their thinking, for example: specific language, using different colors, shading, arrows, labels, notes, diagrams, or drawings. Give students time to investigate each other's work. During the whole-class discussion, ask students, "What kinds of additional details or language helped you understand the display?" or "Did anyone sort the cards the same way, but explain it differently?"
Advances: Representing, Conversing

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support organization, provide students with a graphic organizer with the phrase "shapes of distributions" at the center. During the *Synthesis*, prompt students to draw arrows connecting the central phrase to all the related concepts mentioned during the discussion.
Supports accessibility for: Language, Organization

4.3 Where Did The Distribution Come From?

10 mins

Activity Narrative

The mathematical purpose of this activity is to remind students of the importance of context to statistics. Although some analysis can be done outside of a context, it is often useful to think about the real situations in which the data were collected, to engage student intuition and understanding.

Standards

Addressing	HSS-ID.A.1
Building Towards	HSS-ID.A.2

Instructional Routines

- MLR2: Collect and Display

Launch

Keep students in the same groups. Assign each pair of students one of the completed matches from the *Card Sort* activity. Tell students that there are many possible answers for each representation. After 2 minutes of quiet work time, ask students to compare their responses to their partner's and decide if they are both reasonable. You may need to demonstrate this activity before beginning if you think students may have trouble getting started. After each group finishes with their assigned distribution, assign the group another distribution to consider.

Access for English Language Learners

MLR2 Collect and Display. Circulate, listen for, and collect the language that students use as they share their educated guesses of the survey question that produced the data for their assigned matches. On a visible display, record collected words and phrases, such as "line of symmetry," "bell-shaped," "skewed," "bimodal," or "uniform." Invite students to borrow language from the display as needed, and update it throughout the lesson.
Advances: Conversing, Reading

Access for Students with Disabilities

Action and Expression: Develop Expression and Communication. To help get students started, display sentence frames, such as "A possible survey question that produced this data is _____ because _____."
Supports accessibility for: Language, Organization

Student Task Statement

Your teacher will assign you some of the matched distributions. Using the information provided in the data displays, make an educated guess about the question that produced this data. Be prepared to share your reasoning.

Student Response

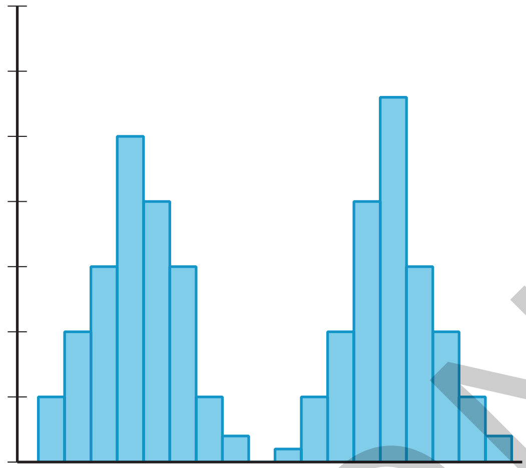
Sample responses:

1. How many feet away did the ball go when Priya tried to throw it at a target 6 feet away?
2. How many meters below sea level are sharks found swimming?

3. How many friends did you know at the birthday party?

Are You Ready for More?

This distribution shows the length in inches of fish caught and released from a nearby lake.



1. Describe the shape of the distribution.
2. Make an educated guess about what could cause the distribution to have this shape.

Extension Student Response

1. Sample response: The distribution is bimodal. It is approximately symmetric, but possibly slightly skewed to the right. Most of the data are not near the center of the distribution.
2. Sample response: There could be two different types of fish in the lake, one of which is typically shorter than the other.

Activity Synthesis

Ask each group to share their response for at least one of the distributions they were assigned. After each group shares, ask the class if the described context is reasonable. Here are some questions for discussion:

- “How did you use the shape of the data to come up with your question?” (Because the data distribution is bell-shaped, I tried to think of situations where most of the data would be similar with a few points a little away from these values.)
- “Would you always expect your question to result in a [symmetric, skewed, bell-shaped, etc.] distribution?” (Not necessarily, but for most cases it would.)

Reveal the actual question that produced the distribution. Actual questions by row:

1. How many points did Kiran score in each of his 22 games this season?
2. What were typical low temperatures in a Siberian town during January?
3. On a scale of 1–8, how was the service at the restaurant?

4. How many questions did people get correct on the vocabulary test the first week of school?
5. How many questions did people get correct on the vocabulary test the second week of school?
6. How many feet below the surface were each of the core samples taken?
7. How many trees are in my backyard at various temperatures?
8. What was the sum when you spun a spinner labeled 0 to 5 twice? (Note that the value at 25 could not actually result from this process, but could be from erroneous data collection. This will be studied with outliers later in the unit.)
9. What was the weight of the crystal you grew in chemistry class?
10. How many questions did students get correct on a 10-item matching test?

Ask students to share what they have learned about the distribution now that they can think of the data in a real situation.

Lesson Synthesis

In this lesson students describe the shape of distributions using formal language and invent contexts for distributions with different shapes. Here are some questions for discussion.

- "What does a symmetric data set look like?" (It will have a line of symmetry in the middle and the left side will look like a reflection of the right side.)
- "What does it mean to say that the shape of a distribution is uniform?" (There will be an equal number of each data value and the shape will look rectangular.)
- "Have you heard of a bell curve before? How does this relate to a bell-shaped distribution?" (Yes. I have heard of it in science class where a bell curve was used to compare data in an experiment.)
- "What is an example of a context where you would expect to find a bimodal distribution?" (You might find it if you measured the weight of a herd of cows in the springtime. The adult cows would be one peak and the calves would be the other peak.)
- "Can a skewed distribution also be symmetric? Why or why not?" (No, because skewed means that one side of the peak of the data has more data values further away from the peak than the other side. There is no line of symmetry.)

4.4 Distribution Types

Cool-down

5 mins

Standards

Addressing HSS-ID.A.1

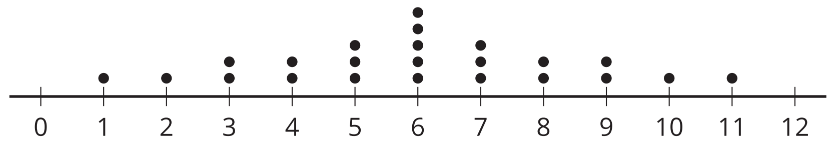
Building Towards HSS-ID.A.2

Student Task Statement

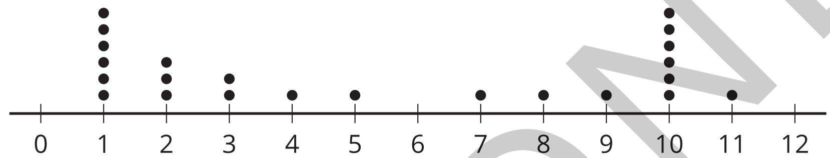
Describe each of these distributions. If more than one term applies, include all the terms that describe each distribution. Where possible, use the terms:

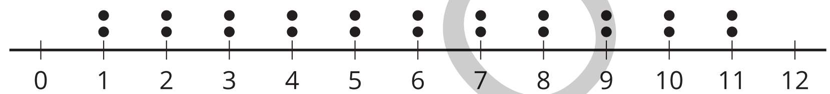
- Symmetric distribution

- Skewed distribution
- Bell-shaped distribution
- Uniform distribution
- Bimodal distribution

1. 

2. 

3. 

4. 

5. Which of these distributions is most likely to show data collected while studying the number of plates people use while eating at an all-you-can-eat buffet? Explain your reasoning.

Student Response

1. Symmetric, bell-shaped with center near 6
2. Skewed right with center near 4
3. Skewed right, bimodal with center near 6
4. Symmetric, uniform with center near 6
5. Sample response: The second dot plot, since most people will use between 1 and 3 plates while a rare person might go back again and again to use up to nine plates. The other distributions don't make much sense for this context.

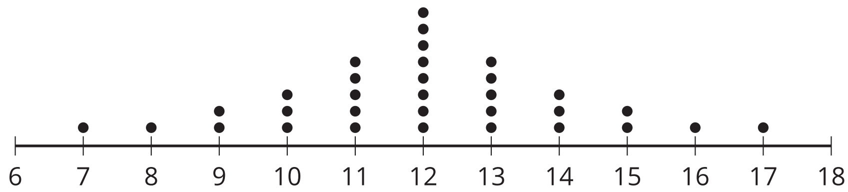
Responding To Student Thinking

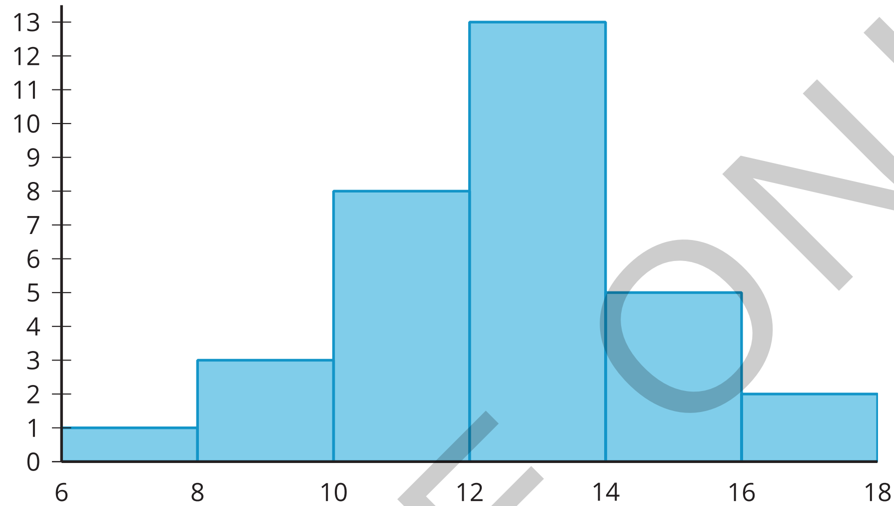
Points to Emphasize
If most students struggle with identifying the shape of the distributions, as opportunities arise, focus on students who are using precise language. For example, as students plot the survey data in the activity referred to here, ask them to identify the shape of the distribution.

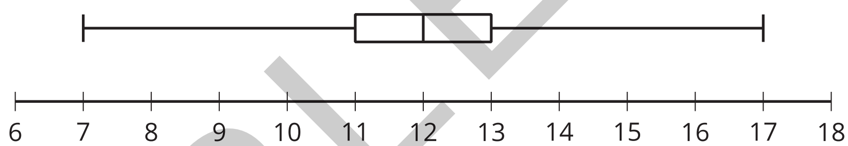
Algebra 1, Unit 1, Lesson 9, Activity 3 Making Digital Displays

Lesson 4 Summary

We can describe the shape of distributions as *symmetric*, *skewed*, *bell-shaped*, *bimodal*, or *uniform*. The dot plot, histogram, and box plot shown here represent the distribution of the same data set. This data set has a symmetric distribution.

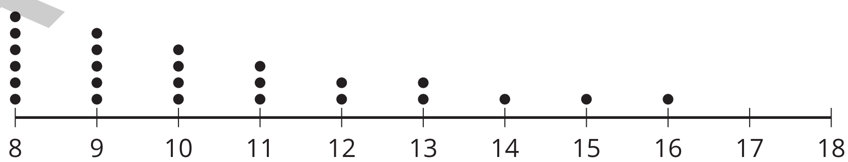


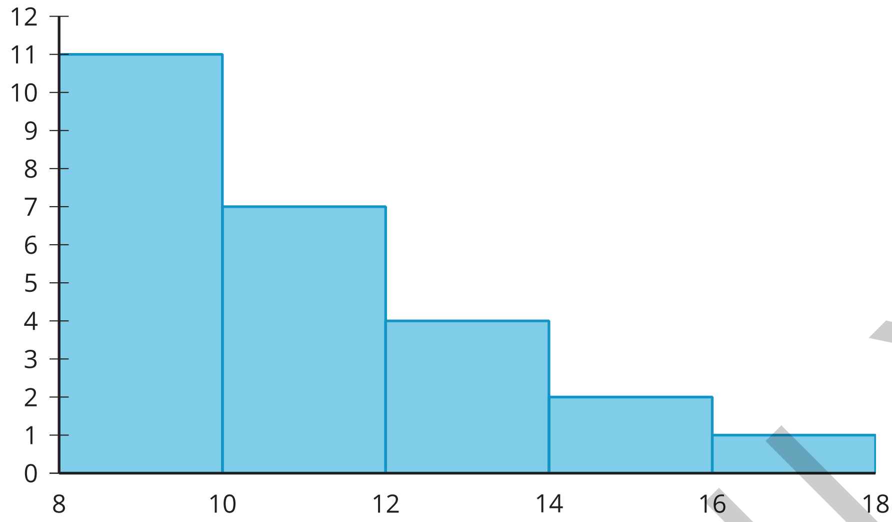


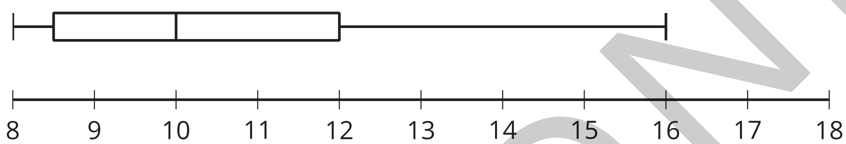


In a **symmetric distribution**, the mean is equal to the median and there is a vertical line of symmetry in the center of the data display. The histogram and the box plot both group data together. Since histograms and box plots do not display each data value individually, they do not provide information about the shape of the distribution to the same level of detail that a dot plot does. This distribution, in particular, can also be called bell-shaped. A **bell-shaped distribution** has a dot plot that takes the form of a bell with most of the data clustered near the center and fewer points farther from the center. This makes the measure of center a very good description of the data as a whole. Bell-shaped distributions are always symmetric or close to it.

The dot plot, histogram, and box plot shown here represent a skewed distribution.

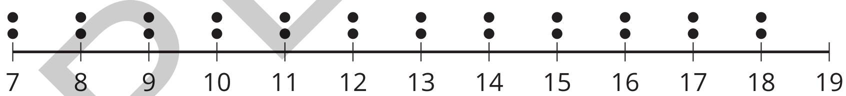


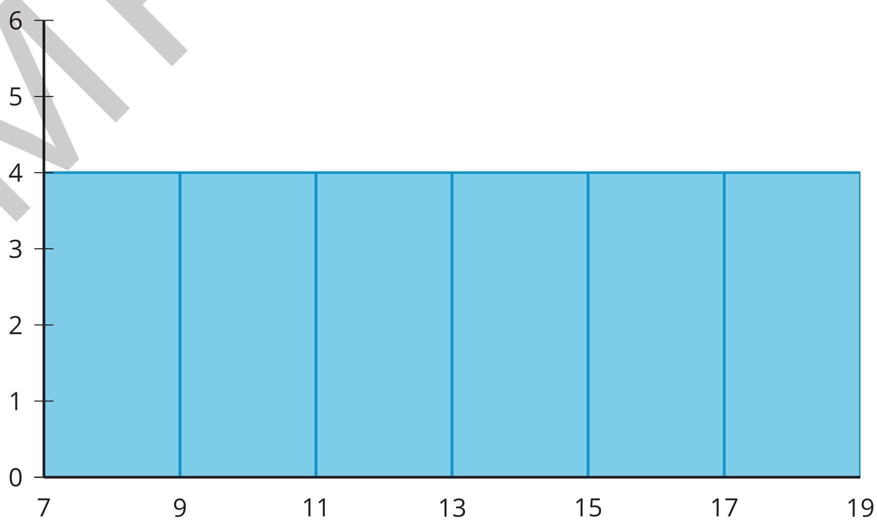


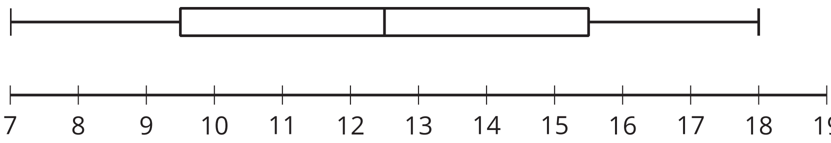


In a **skewed distribution**, one side of the distribution has more values farther from the bulk of the data than the other side has. This results in the mean and median not being equal. In this skewed distribution, the data are skewed to the right because most of the data are near the 8 to 10 interval, but there are many points to the right. The mean is greater than the median. The large data values to the right cause the mean to shift in that direction while the median remains with the bulk of the data. So, the mean is greater than the median for distributions that are skewed to the right. In a data set that is skewed to the left, a similar effect happens but to the other side. Again, the dot plot provides a greater level of detail about the shape of the distribution than do either the histogram or the box plot.

A **uniform distribution** has the data values evenly distributed throughout the range of the data. This causes the distribution to look like a rectangle.

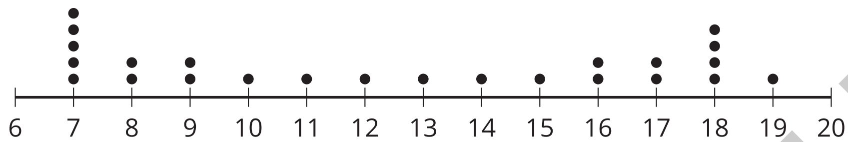


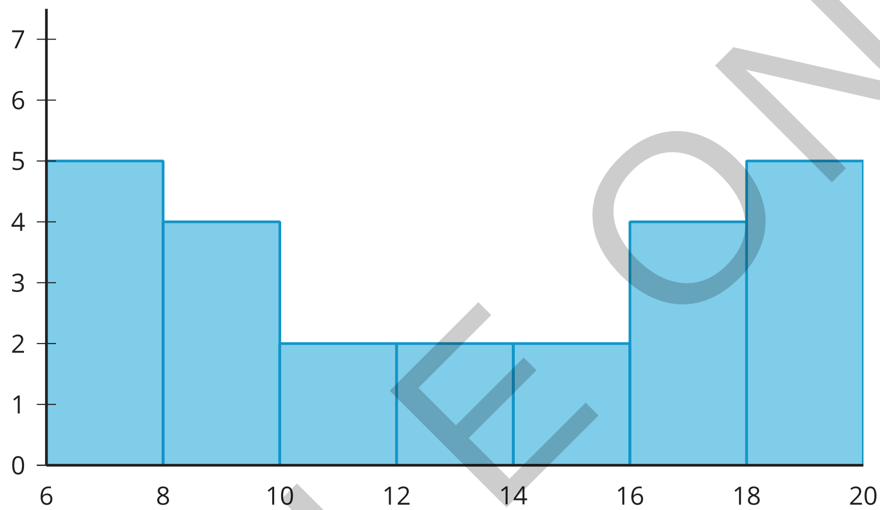


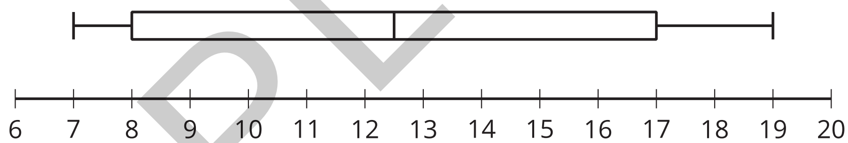


In a uniform distribution, the mean is equal to the median because a uniform distribution is also a symmetric distribution. The box plot does not provide enough information to describe the shape of the distribution as uniform, though the even length of each quarter does suggest that the distribution may be approximately symmetric.

A **bimodal distribution** has two very common data values seen in a dot plot or histogram as distinct peaks.







Sometimes, a bimodal distribution has most of the data clustered in the middle of the distribution. In these cases, the center of the distribution does not describe the data very well. Bimodal distributions are not always symmetric. For example, the peaks may not be equally spaced from the middle of the distribution, or other data values may disrupt the symmetry.

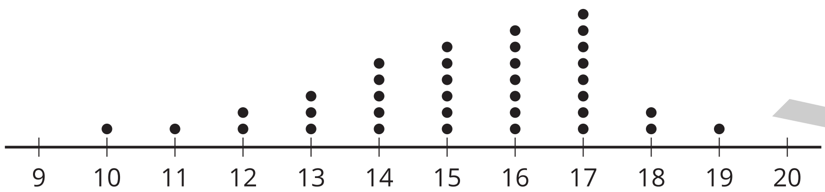
Glossary

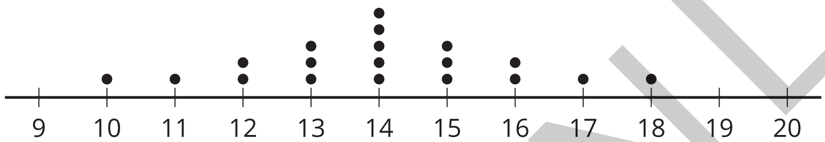
- bell-shaped distribution
- bimodal distribution
- skewed distribution
- symmetric distribution
- uniform distribution

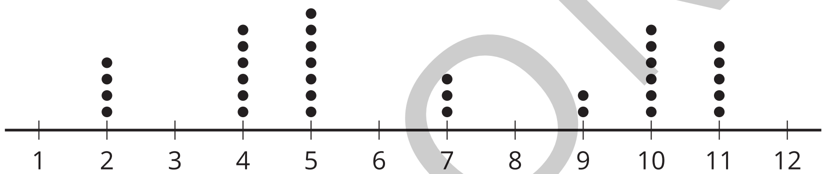
Practice Problems

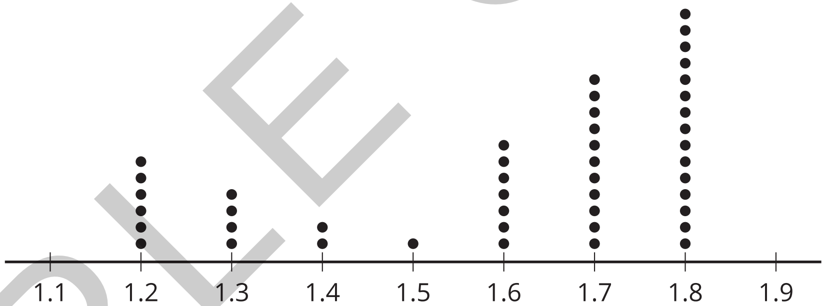
1 Student Task Statement

Which of the dot plots shows a symmetric distribution?

A. 

B. 

C. 

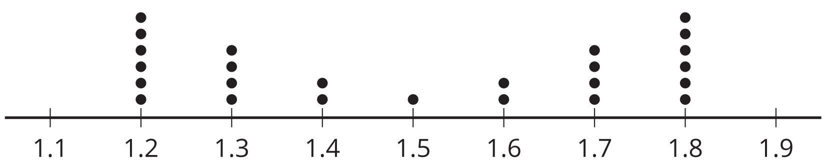
D. 

Solution

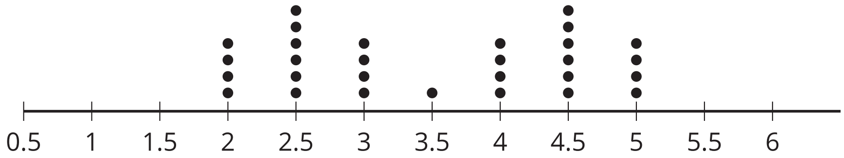
B

2 Student Task Statement

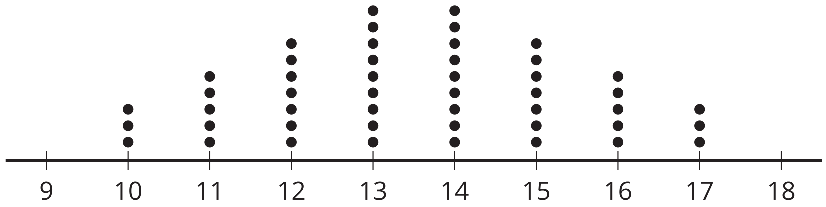
Which of the dot plots shows a skewed distribution?

A. 

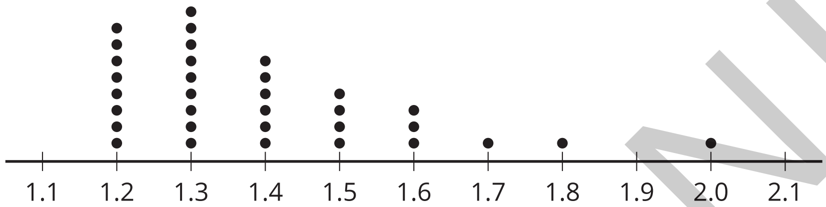
B.



C.



D.



Solution

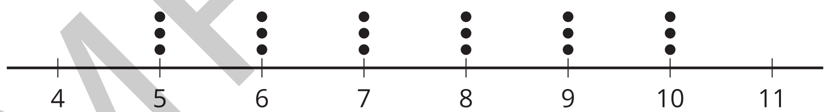
D

3 Student Task Statement

Create a dot plot showing a uniform distribution.

Solution

Sample solution:



4 from Unit 1, Lesson 3

Student Task Statement

The data represent the number of ounces of water that 26 students drank the day before a test at school:

8 8 8 16 16 16 32 32 32 32 32 32 64 64 64 64 64 64 64 80

80 80 80 88 88 88

a. Create a dot plot for the data.

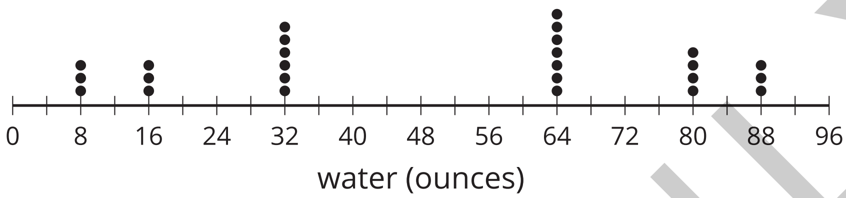
b. Create a box plot for the data.

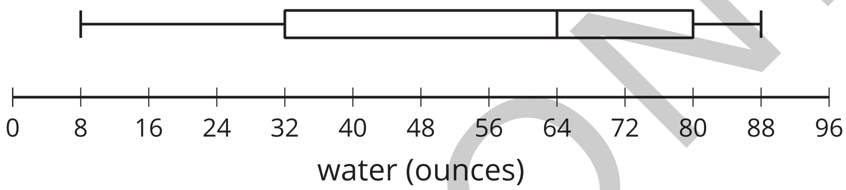
c. What information about the data is provided by the box plot that is not provided by the dot plot?

d. What information about the data is provided by the dot plot that is not provided by the box plot?

e. It was recommended that students drink 48 or more ounces of water. How could you use a histogram to easily display the number of students who drank the recommended amount?

Solution

a. 

b. 

c. Sample response: The box plot displays the median and the quartiles. To find those values using a dot plot would require calculation.

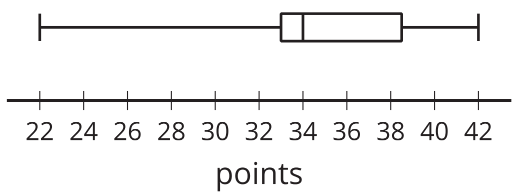
d. Sample response: The dot plot displays all the data and can be used to determine the most frequent value or values.

e. Sample response: A histogram with intervals from 0 to 48 and 48 to 96 would easily display this information.

5 from Unit 1, Lesson 2

Student Task Statement

The box plot represents the distribution of the number of points scored by a cross country team at 12 meets.



a. If possible, find the mean. If not possible, explain why not.

b. If possible, find the median. If not possible, explain why not.

c. Did the cross country team ever score 30 points at a meet?

Solution

a. Sample response: It is not possible to calculate the mean because the box plot does not provide the 12 scores. It provides only a summary of the scores.

b. 34 points

c. Sample response: It is not possible to tell if they scored 30 points.

Calculating Measures of Center and Variability

Goals

- Calculate mean absolute deviation, interquartile range, mean, and median.

Learning Targets

- I can calculate mean absolute deviation, interquartile range, mean, and median for a set of data.

Lesson Narrative

This lesson is optional because it revisits content that is below grade level. If the Check Your Readiness assessment indicates that students know how to calculate the mean, median, mean absolute deviation (MAD), and interquartile range (IQR), then this lesson may be safely skipped. When students explain the MAD using the meter stick example, they are reasoning abstractly and quantitatively by interpreting the MAD in context (MP2).

Standards

Building On	**6.SP.B.5.c**
Building Towards	**HSS-ID.A.1, HSS-ID.A.2**

Instructional Routines

- MLR8: Discussion Supports

Required Materials

Materials To Copy

- Heartbeats Part 1 Handout (1 copy for every 2 students): Activity 2
- Algebra 1 Unit 1 Useful Terms and Displays (1 copy for every 30 students): Activity 3

Required Preparation

Activity 3:

You will need a meter stick and 14 pennies (or other small weights) for a demonstration.

Lesson:

An optional blackline master is included as a graphic organizer for computing interquartile range. One copy of the blackline master contains two graphic organizers.

Student Facing Learning Goals

Let's calculate measures of center and measures of variability and know which are most appropriate for the data.

5.1 Calculating Centers

Warm-up

5 mins

Activity Narrative

The purpose of this *Warm-up* is to encourage students to review how to calculate mean and median.

Standards

Building On **6.SP.B.5.c**

Building Towards **HSS-ID.A.1, HSS-ID.A.2**

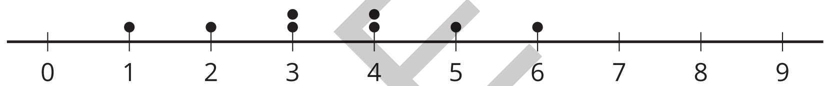
Launch

Display one problem at a time. Give students 1 minute of quiet think time followed by a whole-class discussion.

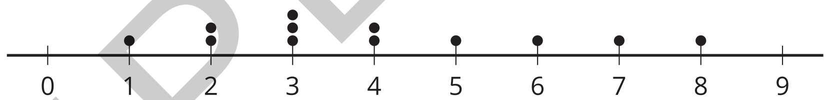
Student Task Statement

Decide if each situation is true or false. Explain your reasoning.

1. The mean can be found by adding all the numbers in a data set and dividing by the number of numbers in the data set.
2. The mean of the data in the dot plot is 4.



3. The median of the data set is 9 for the data: 4, 5, 9, 1, 10.
4. The median of the data in the dot plot is 3.5.



Student Response

1. True. Sample reasoning: The statement is the correct way to calculate the mean.
2. False. Sample reasoning: The mean is 3.5 because the sum is 28 and $28 \div 8 = 3.5$. Do not round to 4.
3. False. Sample reasoning: Nine is not the median because the numbers are not arranged in order.
4. True. Sample reasoning: The median is 3.5 because the two values in the middle are 3 and 4, and their mean is 3.5.

Building on Student Thinking

Some students may forget to sort the data when finding the median. Ask them, “What is a median? What does it tell you about the data?” Some students may not remember how to find the median when there is an even number of data values. Ask them, “What does the median tell you about the data? How could we find a middle number between these two values?”

Activity Synthesis

The goal of this activity is to review how to calculate mean and median and to identify common mistakes in the calculation of mean and median. In the discussion, ask students to recall what information the mean and median reveal about the data.

- “What does the mean tell you about the data?” (On average, where the center of the data is.)
- “What does the median tell you about the data?” (Half of the numbers are greater than or equal to the median, and half are less than or equal to the median.)

Some common mistakes to avoid:

- Not putting the numbers in order when finding the median.
- If two numbers are in the middle, not adding them and dividing by two to find the median.
- Finding the middle number on the horizontal axis rather than in the data.
- Rounding the mean or median (if it must be calculated) to the nearest whole number.

5.2 Heartbeats: Part 1

Optional

10 mins

Activity Narrative

The purpose of this activity is to get students to calculate the median and IQR, and to investigate how those values are affected by outliers.

The data sets in this lesson are small enough that finding summary statistics like measures of center or measures of variability are not necessary. The entire data set could be assessed fairly easily and does not need further analysis. The sets are small here for the purposes of practicing calculating and understanding the statistics. In reality, finding such statistics are much more useful when the data set is much larger. For example, some devices will find heart rate every 10 minutes, giving more than 1,000 values for a week, which might be analyzed to determine health information about a person. It would be difficult to understand all of the data by looking at a table or even a data display. Having some summary statistics, such as mean or median, would be useful to understand a person’s heart rate for the week.

In this activity, an optional graphic organizer is provided in the blackline master to help students compute the interquartile range. The boxes that are shaded are not to be used, and the position of the open boxes are meant to highlight the useful data for calculating the value for the row. For example, to compute the first quartiles, students use the average of the second and third values, so those boxes are left open to indicate that this data are useful.

Standards

Building On	**6.SP.B.5.c**
Building Towards	**HSS-ID.A.2**

Launch

Give students 3–5 minutes of quiet time to answer the first question, and then pause for a brief whole-class discussion about how to calculate the median, quartiles, and IQR.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support organization, provide students with the graphic organizer for students to organize their work with calculating the median, quartiles, and IQR.
Supports accessibility for: Language, Organization

Student Task Statement

The heart rates of eight high school students are listed in beats per minute:

72 75 81 76 76 77 79 78

1. What is the interquartile range?
2. How many values in the data set are:
   a. less than Q1?
   b. between Q1 and the median?
   c. between the median and Q3?
   d. greater than Q3?
3. A pod of dolphins contains 800 dolphins of various ages and lengths. The median length of dolphins in this pod is 5.8 feet. What information does this tell you about the length of dolphins in this pod?
4. The same vocabulary test with 50 questions is given to 600 students from fifth to tenth grades and the number of correct responses is collected for each student in this group. The interquartile range is 40 correct responses. What information does this tell you about the number of correct responses for students taking this test?

Student Response

1. 3 beats per minute
2.
   a. 2
   b. 2
   c. 2
   d. 2
3. Sample response: Half of the dolphins are 5.8 feet or longer, and half are 5.8 feet or shorter.
4. Sample response: There is a lot of variability in the number of correct responses for this group. The middle half of the data spanned an interval of 40 correct responses, and since there are only 50 questions, this indicated there is a lot of variability.

Building on Student Thinking

Students may have difficulty calculating the median of a data set with an even number of data points. Ask them what the median represents for the data set and where that value might be. If they still struggle, remind them that the median is the average of the two middle numbers.

Activity Synthesis

The purpose of this discussion is to discuss the method of calculating median and IQR as well as the interpretation of each. Here are some questions for discussion.

- "What would a box plot look like for the results of the vocabulary test?" (It would depend on the high and the low score, but the box would likely start at 5 and go up to 45 or it could start at 10 and go up to 50.)
- "What does the IQR tell you about a data set?" (The IQR is a measure of variability, which describes the range of the middle half of the data).
- "How do the median and quartiles divide the data?" (Into equal quarters so that the same number of values are in each quarter.)
- "How much of the data are between Q1 and Q3?" (The middle 50% of the data are between these values.)

5.3 Heartbeats: Part 2

Optional

15 mins

Activity Narrative

The purpose of this activity is to get students to calculate and describe the mean absolute deviation. Students are given a data set and an organizer for calculating the MAD. Then they consider questions that are intended to get them thinking about MAD more conceptually as a measure of variability.

Standards

Building On	**6.SP.B.5.c**
Building Towards	**HSS-ID.A.2**

Instructional Routines

- MLR8: Discussion Supports

Launch

Help students understand how to use the table by showing students the example table:

data values	mean	deviation from the mean (data value – mean)	absolute deviation |deviation|
1	5	-4	4
2	5	-3	3
3	5	-2	2
4	5	-1	1
5	5	0	0
6	5	1	1
7	5	2	2
12	5	7	7

This results in a MAD of 2.5 because $\frac{4+3+2+1+0+1+2+7}{8} = 2.5$.

Access for English Language Learners

MLR8 Discussion Supports. To support the transfer of new vocabulary to long-term memory, invite students to chorally repeat these phrases in unison 1–2 times: "deviation from the mean" "absolute deviation."
Advances: Speaking

Access for Students with Disabilities

Representation: Internalize Comprehension. Begin with a physical demonstration of using pennies on a meter stick to support connections between new situations and prior understandings. Consider using the prompts: “What does this demonstration have in common with measures of center and variability?” or “How is the balance point related to the mean or median?”
Supports accessibility for: Conceptual Processing, Visual-Spatial Processing

Student Task Statement

1. Calculate the MAD using the same data from the previous activity by finding the average distance from each data value to the mean. You may find it helpful to organize your work by completing the table provided.

data values	mean	deviation from the mean (data value - mean)	absolute deviation |deviation|
72			
75			
81			
76			
76			
77			
79			
78			

MAD:

2. For another data set, all of the values are either 3 beats per minute above the mean or 3 beats per minute below the mean. Is that enough information to find the MAD for this data set? If so, find the MAD. If not, what other information is needed? Explain your reasoning.
3. Several pennies are placed along a meter stick, and the position in centimeters of each penny is recorded. The mean position is the 50 centimeter mark and the MAD is 10 centimeters. What information does this tell you about the position of the pennies along the meter stick?

Student Response

1. 2 beats per minute
2. It is enough information. The MAD is 3 beats per minute since the average distance to the mean is 3 beats per minute.
3. Sample response: The pennies are spaced so that, on average, they are about 10 centimeters away from the 50 centimeter mark.

Building on Student Thinking

Monitor for students who have trouble finding the mean or who are using negative values for the distance from the mean. Remind them that the $|x|$ symbol represents absolute value. If necessary, ask them to look at a number line and describe the distance between two values to remind them that distances should always be described as positive values.

Are You Ready for More?

Suppose there are 6 pennies on a meter stick so that the mean position is the 50 centimeter mark and the MAD is 10 centimeters.

1. Find possible locations for the 6 pennies.
2. Find a different set of possible locations for the 6 pennies.

Extension Student Response

1. Sample response: 3 pennies at the 40 centimeter mark and 3 pennies at the 60 centimeter mark.
2. Sample response: 1 penny at the 35 centimeter mark, 1 penny at the 40 centimeter mark, 1 penny at the 45 centimeter mark, 1 penny at the 55 centimeter mark, 1 penny at the 60 centimeter mark, 1 penny at the 65 centimeter mark.

Activity Synthesis

Put pennies on a meter stick so that the centers of the pennies are at {20, 40, 40, 45, 45, 45, 45, 55, 55, 55, 55, 55, 55, 90}. Show how the stick balances when you put your finger at the 50 centimeter mark and how some are farther and some are closer than 10 centimeters away from the mean, but they're spread out so that, on average, they're 10 cm away.

Sec B

Create a display that incorporates the measures of center (mean and median) and variability (interquartile range and mean absolute deviation) discussed so far. This display should be posted in the classroom for the remaining lessons within this unit. You will add to the display throughout the unit. The blackline master provides an example of what this display may look like after all items are added.

Here are some questions for discussion.

- "If you put pennies at 45, 35, and 70, where do you need to put a penny for the meter stick to balance at 50 cm? What is the MAD?" (50 cm, 10 cm)
- "If you put two pennies at 60, where do you need to put a penny to make the meter stick balance at 50 cm? What is the MAD?" (30 cm, $\frac{40}{3}$ cm)

Lesson Synthesis

The goal is to make sure students know how to calculate MAD and IQR and that they are measures of variability.

- "How do you calculate the IQR?" (Find the difference between the value of the third and first quartiles.)
- "How do you calculate the MAD?" (Find the distance of each value from the mean. Then take the mean of those values.)
- "One data set has a greater IQR than another. What does this mean about the data in the first data set?" (It has greater variability. This means that the middle half of the data are more spread out from the center than the middle half of the data from the second data set are.)

5.4 Calculating MAD and IQR

Cool-down

5 mins

Standards

Building On 6.SP.B.5.c

Building Towards HSS-ID.A.2

Student Task Statement

- 5
- 18
- 6
- 18
- 13

mean: 12

1. Find the mean absolute deviation for the data.
2. Find the interquartile range for the data.

Student Response

1. 5.2
2. 12.5

Responding To Student Thinking

Points to Emphasize
If most students struggle with calculating the mean absolute deviation or the interquartile range, make time to revisit these calculations as part of the activity referred to here. For example, invite 2–3 students to share how they calculated these values for their survey data and where the values can be seen in the different displays.

Algebra 1, Unit 1, Lesson 9, Activity 3 Making Digital Displays

Lesson 5 Summary

The *mean absolute deviation*, or MAD, and the *interquartile range*, or IQR, are measures of variability. Measures of variability tell you how much the values in a data set tend to differ from one another. A greater measure of variability means that the data are more spread out, while a smaller measure of variability means that the data are more consistent and are closer to the measure of center.

To calculate the MAD of a data set:

1. Find the mean of the values in the data set.
2. Find the distance between each data value and the mean (on the number line):

$$|\text{data value} - \text{mean}|$$

3. Find the mean of the distances. This value is the MAD.

To calculate the IQR, subtract the value of the first quartile from the value of the third quartile. Recall that the first and third quartile are included in the five-number summary.

Sec B

Practice Problems

1 Student Task Statement

The data set represents the number of errors on a typing test.

5 6 8 8 9 9 10 10 10 12

a. What is the median? Interpret this value in the situation.

b. What is the IQR?

Solution

a. 9 errors. Half of the people had 9 errors or more and half had 9 errors or fewer.

b. 2 errors

Sec B

2 Student Task Statement

The data set represents the heights, in centimeters, of ten model bridges made for an engineering competition.

13 14 14 16 16 16 16 18 18 19

a. What is the mean?

b. What is the MAD?

Solution

a. 16 centimeters

b. 1.4 centimeters

3 from Unit 1, Lesson 4

Student Task Statement

Describe the shape of the distribution shown in the dot plot. The dot plot displays the golf scores from a golf tournament.

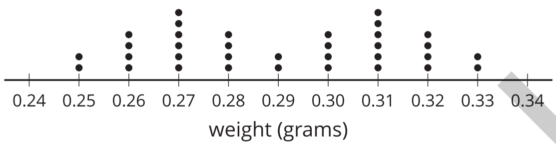
Solution

Skewed right centered near 136 strokes

4 from Unit 1, Lesson 4

Student Task Statement

The dot plot shows the weight, in grams, of several different rocks. Select **all** the terms that describe the shape of the distribution.



A. bell-shaped

B. bimodal

C. skewed

D. symmetric

E. uniform

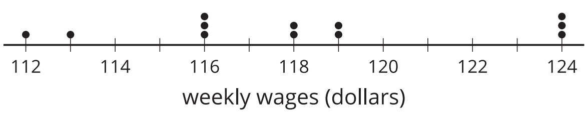
Solution

B, D

5 from Unit 1, Lesson 3

Student Task Statement

The dot plot represents the distribution of wages earned during a one-week period by 12 college students.



a. What is the mean? Interpret this value based on the situation.
b. What is the median? Interpret this value based on the situation.
c. Would a box plot of the same data have allowed you to find both the mean and the median?

Solution

a. $118.25. Sample interpretation: This means that if all 12 students put their earnings for the week together and then redistributed the money equally, each person would get $118.25.

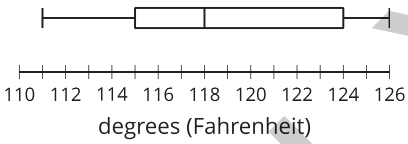
b. $118. This means that half of the students earned $118 or more, and half earned $118 or less.

c. No, it would not have been possible to find the mean.

6 from Unit 1, Lesson 2

Student Task Statement

The box plot displays the temperature of saunas in degrees Fahrenheit. What is the median?



Sec B

Solution

118 degrees Fahrenheit

Section C: How to Use Spreadsheets

Goals

- Use a spreadsheet as a tool for calculating values quickly.

Section Narrative

In this optional section students are introduced to using spreadsheets to create basic formulas and interpret results. It is assumed that students know nothing about spreadsheets. If students are proficient at using spreadsheets as computation tools, this sequence of lessons can be safely skipped. This technology is useful for interpreting data in this unit and is also used to examine functions in a subsequent unit.

Because this entire section is optional, no checkpoint questions are included.

	A	B	C	D	E
1				=B2-D5	
2		99			
3					
4	-17				
5				0.25	
6					
7					

Teacher Reflection Questions

- **Math Content and Student Thinking:** What understandings about spreadsheets are important for students to bring to future work on analyzing data and interpreting functions?
- **Pedagogy:** Think about a recent activity that went longer than planned. If you could go back and teach it again, would you make the same choices?
- **Access and Equity:** In Lesson 6, Activity 2: Mystery Operations, some sentence frames are offered as a support to encourage peer interaction and discourse. Which of your students could this accommodation support and in what ways?

Section C Checkpoint

Teacher Instructions

There are no Checkpoint questions for this optional section.

Unit 1, Lesson 6

Mystery Computations

Sec C

Goals

- Use cell references in a spreadsheet to provide arguments for mathematical operations.

Learning Targets

- I can determine basic relationships between cell values in a spreadsheet by changing the values and noticing what happens in another cell.

Lesson Narrative

This lesson is optional because it provides instructions for using spreadsheets.

To engage in this lesson, students must have access to internet-enabled devices. Ideally, they should navigate to the digital version of the materials with the embedded applets. A direct link to each applet is also provided.

Students experiment by changing numbers in a provided spreadsheet and discover what it does and how. Then, they have an opportunity to write their own formula. When students observe the outcome of different inputs and then generalize the operations in the formula cell, they are expressing regularity in repeated reasoning (MP8).

Instructional Routines

- MLR8: Discussion Supports

Required Materials

Materials To Gather

- Internet-enabled device: Activity 2, Activity 3
- Spreadsheet technology: Activity 2, Activity 3

Student Facing Learning Goals

Let's explore spreadsheets.

6.1 Make 24

Warm-up

5 mins

Activity Narrative

In the next activity, students are asked to look for a rule that puts together some numbers to make a target number. This *Warm-up* activates that sort of thinking.

Launch

Arrange students in groups of 2.

Ask each student to think of a whole number. (Make this more straightforward by limiting the choices to single-digit numbers, more challenging by limiting to 0–20 or 0–100, or really challenging by placing no restrictions on the numbers.) Select four students at random to share their number, and display these four numbers for all to see.

Student Task Statement

Your teacher will give you 4 numbers. Use these numbers, along with mathematical operations like addition and multiplication, to make 24.

Student Response

Sample responses:

- If the 4 shared numbers are 1, 2, 3, and 4. $1 \cdot 2 \cdot 3 \cdot 4 = 24$.
- If the 4 shared numbers are 1, 2, 3, and 5. $2 \cdot 3 \cdot (5 - 1)$.
- If the 4 shared numbers are 5, 5, 5, and 5. $5 \cdot 5 - \frac{5}{5} = 24$.

Sec C

Activity Synthesis

Ask students to share their expression with a partner, so that their partner can verify that their expression is equal to 24. Then, select a few students to share their expression with the whole class and record these for all to see. How many different ways did the class find to make 24?

6.2 Mystery Operations

Optional

10 mins

Activity Narrative

There is a digital version of this activity.

In this activity, students experiment with an applet in which the value displayed in some cells depends on other cells. Monitor for students who notice interesting things.

Instead of writing down responses to the questions, it would make sense for students to share their observations verbally with a partner.

Some students will probably notice that the answers are in the spreadsheet. For example, if you click on cell B2, you can see the formula that was used to generate the number that appears in that cell. They should be encouraged to use the values of the other cells to verify what they notice. Ultimately, all students should be able to interpret the spreadsheet formula in the cell.

In the digital version of the activity, students use an applet to discover the formulas within spreadsheet cells. The applet allows students to work with a Geogebra spreadsheet directly. Use the digital version if it is available so that students are more familiar with Geogebra.

Instructional Routines

- MLR8: Discussion Supports

Launch

Distribute internet-enabled devices, and give students instructions to navigate to this lesson in the digital version of the materials.

Ensure that students have access to the spreadsheet applet and that they understand how to type entries into cells. Call their attention to the *Task Statement*, which instructs them to try different numbers in column A and conjecture about how the numbers in column B are calculated.

Monitor for students who try interesting things, are able to predict what will be in a cell in column B, or can explain how a cell in column B is calculated.

Access for English Language Learners

MLR8 Discussion Supports. Display sentence frames to support partner discussion:

- "I noticed that ________, so I ________."
- "First, I ________ because ________."

Advances: Speaking, Representing

Access for Students with Disabilities

Engagement: Develop Effort and Persistence. Encourage and support opportunities for peer interactions. Invite students to talk about their ideas with a partner before writing them down. Display sentence frames to support students when they explain their strategy. For example,

- "I notice that the cells in column B ________."
- "I tried ________, and what happened was ________."

Supports accessibility for: Language, Social-Emotional Functioning

Student Task Statement

Navigate to this activity in the digital version of the materials, or to ggbm.at/fjcybyqf.

Input different numbers in column A, and try to predict what will happen in column B. (Do not change anything in column B.)

1. How is the number in cell B2 related to all or some of the numbers in cells A2, A3, A4, and A5?
2. How is the number in cell B3 related to all or some of the numbers in cells A2, A3, A4, and A5?
3. How is the number in cell B4 related to all or some of the numbers in cells A2, A3, A4, and A5?
4. How is the number in cell B5 related to all or some of the numbers in cells A2, A3, A4, and A5?

Student Response

1. The number in cell B2 is the sum of the numbers in A3 and A4.
2. The number in cell B3 is the product of the numbers in A3 and A4.

3. The number in cell B4 is the sum of all four numbers in column A.
4. The number in cell B5 is the product of all four numbers in column A.

Building on Student Thinking

Students may be confused about how the term "related" is being used in the task statement. Explain that students should be trying to figure out what calculations take place with the numbers in column A in order to get the numbers in column B. Students may have difficulty figuring out the relationships between the numbers in column A and the numbers in column B. Suggest they try different kinds of numbers in column A, like small whole numbers, 0, multiples of 10, or decimal values.

Activity Synthesis

After students have had a chance to experiment with the applet and make conjectures about how the cells in column B are calculated, invite them to compare their conjectures with a partner.

Before selecting students to share their conclusions, select students to share their predicted value for various cells in column B based on combinations of values in column A. For example, "What would be the value in cell B4 if column A has the values 2, 4, 5, and 9?" (20)

Select students who tried or noticed interesting things to share with the whole class.

The words "sum" and "product" appear in the formulas. Briefly ask students to recall the meaning of these words. Invite them to guess what using the : symbol in the cell formula does. For example, =Sum(A2:A5) calculates A2 + A3 + A4 + A5.

6.3 More Spreadsheets!

Optional

15 mins

Activity Narrative

There is a digital version of this activity.

The purpose of this activity is to improve student familiarity with writing spreadsheet formulas based on references to other cells. In this activity, students have an opportunity to write their own spreadsheet formula in a cell and trade with a partner to guess their formula.

In the digital version of the activity, students use an applet to play with formulas in a spreadsheet. The applet allows students to directly manipulate the formulas and see the results immediately. Use the digital version if it is available so that students can be familiar with the spreadsheets available to them in this course.

Instructional Routines

- MLR8: Discussion Supports

Launch

Access for English Language Learners

MLR8 Discussion Supports. During group work, invite students to take turns sharing their responses. Ask students

to restate what they heard using precise mathematical language and their own words. Display the sentence frame: "I heard you say . . . ." Original speakers can agree or clarify for their partner.
Advances: Listening, Speaking

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support development of organizational skills in problem solving, chunk this task into more manageable parts. For example, present one question at a time.
Supports accessibility for: Memory, Organization

Student Task Statement

Navigate to this activity in the digital version of the materials or to ggbm.at/wu9t7kkd.

1. Change the spreadsheet so that B2 contains =A2+A4. To edit the formula in B2, you may have to click it twice.
2. Change the numbers in A2 through A5. Make sure that your new formula does what it is supposed to do by doing a mental calculation and checking the result in B2.
3. Change the contents of B3 so that B3 does something different.
4. Before trading with a partner, make sure your new formula is not visible by clicking in a different cell.
5. Trade with your partner.
6. Change the numbers in Column A to try and figure out your partner's new rule.

Student Response

No written responses are required.

Building on Student Thinking

Students may have difficulty inputing a new formula in a cell. For example, students may forget to begin the formula with =. Ensure that students know how to input the symbols that represent each operation.

Activity Synthesis

Ask students to report on the work they did with their partner.

- "Which formulas were interesting to guess?"
- "Which formulas were difficult to guess? What made them difficult?"
- "Which formulas were easy to guess? What made them easy?"

6.4 What Does This Do?

Cool-down

5 mins

Student Task Statement

	A	B	C
1	2		=(A1 + A2)*A3
2	5		
3	7		
4	12		
5	18		
6			
7			

1. What value will be in cell C1 after the formula is entered?
2. What value will be in cell C1 if the numbers in column A are changed to 1, 2, 3, 4, and 5 in that order?

Sec C

Student Response

1. 49
2. 9

Responding To Student Thinking

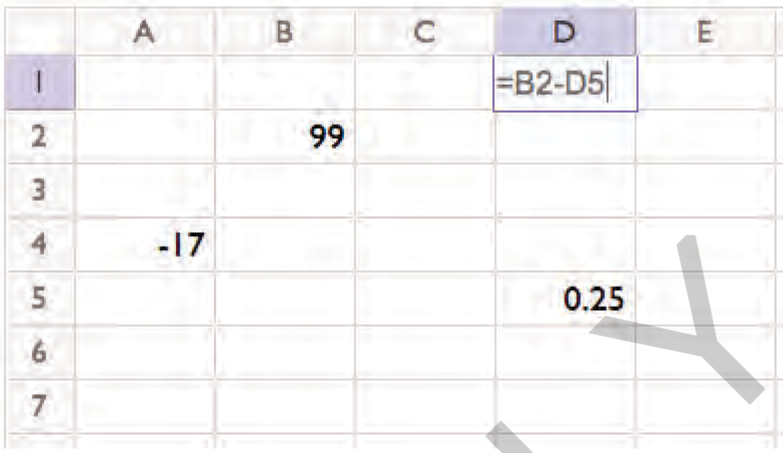
More Chances
Students will have more opportunities to understand the mathematical ideas addressed here. There is no need to slow down or add additional work to the next lessons.

Lesson 6 Summary

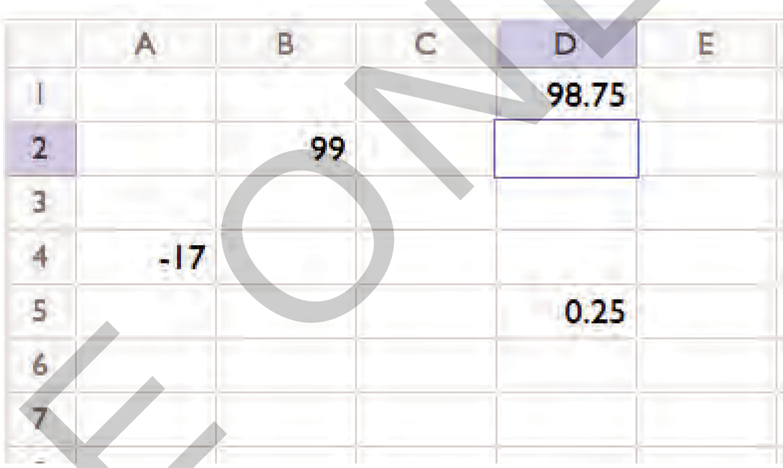
Spreadsheets are useful mathematical and statistical tools. Here is an example of a spreadsheet. Each cell in the spreadsheet can be named with its column and row. For example, cell B2 contains the value 99. Cell A4 contains the value -17. Cell D1 is selected.

	A	B	C	D	E
1					
2		99			
3					
4	-17				
5				0.25	
6					
7					

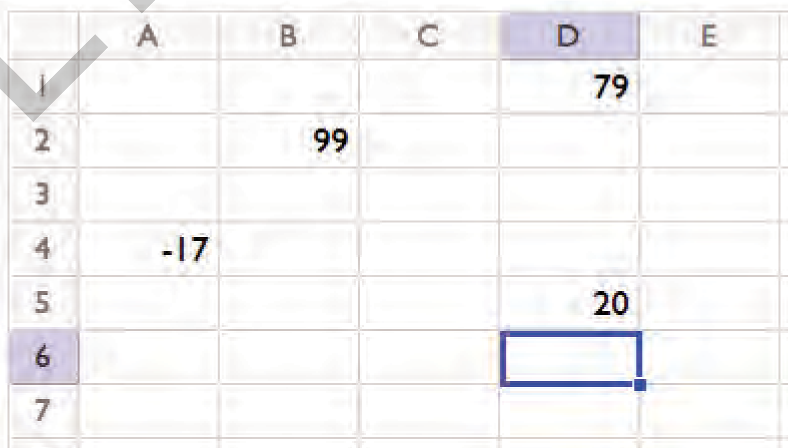
It is possible for the value in a cell to depend on the value in other cells. Let's type the formula $= B2 - D5$ into cell D1.



When we press enter, D1 will display the result of subtracting the number in cell D5 from the number in cell B2.



If we type new numbers into B2 or D5, the number in D1 will automatically change.



Sec C

Practice Problems

1 Student Task Statement

What could be the formula used to compute the value shown in cell B3?

	A	B
1	change these	what happens here?
2	7	20
3	0	350
4	13	0
5	50	69
6	-1	

A. = B3 * B4

B. = A2 + A5

C. = A2 * A5

D. = Sum(A2:A6)

Solution

C

2 Student Task Statement

What number will appear in cell B2 when the user presses Enter?

	A	B
1	change these	what happens here?
2	10	=Sum(A3:A5)
3	5	
4	0	
5	-7	

Solution

-2

3 Student Task Statement

Select **all** the formulas that could be used to calculate the value in cell B4.

	A	B
1	change these	what happens here?
2	7	20
3	0	350
4	13	0
5	50	69
6	-1	

A. =Product(A2:A6)

B. =Sum(A2:A6)

C. = A2 + A3

D. = A2 * A3

E. = A3 * A4 * A5

F. = A3 + A4 + A5

Solution

A, D, E

4 Student Task Statement

The formula in cell B2 is = Product(A2 : A5). Describe a way to change the contents of column A so that the value in cell B2 becomes -70.

	A	B
1	change these	what happens here?
2	10	0
3	5	
4	0	
5	-7	

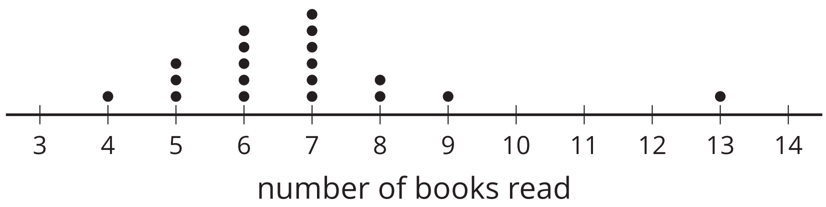
Solution

Sample response: Change A3 to 1 and A4 to 1.

5 from Unit 1, Lesson 5

Student Task Statement

The dot plot displays the number of books read by students during the semester.



a. Which measure of center would you use given the shape of the distribution in the dot plot? Explain your reasoning.
b. Which measure of variability would you use? Explain your reasoning.

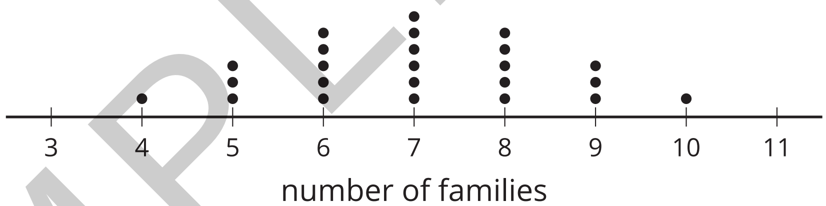
Solution

a. Sample response: I would use the median because the data set has a skewed distribution. The median is not impacted as much by the presence of outliers as is the mean.
b. Sample response: I would use the IQR because it is the measure of variability used with the median.

6 from Unit 1, Lesson 5

Student Task Statement

The dot plot displays the number of families living in different blocks of a town.



a. Which measure of center would you use, given the shape of the distribution in the dot plot? Explain your reasoning.
b. Which measure of variability would you use? Explain your reasoning.

Solution

a. Sample response: I would use the mean, because the data are symmetric.
b. Sample response: I would use the MAD, because it is calculated using the mean.

Unit 1, Lesson 7

Spreadsheet Computations

Goals

- Use a spreadsheet to calculate values in word problems.

Learning Targets

- I can use a spreadsheet as a calculator to find solutions to word problems.

Lesson Narrative

This lesson is optional.

In this lesson, students learn to use a spreadsheet to perform basic operations with numbers. They also begin to notice the power of spreadsheets in such calculations by using the dynamic nature of the autocalculating features.

Instructional Routines

- Analyze It
- MLR7: Compare and Connect

Required Materials

Materials To Gather

- Internet-enabled device: Activity 2, Activity 3, Activity 4
- Spreadsheet technology: Activity 2, Activity 3, Activity 4

Student Facing Learning Goals

Let's use spreadsheets as calculators.

7.1 Dust Off Those Cobwebs

Warm-up

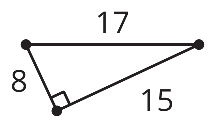
5 mins

Activity Narrative

The purpose of this *Warm-up* is to remind students about some familiar mathematical contexts, to prepare them for engaging with some word problems later in the lesson.

Student Task Statement

1. A person walks 4 miles per hour for 2.5 hours. How far does the person walk?
2. A rectangle has an area of 24 square centimeters. What could be its length and width?
3. What is the area of this triangle?



Student Response

1. 10 miles
2. Any two lengths in centimeters whose product is 24 square centimeters
3. 60 square units

Sec C

7.2 A Spreadsheet Is a Calculator

Optional

10 mins

Activity Narrative

The purpose of this activity is for students to learn and practice how to use a spreadsheet to do calculations using some common operations, including some illustrations of how parentheses can be used to indicate order of operations.

Instructional Routines

- Analyze It
- MLR7: Compare and Connect

Launch

Ensure that each student has access to a spreadsheet. If using the digital version of the materials, there is a spreadsheet in the Math Tools.

Display the same type of spreadsheet students are using (or the student task) on your projector, and demonstrate how to use it as a calculator. Be sure to communicate the following.

- The = symbol often must be typed as the very first thing in the cell. (Demonstrate what happens if the = symbol is not included.)
- How to "submit" the formula so the computation takes place. If student devices have a keyboard, it's likely the Enter key. On a tablet, they may have to tap a check mark.
- Symbols to use for various operations, and how students can find them on their keyboards. Be sure to include these symbols:
  - + for add
  - - for subtract or for a negative number (this symbol does double duty in most spreadsheets)

- * for multiply
- / for divide or for a fraction
- ^ for an exponent
- . for a decimal point
- ( ) to tell it what to compute first (often needed around fractions)

Consider displaying these commands for all to see and leaving them visible while students work on the activity.

Access for Students with Disabilities

Representation: Develop Language and Symbols. Provide students with access to charts with symbols and meanings. The chart should include all the symbols to use for various operations in a spreadsheet. Be sure to demonstrate using the commands in the spreadsheet as a calculator.
Supports accessibility for: Conceptual Processing, Memory

Student Task Statement

Use a spreadsheet to compute each of the following. Type each computation in a new cell, instead of erasing a previous computation.

- $2+7$
- $2-7$
- $7 \cdot 2$
- 7^2
- $7 \div 2$
- $\frac{1}{7}$ of 91
- $0.1 \cdot 2 + 3$
- $0.1(2+3)$
- $13 \div \frac{1}{7}$
- The average of 2, 7, 8, and 11

Student Response

Here is an example of what you might type to do each computation:

- = 2 + 7
- = 2 - 7
- = 7 * 2
- = 7 ^ 2
- = 7 / 2
- = 1 / 7 * 91
- = 0.1 * 2 + 3
- = 0.1 * (2 + 3)
- = 13 / (1 / 7)
- = MEAN(2,7,8,11), =AVERAGE(2,7,8,11), or = (2+7+8+11)/4

Building on Student Thinking

If students want to modify their formula, they should try to double click on the cell. If students forget which symbol to use for an operation or how to use the keyboard to type it, draw their attention to the display from the *Launch*.

Activity Synthesis

Ask students to compare their answers with a partner and resolve any discrepancies. Help students understand how to use parentheses to get the spreadsheet to perform the desired calculation. For example, to compute $7 \div \frac{1}{2}$, you may have to type = 7 / (1 / 2).

Note that the average of 2, 7, 8, and 11 can be calculated by typing =(2+7+8+11)/4 but also by typing =MEAN(2,7,8,11) or =Average(2,7,8,11), depending on the spreadsheet software.

Access for English Language Learners

MLR7 Compare and Connect. After all strategies for what was typed to do each computation have been presented, lead a discussion comparing, contrasting, and connecting the different approaches. Ask, “Did anyone solve the problem the same way, but would explain it differently?” or “Why did the different approaches lead to the same outcome?”
Advances: Representing, Conversing

7.3 Use the Contents of a Cell in a Calculation

5 mins

Optional

Sec C

Activity Narrative

The purpose of this activity is for students to understand how to use references to cells to calculate values. In later lessons, students may wish to use spreadsheets to perform similar calculations on different numbers and using references can save time so that students do not need to retype all the calculations.

Launch

Display a blank spreadsheet. In cell B3, type a 2 and press enter. Point to that cell and ask students how to refer to that cell. Tell students that people often call B3 the "address" of the cell, and 2 the "contents" of the cell.

Student Task Statement

1. Type any number in cell A1, and another number in cell A2. Then in cell A3, type =A1+A2. What happens?
2. In cell A4, compute the product of the numbers in A1 and A2.
3. In cell A5, compute the number in A1 raised to the power of the number in A2.
4. Now, type a new number in cell A1. What happens?
5. Type a new number in cell A2. What happens?
6. Use nearby cells to label the contents of each cell. For example, in cell B3, type "the sum of A1 and A2." (This is a good habit to get into. It will remind you and anyone else using the spreadsheet what each cell means.)

Student Response

1. A3 shows the result of adding the numbers in A1 and A2.
2. Type = A1 * A2 in cell A4.
3. Type = A1^A2 in cell A5.

4. Cells A3, A4, and A5 update to use the new value in A1.
5. Cells A3, A4, and A5 update to use the new value in A2.
6. No response needed.

7.4 Solve Some Problems

Optional

15 mins

Activity Narrative

The purpose of this activity is to demonstrate that spreadsheets can be efficient when repeating similar calculations with different values. A single spreadsheet formula can refer to the contents of cells and the calculation will automatically update when the values change.

If time is short, the first two questions can be skipped or you can tell students to choose only one of them. It's important that all students complete the last two questions to illustrate why a spreadsheet can be more useful than a handheld calculator when you need to repeat the same calculation for different values.

Launch

Access for Students with Disabilities

Representation: Access for Perception. Read all situations aloud. Students who both listen to and read the information will benefit from extra processing time.
Supports accessibility for: Language

Student Task Statement

For each problem:

- Estimate the answer before calculating anything.
- Use the spreadsheet to calculate the answer.
- Write down the answer and the formula that you used in the spreadsheet to calculate it.

1. The speed limit on a highway is 110 kilometers per hour. How much time does it take a car to travel 132 kilometers at this speed?
2. In a right triangle, the lengths of the sides that make a right angle are 98.7 cm and 24.6 cm. What is the area of the triangle?
3. A recipe for fruit punch uses 2 cups of seltzer water, $\frac{1}{4}$ cup of pineapple juice, and $\frac{2}{3}$ cup of cranberry juice. How many cups of fruit punch are in 5 batches of this recipe?
4. Check in with a partner, and resolve any discrepancies with your answer to the last question. Next, type 2, $\frac{1}{4}$, $\frac{2}{3}$, and 5 in separate cells. (You may find it helpful to label cells next to them with the meaning of each number.) In a blank cell, type a formula for the total amount of fruit punch that uses the values in the other four cells. Now you should be able to easily figure out:
   a. How much in 7.25 batches?

b. How much in 5 batches if you change the recipe to 1.5 cups of seltzer water per batch?

c. Change the ratio of the ingredients in the fruit punch so that you would like the flavor. How many total cups are in $\frac{1}{2}$ batch?

Student Response

1. 1.2 (= 132 / 110)
2. 1,214.01 square centimeters (= 0.5 * 98.7 * 24.6)
3. approximately 14.58 (= 5 * (2 + 1/4 + 2/3))
4. If A1=2, A2=1/4, A3=2/3, A4=5, then the formula is = (A1 + A2 + A3) * A4
   a. about 21.15 cups
   b. about 12.08 cups
   c. Answers vary.

Sec C

7.5 Good Old Raisins and Peanuts

Cool-down

5 mins

Student Task Statement

Diego's family is going on a camping trip and his job is to make a batch of GORP (Good Old Raisins and Peanuts) for a snack to take on the trip. In his kitchen, he finds many identical boxes of raisins and many identical bags of peanuts. He puts all of this information in a spreadsheet:

	A	B	C	D
1	ounces of raisins in each box	3.5		
2	number of boxes	12		
3	ounces of peanuts in each bag	4		
4	number of bags	18		
5				
6				
7				

1. Explain how Diego could use the spreadsheet to figure out how many total ounces of GORP he can make.
2. Diego decides that he doesn't need that much GORP. Explain how he could use the spreadsheet to figure out how many boxes of raisins and how many bags of peanuts he needs to make around 60 ounces of GORP.

Student Response

Possible responses:

1. In cell B5, type = B1 * B2 + B3 * B4
2. He could decrease the numbers in cells B2 and B4 until the total number of ounces was around 60.

Responding To Student Thinking

More Chances
Students will have more opportunities to understand the mathematical ideas addressed here. There is no need to slow down or add additional work to the next lessons.

Lesson 7 Summary

A spreadsheet can be thought of as a type of calculator. For example, in a cell, you could type $= 2 + 3$, and then the sum of 5 is displayed in the cell. You can also perform operations on the values in other cells. For example, if you type a number in A1 and a number in A2, and then in A3 type $= A1 + A2$, cell A3 will display the sum of the values in cells A1 and A2.

Familiarize yourself with how your spreadsheet software works on your device.

- On some spreadsheet programs, an = symbol must be typed before the expression in the cell. (On others, it does not matter if your expression begins with =.)
- Know how to "submit" the expression so the computation takes place. If your device has a keyboard, it's likely the Enter key. On a touchscreen device, you may have to tap a check mark.
- Learn symbols to use for various operations, and how to find them on your keyboard. Here are the symbols used for some typical operations:
    - + for add
    - - for subtract or for a negative number (this symbol does double duty in most spreadsheets)
    - * for multiply
    - / for divide
    - a / b for the fraction $\frac{a}{b}$
    - ^ for an exponent
    - . for a decimal point
    - ( ) to tell it what to compute first (often needed around fractions)

Practice Problems

1 Student Task Statement

Write a formula that you could type into a spreadsheet to compute the value of each expression.

a. $(19.2) \cdot 73$

b. 1.1^5

c. $2.34 \div 5$

d. $\frac{91}{7}$

Solution

Sample responses:

a. = 19.2 * 73

b. = 1.1 ^ 5

c. = 2.34 / 5

d. = 91 / 7

Sec C

2 Student Task Statement

A long-distance runner jogs at a constant speed of 7 miles per hour for 45 minutes. Which spreadsheet formula would give the distance she traveled?

A. = 7 * 45

B. = 7 / 45

C. = 7 * (3 / 4)

D. = 7 / (3 / 4)

Solution

C

3 Student Task Statement

In a right triangle, the lengths of the sides that make a right angle are 3.4 meters and 5.6 meters. Select **all** the spreadsheet formulas that would give the area of this triangle.

A. = 3.4 * 5.6

B. = 3.4 * 5.6 * 2

C. = 3.4 * 5.6 / 2

D. = 3.4 * 5.6 * (1/2)

E. = (3.4 * 5.6) / 2

Solution

C, D, E

4

Student Task Statement

This spreadsheet should compute the total ounces of sparkling grape juice based on the number of batches, ounces of grape juice in a single batch, and ounces of sparkling water in a single batch.

	A	B
1	number of batches	4
2	ounces of grape juice in 1 batch	3
3	ounces of sparkling water in 1 batch	7
4	total ounces	

a. Write a formula for cell B4 that uses the values in cells B1, B2, and B3, to compute the total ounces of sparkling grape juice.

b. How would the output of the formula change if the value in cell B1 was changed to 10?

c. What would change about the sparkling grape juice if the value in B3 was changed to 10?

Solution

Sample responses:

a. = B1 * (B2 + B3)

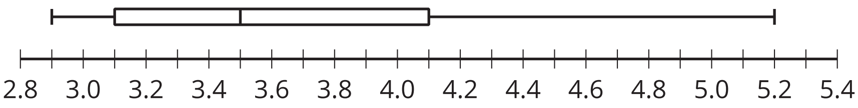
b. It would be 100 instead of 40.

c. The sparkling grape juice would be waterier or less strongly flavored.

5 from Unit 1, Lesson 5

Student Task Statement

The dot plot and the box plot represent the same distribution of data.



a. How does the median change when the highest value, 5.2, is removed?

b. How does the IQR change when the highest value, 5.2, is removed?

Solution

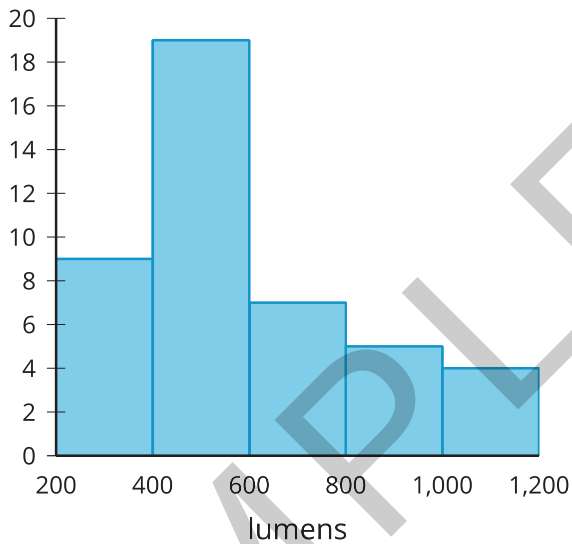
a. Sample response: The median remains 3.5.

b. Sample response: The IQR changes from 1 to 0.9 when 5.2 is removed from the data set.

6 from Unit 1, Lesson 4

Student Task Statement

Describe the shape of the distribution shown in the histogram, which displays the light output, in lumens, of various light sources.



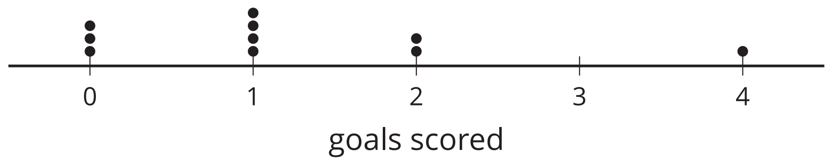
Solution

Skewed right with a center near 600 lumens.

7 from Unit 1, Lesson 2

Student Task Statement

The dot plot represents the distribution of the number of goals scored by a soccer team in 10 games.



a. If possible, find the mean. If not possible, explain why not.
b. If possible, find the median. If not possible, explain why not.
c. Did the soccer team ever score exactly 3 goals in one of the games?

Solution

a. 1.2 goals
b. 1 goal
c. No

Unit 1, Lesson 8

Spreadsheet Shortcuts

Goals

- Use spreadsheet shortcuts such as clicking on cells or filling cells with a pattern.

Learning Targets

- I can use shortcuts to fill in cells on a spreadsheet.

Lesson Narrative

This lesson is optional.

In this lesson, students learn to include cell references in formulas by clicking on a cell instead of typing its address and also how to continue patterns by dragging to generate a sequence of numbers based on a formula.

To engage in this lesson, students must have access to internet-enabled devices. Ideally, they should navigate to the digital version of the materials with the embedded applets. A direct link to each applet is also provided.

Sec C

Required Materials

Materials To Gather

- Internet-enabled device: Activity 2, Activity 3
- Spreadsheet technology: Activity 2, Activity 3

Student Facing Learning Goals

Let's explore recursive formulas in spreadsheets.

8.1 Tables of Equivalent Ratios

Warm-up

10 mins

Activity Narrative

The purpose of this activity is to prepare students to work with a table of equivalent ratios that appears in a later activity in this lesson.

Launch

Give students a few minutes to complete the table. Ensure that they found the correct values before they work on the second question.

Student Task Statement

Here is a table of equivalent ratios:

a	b
3	15
10	50
6	30
1	
	80

1. Complete the table with the missing values.
2. Explain what it means to say that the pairs of numbers are equivalent ratios.

Student Response

a	b
3	15
10	50
6	30
1	5
16	80

1. See table.
2. Possible response: Each value in column b results from multiplying the value in column a by 5.

Activity Synthesis

Invite students to share how they reasoned about the values needed to complete the table. Keep asking "Did anyone think about it a different way?" until a few ways of reasoning come to light. Some methods might include:

- Multiplying any number in the first column by 5 to find the number in the second column
- Dividing any number in the second column by 5 (or multiplying by $\frac{1}{5}$ to find the number in the first column)
- Using a scale factor to move from row to row, for example, starting with 6 and 30, you can multiply each value by $\frac{1}{6}$ to find 1 and 5 in the next row.

8.2

The Birthday Trick

Optional

15 mins

Activity Narrative

There is a digital version of this activity.

The purpose of this activity is to learn that you can click on a cell to use its contents in a new expression rather than typing the address of the cell.

In the digital version of the activity, students use an applet to calculate values from other values in a spreadsheet. The applet allows students to directly type into the spreadsheet cells without navigating somewhere or opening another application. Use the digital version if it is available so that students can practice using the spreadsheet tools available in this course.

Launch

Ask students to write down the day and month of the birthday of one of their heroes, such as a local community member, famous person from history, favorite athlete, or favorite celebrity. Then, give them this sequence of instructions. Allow them to use a calculator if they wish. (Alternatively, you could have one student demonstrate the procedure on the birthday of a local community hero for the group, and then re-run the computations with each student using the birthday of a hero.)

- Multiply the month by 5.
- Add 6.
- Multiply by 4.
- Add 9.
- Multiply by 5.
- Add the day.
- Subtract 165.

Distribute internet-enabled devices, and give students instructions to navigate to this lesson in the digital version of the materials.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support organization, provide students with grid or graph paper to organize their work with the sequence of steps for the calculations.
Supports accessibility for: Language, Organization

Student Task Statement

Navigate to this activity in the digital version of the materials or to ggbm.at/djcz6fjf.

1. In cell B4, we want to enter = B1 * 5 to multiply the month by 5. Begin by entering the "=" sign, but when you are about to type B1, instead, click on cell B1. Then type in the "*5." This shortcut can be used any time: Click on a cell instead of typing its address.
2. Practice this technique as you program each cell in B5 through B10 to perform the right computation.

3. When you are finished, does cell B10 show a number that contains the month and day of your birthday? If not, troubleshoot your computations.
4. Try changing the month and day in cells B1 and B2. The rest of the computations should automatically update. If not, troubleshoot your computations.

Student Response

No written responses are required. When all the expressions are in place, the spreadsheet will look like:

	A	B	C
1	month	11	
2	day	27	
3			
4	multiply month by 5	55	
5	add 6	61	
6	multiply by 4	244	
7	add 9	253	
8	multiply by 5	1265	
9	add the day	1292	
10	subtract 165	1127	
11			

Building on Student Thinking

Some students may misunderstand the instructions and continue to refer each cell back to the number of their month. Each row after the first instruction is meant to refer back to the previous value.

Are You Ready for More?

Why does this trick work? Try using m for the month and d for the day, and writing the entire computation as an algebraic expression. Can you see why the resulting number contains the month and day?

Extension Student Response

Using m for the month and d for the day, here is how we can build an expression:

Multiply the month by 5: $5m$

Add 6: $5m + 6$

Multiply by 4: $4(5m + 6)$

Add 9: $4(5m + 6) + 9$

Multiply by 5: $5(4(5m + 6) + 9)$

Add the day: $5(4(5m + 6) + 9) + d$

Subtract 165: $5(4(5m + 6) + 9) + d - 165$

We can write equivalent expressions so that we can see what is going on. Here is one way to do that:

$$5(4(5m + 6) + 9) + d - 165$$

$$5(20m + 24 + 9) + d - 165$$
$$5(20m + 33) + d - 165$$
$$100m + 165 + d - 165$$
$$100m + d$$

Multiplying the month by 100 puts the month into the hundreds (and possibly thousands, if it is a two-digit number) place, with the d in the tens and ones place.

Activity Synthesis

Ask students what new shortcut they learned in the spreadsheet. Namely, when students want to use the contents of the cell in a computation, they can click on a cell rather than type in its address. Ask students what other things they had to remember about using spreadsheets.

Sec C

8.3 Using Spreadsheet Patterns

Optional

10 mins

Activity Narrative

There is a digital version of this activity.

The mathematical purpose of this activity is for students to become more efficient at entering values into a spreadsheet. In later units, students may wish to use the techniques in this activity to quickly compute the values of sequences by continuing the pattern.

Monitor for students who use a recursive formula, for example, typing = B2 + 2 in cell B3. Also look for students who use an explicit formula, like typing = A3 / 3 in cell B3. After students complete the first question, ask students to share their various spreadsheet formulas that work. Assure them that either is okay.

In the digital version of the activity, students use an applet to fill in many spreadsheet cells with a pattern. The applet allows students to work directly in the spreadsheet without navigating away or using another application. Use the digital version if it is available so that students may have more practice using the spreadsheet tools available in this course.

Launch

Access for Students with Disabilities

Representation: Internalize Comprehension. Use color coding and annotations to highlight connections between representations in a problem. For example, use the same color to highlight important connections between cells and formulas.
Supports accessibility for: Visual-Spatial Processing

Student Task Statement

Navigate to this activity in the digital version of the materials or to the URL, ggbm.at/wu9t7kkd.

The spreadsheet contains a table of equivalent ratios.

1. Use spreadsheet calculations to continue the pattern in columns A and B, down to row 5. Pause for discussion.
2. Click on cell A5. See the tiny blue square in the bottom right corner of the cell? Click it and drag it down for several cells and let go.
3. Repeat this, starting with cell B5.

Student Response

	A	B
1	6	2
2	12	4
3	18	6
4	24	8
5	30	10
6	36	12
7	42	14
8	48	16
9	54	18

No written response required. When finished, the spreadsheet will look like this:

Activity Synthesis

Tell students that the technique used in this activity extends a set of operations to additional cells in the spreadsheet. Invite students to experiment with extending some other patterns in the spreadsheet.

8.4

Doubling in a Spreadsheet

Cool-down

5 mins

Student Task Statement

A list of numbers is made with this pattern: Start with 3, and multiply by 2 each time.

Here is the beginning of the list of numbers: 3, 6, 12, . . .

Explain how you could use "fill down" in a spreadsheet to find the tenth number in this list. (You do *not* need to actually find this number.)

Student Response

Sample response: In cells A1, A2, and A3, type 3, 6, and 12. In cell A4, type = A3 * 2. Then, click the little square in the corner of A4 and drag it down far enough so that you can see the contents of A10.

Responding To Student Thinking

More Chances
Students will have more opportunities to understand the mathematical ideas addressed here. There is no need to slow down or add additional work to the next lessons.

Lesson 8 Summary

Sometimes you want to create a list of numbers based on a rule. For example, let's say that the cost of a gym membership is a $25 sign-up fee followed by monthly dues of $35. We may want to know how much the membership will cost over the course of 6 months. We could use a spreadsheet and set it up this way:

	A	B	C
1	sign-up fee	25	
2	total cost after 1 month	=B1+35	
3	total cost after 2 months		
4	total cost after 3 months		
5	total cost after 4 months		
6	total cost after 5 months		
7	total cost after 6 months		
8			
9			

Which results in:

	A	B	C
1	sign-up fee	25	
2	total cost after 1 month	60	
3	total cost after 2 months		
4	total cost after 3 months		
5	total cost after 4 months		
6	total cost after 5 months		
7	total cost after 6 months		
8			
9			

See the little square on the lower-right corner of cell B2? If we click and drag that down, it will keep adding 35 to the value above to find the value in the next row. Drag it down far enough, and we can see the total cost after 6 months.

	A	B	C
1	sign-up fee	25	
2	total cost after 1 month	60	
3	total cost after 2 months	95	
4	total cost after 3 months	130	
5	total cost after 4 months	165	
6	total cost after 5 months	200	
7	total cost after 6 months	235	
8			
9			

Anytime you need to repeat a mathematical operation several times, continuing a pattern by dragging in a spreadsheet might be a good choice.

Practice Problems

1 Student Task Statement

Technology required. Open a blank spreadsheet. Use "fill down" to recreate this table of equivalent ratios. You should not need to type anything in rows 3–10.

	A	B
1	3	7
2	6	14
3	9	21
4	12	28
5	15	35
6	18	42
7	21	49
8	24	56
9	27	63
10	30	70

Solution

No written response necessary.

2 Student Task Statement

A list of numbers is made with this pattern: Start with 11, and subtract 4 to find the next number.

Here is the beginning of the list: 11, 7, 3, . . .

Explain how you could use "fill down" in a spreadsheet to find the tenth number in this list. (You do *not* need to actually find this number.)

Solution

Sample response: In cell A1, type 11. In cell A2, type = A1 - 4 and press enter. Then, click the little square in the corner of A2 and drag it down far enough so that you can see the contents of A10.

3 Student Task Statement

Here is a spreadsheet showing the computations for a different version of the birthday trick:

	A	B
1	month	7
2	day	4
3		
4	multiply month by 50	
5	add 30	
6	multiply by 2	
7	add the day	
8	subtract 60	
9		
10		

Explain what formulas you would enter in cells B4 through B8 so that cell B8 shows a number representing the month and day. (In this example, cell B8 should show 704.) If you have access to a spreadsheet, try your formulas with a month and day to see whether it works.

Solution

Sample response, in cells B4 through B8:

- = B1 * 50
- = B4 + 30
- = B5 * 2
- = B6 + B2
- = B7 - 60

4 from Unit 1, Lesson 7

Student Task Statement

Write a formula that you could type into a spreadsheet to compute the value of each expression.

a. $\frac{2}{5}$ of 35

b. $25 \div \frac{5}{3}$

c. $\left(\frac{1}{11}\right)^4$

d. The average of 0, 3, and 17

Solution

Sample response:

a. = 2 / 5 * 35
b. = 25 / (5 / 3)
c. = (1 / 11) ^ 4
d. = MEAN (0, 3, 17)

5 from Unit 1, Lesson 5

Student Task Statement

The values represent the number of cars in a town given a speeding ticket each day for 10 days.

2 4 5 5 7 7 8 8 8 12

a. What is the median? Interpret this value in the situation.
b. What is the IQR?

Solution

a. 7 cars. On half of the days, 7 or fewer cars got a speeding ticket. On half of the days, 7 or more cars got a speeding ticket.
b. 3 cars

6 from Unit 1, Lesson 5

Student Task Statement

The values represent the most recent sale price, in thousands of dollars, of ten homes on a street.

85 91 93 99 99 99 102 108 110 115

a. What is the mean?
b. What is the MAD?

Solution

a. 100.1 (in thousands of dollars) or $100,100
b. 6.92 (in thousands of dollars) or $6,920

Section D: Manipulating Data

Goals

- Describe the effect of outliers on a distribution of data.
- Recognize standard deviation as a measure of variability.

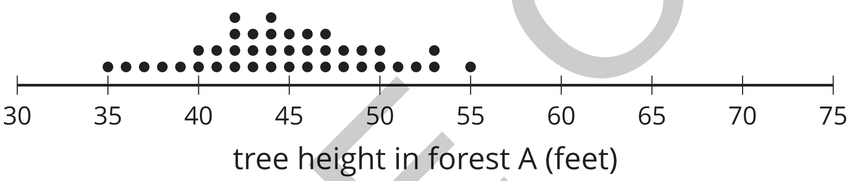
Section Narrative

In this section, students are introduced to additional tools to analyze data such as standard deviation and outliers. They finish the section by using all of these tools to compare similar data sets.

Note that in this unit, all standard deviations refer to the population standard deviation (σ) calculation rather than the sample standard deviation (s).

If students are using GeoGebra or the Spreadsheet tool in Math Tools to create a histogram, there is an issue when the maximum value is on the boundary of the greatest interval. In this case, GeoGebra includes the maximum value with the previous interval rather than following the convention of creating a new bar for the next interval. Address this issue with students if they are using either of those tools.

mean: 44.8 feet, standard deviation: 4.72 feet



Teacher Reflection Questions

- **Math Content and Student Thinking:** What prior knowledge about extremes was evident as students explored outliers? What understandings did students show about the impact of extreme values in data sets?
- **Pedagogy:** Reflect on the descriptors that students contributed in the “sounds like” and “looks like” categories in your Math Community Chart. As the year progresses, what new descriptors do you hope that students might add based on their experiences in your classroom?
- **Access and Equity:** Identify who has been sharing their ideas in class lately. Make a note of students whose ideas have not been shared, and look for an opportunity for them to share their thinking in tomorrow’s lesson.

Section D Checkpoint

1

Goals Assessed

- Recognize standard deviation as a measure of variability.

Student Task Statement

What does standard deviation measure? How would a distribution with a large standard deviation compare to a distribution with a small standard deviation?

Solution

Standard deviation measures variability. A distribution with a large standard deviation would be more spread out than a distribution with a small one.

Responding To Student Thinking

Press Pause
If most students struggle to describe the meaning of standard deviation, make time to revisit the concept. For example, plan to do the referenced optional activity about interpreting the mean and standard deviation in context. The Course Guide provides additional ideas for revisiting earlier work.

Algebra 1, Unit 1, Lesson 13, Activity 3 Interpreting Measures of Center and Variability

2

Goals Assessed

- Describe the effect of outliers on a distribution of data.

Student Task Statement

Describe what an outlier is. Do outliers affect mean or median more?

Solution

Sample response: An outlier is an extreme value in a data set that is more than 1.5 times the interquartile range away from the nearest quartile. Outliers affect mean more than median.

Responding To Student Thinking

Press Pause
If most students struggle to describe the meaning of outliers, make time to revisit the concept. For example, plan to do the referenced practice problems and ask students to describe why the values are considered outliers. The

Course Guide provides additional ideas for revisiting earlier work.

Algebra 1, Unit 1, Lesson 14, Practice Problem 1

Unit 1, Lesson 9

Technological Graphing

Goals

- Use technology tools to graphically represent data and calculate useful statistics.

Learning Targets

- I can use technology to create graphic representations of data and calculate statistics.

Lesson Narrative

In this lesson, students use technology to create data displays and encounter the term **statistic** which is a quantity that is calculated from sample data. They enter data into a spreadsheet, find statistics, and create box plots using technology. When students investigate and explain how changing values in a data set affect the mean and the median they are looking for and expressing regularity in repeated reasoning (MP8).

Math Community

Today's community building centers on the teacher sharing their draft commitments as part of the mathematical community. At the end of the lesson, students are invited to suggest additions to the teacher sections of the chart.

Standards

Building On	**HSS-ID.A.2**
Addressing	**HSS-ID.A.1, HSS-ID.A.2**
Building Towards	**HSS-ID.A.2, HSS-ID.A.3**

Instructional Routines

- Analyze It
- MLR2: Collect and Display

Required Materials

Materials To Gather

- Math Community Chart: Activity 1
- Statistical technology: Activity 1, Activity 2

Required Preparation

Activity 1:

In the "Doing Math" teacher section of the Math Community Chart, add 2–5 commitments you have for what your teaching practice "looks like" and "sounds like" this year.

Activity 3:

Students require access to data collected from survey questions in a previous lesson.

Student Facing Learning Goals

Let's use technology to represent data.

9.1 It Begins With Data

Warm-up

10 mins

Activity Narrative

The mathematical purpose of this activity is to gain familiarity with entering data into a spreadsheet and to prepare students for finding statistics using technology.

Standards

Building On **HSS-ID.A.2**

Launch

Arrange students in groups of 2. If students are using the digital version of the materials, show them how to open the GeoGebra spreadsheet app in the math tools. If students are using the print version of the materials, they can access the GeoGebra spreadsheet app at www.geogebra.org/spreadsheet. If students will be using statistical technology other than GeoGebra for this activity, prepare alternate instructions.

Make sure students input the data in one column, even though the data are represented in two columns in the task statement.

Student Task Statement

Open a spreadsheet window and enter the data so that each value is in its own cell in column A.

1. How many values are in the spreadsheet? Explain your reasoning.
2. If you entered the data in the order that the values are listed, the number 7 is in the cell at position A1 and the number 5 is in cell A5. List all of the cells that contain the number 13.
3. In cell C1 type the word "Sum," in C2 type "Mean," and in C3 type "Median." You may wish to double-click or drag the vertical line between columns C and D to allow the entire words to be seen.

	A
1	7
2	8
3	4
4	13
5	5
6	15
7	14
8	8
9	12
10	2

	A
11	8
12	13
13	12
14	13
15	6
16	1
17	9
18	4
19	9
20	15

Student Response

1. 20 since the rows are numbered and the last value is next to 20.
2. A4, A12, and A14.
3. No response required.

Activity Synthesis

The goal is to make sure that students know how to type data into a spreadsheet and to locate values in the spreadsheet by row and column. The locations will be referred to with spreadsheet functions in upcoming activities. Here are some questions for discussion.

- "What value is in cell A7?" (14)
- "What was interesting or challenging about this activity?" (I never knew that you could describe each cell in a spreadsheet using the row and column labels.)

Math Community

After the *Warm-up*, display the Math Community Chart with the "Doing Math" actions added to the teacher section for all to see. Give students 1 minute to review. Then share 2–3 key points from the teacher section and your reasoning for adding them. For example,

- If "questioning vs. telling," a shared reason could focus on your belief that students are capable mathematical thinkers and your desire to understand how students are making meaning of the mathematics.
- If "listening," a shared reason could be that sometimes you want to sit quietly with a group just to listen and hear student thinking and not because you think the group needs help or is off-track.

After sharing, tell students that they will have the opportunity to suggest additions to the teacher section during the *Cool-down*.

9.2 Finding Spreadsheet Statistics

15 mins

Activity Narrative

The mathematical purpose of this activity is to calculate statistics, create data displays, and to investigate how those change when values are added or removed from the data set. Monitor for students discussing the relationship between outliers and the measure of center.

Standards

Addressing	**HSS-ID.A.1, HSS-ID.A.2**
Building Towards	**HSS-ID.A.3**

Launch

Keep students in the same groups. They will continue working using the spreadsheet they started in the previous activity.

Tell students that **statistics** are values that are calculated from data, such as the mean, median, or interquartile range.

Tell students that after they change the value in A1 to change the mean in the first set of questions, they should continue to use the changed values for the second set of questions rather than reset them to the values from the warm-up.

Note that GeoGebra is like any other computer program. It needs directions written in a specific way for it to execute a command. For example, if students forget to type the "=" symbol or don't capitalize "Sum," the formula won't work. Ask students to pause after typing the formulas and to ensure that cells D1, D2, and D3 display numbers for each statistic. If not, ask students to delete the contents of the cell and retype the formula, ensuring that they start with an "=" symbol and capitalize "Sum," "Mean," and "Median." If students will be using statistical technology other than GeoGebra for this activity, prepare alternate instructions.

Access for English Language Learners

Action and Expression: Internalize Executive Functions. To support development of organizational skills in problem-solving, chunk this task into more manageable parts. For example, present one set of questions at a time.
Supports accessibility for: Organization, Attention

Student Task Statement

Using the data from the *Warm-up*, we can calculate a few **statistics** and look at the data.

- Next to the word Sum, in cell D1, type =Sum(A1:A20)
- Next to the word Mean, in cell D2, type =Mean(A1:A20)
- Next to the word Median, in cell D3, type =Median(A1:A20)

1. What are the values for each of the statistics?
2. Change the value in A1 to 8. How does that change the statistics?
3. What value can be put into A1 to change the mean to 10.05 and the median to 9?

We can also use Geogebra to create data displays.

- Click on the letter A for the first column so that the entire column is highlighted.
- Click on the button that looks like a histogram to get a new window labeled One Variable Analysis .
- Click Analyze to see a histogram of the data.

1. Click the button Σx to see many of the statistics.
   a. What does the value for n represent?
   b. What does the value for Σx represent?
   c. What other statistics do you recognize?
2. Adjust the slider next to the word Histogram. What changes?
3. Click on the button to the right of the slider to bring in another window with more options. Then, click the box next to Set Classes Manually and set the Width to 5. What does this do to the histogram?
4. Click the word Histogram, and look at a box plot and dot plot of the data. When looking at the box plot, notice there is an x on the right side of the box plot. This represents a data point that is considered an outlier. Click on the button to the right of the slider, and uncheck the box labeled Show Outliers to include this point in the box plot. What changes? Why might you want to show outliers? Why might you want to include or exclude outliers?

Student Response

1. The sum is 178, the mean is 8.9, the median is 8.5.
2. The sum is now 179, the mean is 8.95, the median is still 8.5
3. 30

1. a. The number of values being analyzed.
   b. The sum of the values being analyzed.
   c. Mean, minimum, Q1 (first quartile), median, Q3 (third quartile), and maximum.
2. A number shows up to the right of the slider and the number and size of the bars change.
3. It makes each bar have a width of 5.
4. The x becomes part of the box plot. Reasoning varies. Sample response: It might be helpful to show outliers if we are unsure of the extreme data values and want to see what might happen if those values are not included. It might be better to include extreme values in the data if we do not see any error in the data and should not throw out data. If any extreme values are in doubt, or if there is a reason to call attention to outliers, it might make sense to separate them from the box plot.

Activity Synthesis

The purpose of this discussion is for students to create data displays using technology and to analyze what happens to the displays and the statistics when changes are made to the data set. Here are some questions for discussion.

- What happened to the statistics when you changed the value for A1 to 8 in the spreadsheet?" (When it was changed to 8, the mean increased slightly but the median stayed the same.)
- "Why did the mean increase?" (The sum of the data increased but the number of numbers stayed the same, so the mean had to increase.)
- "Why did the median stay the same?" (Changing a 7 to an 8 in the data set did not change the middle numbers, 8 and 9, in the data set).
- "What did you notice when you changed the width of the classes for the histogram?" (This changed the intervals for each bar to a width of 5, and the data were resorted into those intervals.)

Select students who were previously identified as discussing the relationship between outliers and the measures of center. Ask, "what is the relationship between outliers and the measures of center?" (When outliers are present the median is the preferred measure of center because it is less affected by outliers than is the mean.)

9.3 Making Digital Displays

10 mins

Activity Narrative

The mathematical purpose of this activity is for students to create data displays and calculate statistics using technology. Students plot the survey data that they collected from a statistical question in a previous lesson.

If students are using GeoGebra or the Spreadsheet tool in Math Tools to create a histogram, there is an issue when the maximum value is on the boundary of the greatest interval. In this case, GeoGebra includes the maximum value with the previous interval rather than following the convention of creating a new bar for the next interval. Address this issue during the Launch.

Access for English Language Learners

This activity uses the *Collect and Display* math language routine to support students as they develop their mathematical language.

Standards

Addressing	**HSS-ID.A.1**
Building Towards	**HSS-ID.A.2**

Instructional Routines

- Analyze It
- MLR2: Collect and Display

Launch

Arrange students in groups of 2. Tell them that they will be using technology to create data displays and to calculate statistics for data that they collected from a survey question in a previous lesson.

If students are using GeoGebra or the Spreadsheet tool in Math Tools to create a histogram for the data set in the Student Task Statement, draw their attention to the last interval in the created histogram. Tell students that GeoGebra has an issue when the maximum value of the data set is on the boundary of an interval. In that case, it includes the maximum value in the count for the previous bar rather than creating a new bar for that value, which is the convention it uses for other data. Demonstrate using the sample data set to show that this happens with the value "12".. They should check the maximum value, and adjust (either by hand or by redrawing—or just mentally) their histogram based on this issue when using GeoGebra or the Spreadsheet tool from Math Tools. If students will be using statistical technology other than GeoGebra for this activity, prepare alternate instructions.

Use *Collect and Display* to create a shared reference that captures students' developing mathematical language. Collect the language that students use to talk about technology and data representations. Display words and phrases such as "data set," "skewed distribution," "median," or "the IQR on the box plot."

Access for English Language Learners

Engagement: Develop Effort and Persistence. Provide prompts, reminders, guides, rubrics, or checklists that focus on increasing the length of on-task orientation in the face of distractions. For example, create an exemplar dot plot highlighting specific features, such as the scale and labels, as well as calculations for the mean and median. *Supports accessibility for: Attention, Social-Emotional Functioning*

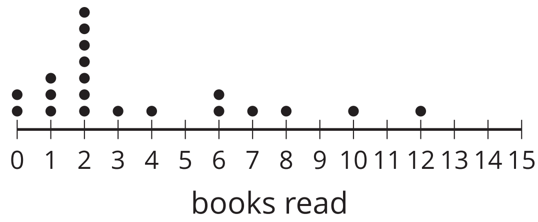
Student Task Statement

Use the data that you collected from the numerical, statistical question from a previous lesson. Use technology to create a dot plot, boxplot, and histogram for your data. Then find the mean, median, and interquartile range for the data.

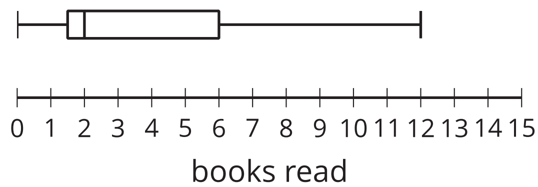
Student Response

Sample response:

- Question: On average, how many books did each student in the class read this summer?
- Data: 0, 0, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 4, 6, 6, 7, 8, 10, 12
- The mean is 3.65, the median is 2, and the IQR is 4.5.



•



•

•

Building on Student Thinking

Students may lose one data display when they begin to create the next one. Explain to students that it is important to copy their solutions into a more permanent place so they can refer to them later.

Are You Ready for More?

A stem and leaf plot is a table in which each data point is indicated by writing the first digit(s) on the left (the stem) and the last digit(s) on the right (the leaves). Each stem is written only once and shared by all data points with the same first digit(s). For example, the values 31, 32, and 45 might be represented like:

3	1 2
4	5

Key: 3 | 1 means 31

A class took an exam and earned the scores:

86	73	85	86	72	94	88	98	87
86	85	93	75	64	82	95	99	76

84 68

1. Use technology to create a stem and leaf plot for this data set.
2. How can we see the shape of the distribution from this plot?
3. What information can we see from a stem and leaf plot that we cannot see from a histogram?
4. What do we have more control of in a histogram than in a stem and leaf plot?

Extension Student Response

1. Key: 6|4 means 64

6	4 8
7	2 3 5 6
8	2 4 5 5 6 6 6 7 8
9	3 4 5 8 9

2. Sample response: The length of each leaf is like the height of a bar in a histogram except horizontal instead of vertical, so we may still be able to see peaks and tails.
3. We can see the exact values of each data point.
4. A histogram gives much greater control over the lengths of the intervals used.

Activity Synthesis

The goal of this activity was for students to create graphs and find statistics using technology. Here are some questions for discussion.

- “What were some challenges that you faced using technology, and how did you overcome them?” (I was not sure what buttons to press to get to the spreadsheet. I checked with my partner and figured it out.)
- “What width did you use for your histogram? Why?” (I used 5 because my data set has values ranging from 1 to 42. I could have used 10 but then I would have only had 5 bars.)
- “What is the appropriate measure of center for your data set?” (The median was appropriate because my data set has a skewed distribution.)
- “Which display allows you to calculate the IQR the most easily?” (The box plot because it displays Q1 and Q3.)
- “Can you find the median using your histogram?” (No, the data is grouped into intervals, so a histogram cannot be used to find the middle value for the median.)

Lesson Synthesis

The goal of this lesson is for students to display and investigate data using technology. Here are some questions for discussion.

- “How do you create data displays using technology?” (You type the data into the spreadsheet and then click the appropriate buttons.)
- “What are some advantages of using technology to display data and calculate statistics?” (You can easily switch between different data displays and you can change the intervals on histograms without having to sort through the data again. The advantage of having the technology calculate the statistics is that I can see how the statistics

change as I enter or make changes to the data.)

- "When do you think it is appropriate to use technology to display data or to calculate statistics?" (Graphing technology makes it easier to determine the shape of a distribution. I might use it to determine the most appropriate measure of center for a data set. Using technology to calculate statistics makes sense in most situations because statistics are calculated using algorithms that can get complicated when there are many values in the data set. The chance of making a mistake while calculating statistics by hand makes using technology a good choice.)

9.4 What Are These Values?

Cool-down

5 mins

Standards

Addressing **HSS-ID.A.1**

Launch

Math Community

Before distributing the *Cool-downs*, display the Math Community Chart and the community building question "What additions would you make to the teacher 'Doing Math' section of the Math Community Chart?" Ask students to respond to the question after completing the *Cool-down* on the same sheet.

After collecting the *Cool-downs*, identify themes from the community building question. Use them to add to or revise the teacher "Doing Math" section of the Math Community Chart before Exercise 4.

Student Task Statement

Here are some statistics given by a spreadsheet program:

- n: 100
- Mean: 51.68
- σ: 29.2957
- s: 29.4433
- Σx: 5168
- Σx^2: 352906
- Min: 1
- Q1: 23
- Median: 51
- Q3: 77
- Max: 105

1. What are the mean and median for the data?
2. How many values are in the data set?
3. What is the interquartile range for the data? Explain or show your reasoning.

Student Response

1. mean: 51.68, median: 51
2. 100
3. 54 since $77 - 23 = 54$

Responding To Student Thinking

Points to Emphasize
If students struggle with calculating an IQR, revisit how to do so in the activity referred to here. For example, launch the activity with a brief review of IQR and display student work from this *Cool-down* to highlight how to "see" the IQR in a box plot

Algebra 1, Unit 1, Lesson 11, Activity 3 Visual Variability and Statistics

Lesson 9 Summary

Data displays (like histograms or box plots) are very useful for quickly understanding a large amount of information, but often take a long time to construct accurately using pencil and paper. Technology can help create these displays as well as calculate useful *statistics* much faster than doing the same tasks by hand. Especially with very large data sets (in some experiments, millions of pieces of data are collected), technology is essential for putting the information into forms that are more easily understood.

A **statistic** is a quantity that is calculated from sample data as a measure of a distribution. *Mean* and *median* are examples of statistics that are measures of center. *Mean absolute deviation (MAD)* and *interquartile range (IQR)* are examples of statistics that are measures of variability. Although the interpretation must still be done by people, using the tools available can improve the accuracy and speed of doing computations and creating graphs.

Glossary

- statistic

Practice Problems

1 Student Task Statement

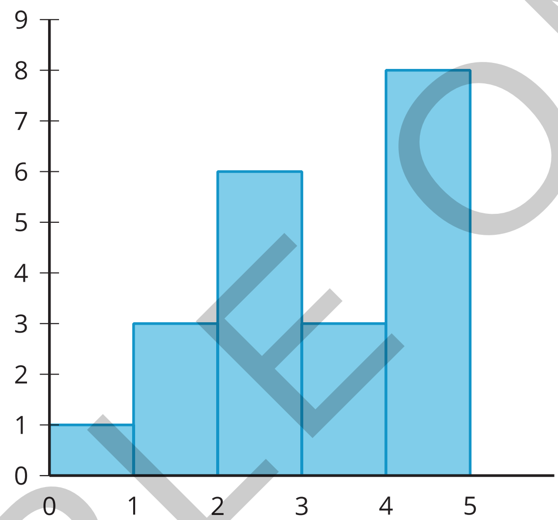
Technology required. The data represent the average customer ratings for several items sold online.

0.5 1 1.2 1.3 2.1 2.1 2.1 2.3 2.5 2.6 3.5 3.6 3.7 4 4.1 4.1 4.2 4.2

4.5 4.7 4.8

a. Use technology to create a histogram for the data with intervals 0–1, 1–2, and so on.
b. Describe the shape of the distribution.
c. Which interval has the highest frequency?

Solution



a.
b. The shape is skewed left with a center near 3.
c. 4 to 5

2 Student Task Statement

Technology required. The data represent the amount of corn, in bushels per acre, harvested from different locations.

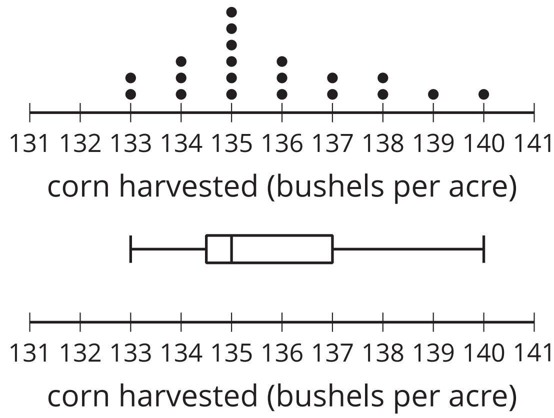
133 133 134 134 134 135 135 135 135 135 135 136 136 136 137 137

138 138 139 140

a. Use technology to create a dot plot and a box plot.
b. What is the shape of the distribution?
c. Compare the information displayed by the dot plot and box plot.

Solution

a.



b. The distribution is skewed right with a center near 135 bushels per acre.

c. Sample response: The box plot displays the median and quartiles and shows that the data are skewed right. Both graphs display the minimum and maximum. The dot plot shows all the data values but no measures of center directly. The dot plot displays the shape of the data more precisely than the box plot.

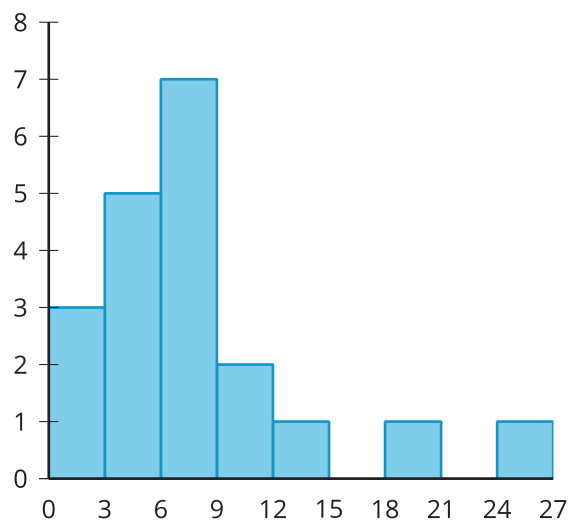
3 from Unit 1, Lesson 4

Student Task Statement

a. Describe the shape of the distribution.

b. How many values are represented by the histogram?

c. Write a statistical question of interest that could have produced the data set summarized in the histogram.



Solution

a. Skewed right with a center near 9.

b. 20

c. Sample response: What is the typical length, in feet, of a marine mammal?

4 from Unit 1, Lesson 3

Student Task Statement

The dot plot represents the distribution of satisfaction ratings for a landscaping company on a scale of 1 to 10. Twenty-five customers were surveyed.

On average, what was the satisfaction rating of the landscaping company?

Solution

Sample response: 8 if using the median or mode. 7.2 if using the mean.

Unit 1, Lesson 10

The Effect of Extremes

Goals

- Recognize the relationship between mean and median based on the shape of the distribution.
- Understand the effects of extreme values on measures of center.

Learning Targets

- I can describe how an extreme value will affect the mean and median.
- I can use the shape of a distribution to compare the mean and median.

Lesson Narrative

Students use dot plots to investigate the relationship between the shape of a distribution and the mean and median.

They begin by using aspects of mathematical modeling (MP4) to select appropriate variables to compare. Then, students make use of structure (MP7) and appropriate tools (MP5) to construct dot plots of data that have prescribed measures of center.

One of the activities in this lesson works best when each student has access to technology that will easily compute measures of center to produce dot plots or histograms because it will help students focus on understanding the relationship between extreme values and the measure of center without distracting with lengthy computations.

Standards

Building On	**HSS-ID.A.1**
Addressing	**HSS-ID.A.1, HSS-ID.A.2, HSS-ID.A.3**
Building Towards	**HSS-ID.A.3**

Instructional Routines

- Analyze It
- Aspects of Mathematical Modeling
- MLR5: Co-Craft Questions
- MLR8: Discussion Supports

Required Materials

Materials To Gather

- Internet-enabled device: Activity 2
- Statistical technology: Activity 2

Student Facing Learning Goals

Let's see how statistics change with the data.

10.1 At Least I Did Well

Warm-up

5 mins

Activity Narrative

This *Warm-up* prompts students to think about what variables they may use to analyze a situation. Then, students describe data displays they may use to compare two sets of data. Choosing variables and planning a process for comparing data sets engage students in aspects of mathematical modeling (MP4).

Listen for groups that choose a variable other than the number of wins to determine the top players in the game and for groups that select different data displays or ways of comparing data sets, and ask them to share with the whole group.

Standards

Addressing	**HSS-ID.A.2**
Building Towards	**HSS-ID.A.3**

Instructional Routines

- Aspects of Mathematical Modeling

Launch

Arrange students in groups of 2.

Ask students, "Have you ever played a game where you did really well, but still lost?" Consider looking up or sharing your own stories of great individual performances that still resulted in a loss. Tell students, "There are often many ways to measure performance and compare results."

Tell students to think quietly about their answers to the questions for about 1 minute before discussing with their partner and then sharing with the whole group.

Student Task Statement

Several online video games match players with other players at random to compete in a team game.

1. What information could you use to determine the top players in a team game like this? Explain your reasoning.
2. There are two games of a similar type. One person claims that the best players play game A. Another person claims that game B has better players. How could you display data to help inform their discussion? Explain your reasoning.

Student Response

Sample responses:

1. The number of points obtained by an individual player over several games could help determine the top players. Although wins might also be considered, good players will generally earn more points over several games even if they lose when matched with worse teammates.
2. I would create dot plots of the number of points gained by the top 100 players in each game. Then I would compare the center and variability of each distribution.

Activity Synthesis

Select previously identified students to share their solutions. If it does not come up in discussion, ask students how they might interpret a situation in which a small group of players has a significantly greater number of wins or points than the rest of the group. It might mean that most of the players are not very good and there are a few who are, or it might mean that there are a few dominant players who are much better than average players.

10.2 Separated by Skew

20 mins

Activity Narrative

There is a digital version of this activity.

The mathematical purpose of this activity is to help students understand how measures of center for distributions with different shapes are impacted by changes in the data. Students will create a dot plot, then describe the shape of the distribution and find measures of center. They will investigate how the measures of center change when the data set changes. Students create and investigate a data set from a given set of parameters including the shape of the distribution. Monitor for students using the correct terminology to describe the shape of the distribution.

This activity works best when each student has access to technology that computes measures of center and displays dot plots or histograms easily because students will benefit from seeing the relationship in a dynamic way. If students don't have individual access, projecting the distributions would be helpful during the *Launch*.

In the digital version of the activity, students use an applet to do much of the visualization and calculation of statistics. The applet allows students to focus on the impact of the values in the data set rather than creating visualizations or doing calculations themselves. Use the digital version if it is available so that students can dynamically see the impact of different values on various statistics.

This is the first time Math Language Routine 5: *Co-Craft Questions* is suggested in this course. In this routine, students are given a context or situation, often in the form of a problem stem (for example, a story, image, video, or graph) with or without numerical values. Students develop mathematical questions that can be asked about the situation. A typical prompt is, "What mathematical questions could you ask about this situation?" The purpose of this routine is to allow students to make sense of a context before feeling pressure to produce answers, and to develop students' awareness of the language used in mathematics problems.

Access for English Language Learners

This activity uses the *Co-Craft Questions* math language routine to advance reading and writing as students make sense of a context and practice generating mathematical questions.

Standards

Addressing **HSS-ID.A.1, HSS-ID.A.3**

Instructional Routines

- Analyze It
- MLR5: Co-Craft Questions

Launch

Tell students to close their other books or devices.

Sec D

Arrange students in groups of 2. Use *Co-Craft Questions* to give students an opportunity to familiarize themselves with the context, and to practice producing the language of mathematical questions.

- Display only the data, without revealing the question.
  Ask students, "What mathematical questions could you ask about this situation?"
- Give students 1–2 minutes to write a list of mathematical questions that could be asked about the situation before their comparing questions with a partner.
  As pairs discuss, support students in using conversation skills to generate and refine their questions collaboratively by seeking clarity, referring to students' written notes, and revoicing oral responses, as necessary. Listen for how students use language about measures of center and variability.
- Invite several groups to share one question with the class, and record responses. Ask the class to make comparisons among the shared questions and their own. Ask, "What do these questions have in common? How are they different?" Listen for and amplify questions that focus on center and spread.
- Reveal the questions, and give students a couple of minutes to compare them to their own question and those of their classmates. Identify similarities and differences.
- Consider providing these prompts:
  - "Which of your questions is most similar to or different from the ones provided? Why?"
  - "Is there a main mathematical concept that is present in both your questions and the ones provided? If so, describe it."
  - "How do your questions relate to one of the lesson goals of measuring corresponding points to determine congruence?"
- Invite students to choose one question to answer (from the class or from the curriculum), and then have students move on to the *Activity Synthesis*.

Provide access to devices that can run GeoGebra or other statistical technology.

After students have completed the problem that asks them to add values to the original data, give each group one of these distribution descriptions:

1. Uniform distribution with data between 4 and 12
2. Skewed right with most of the values at 10
3. Skewed left with most of the values at 10
4. Symmetric with most of the values at 4 and 16

Remind students of the words used to describe shapes of distributions: "symmetric," "skewed," "bell-shaped," "uniform," and "bimodal."

Student Task Statement

1. Use technology to create a dot plot that represents the distribution of the data, then describe the shape of the distribution.

   6 7 8 8 9 9 9 10 10 10 10 11 11 11 12 12 13 14

2. Find the mean and median of the data.
3. Find the mean and median of the data with 2 additional values included as described.
   a. Add 2 values to the original data set that are greater than 14.
   b. Add 2 values to the original data set that are less than 6.

c. Add 1 value that is greater than 14 and 1 value that is less than 6 to the original data set.

d. Add the two values, 50 and 100, to the original data set.

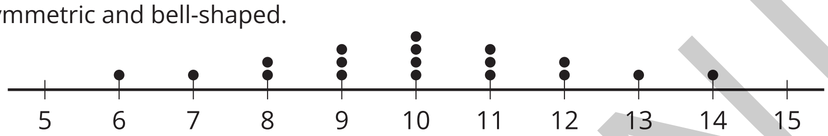
4. Share your work with your group. What do you notice is happening with the mean and median based on the additional values?

5. Change the values so that the distribution fits the description given to you by your teacher, then find the mean and median.

6. Find another group that created a distribution with a different description. Explain your work and listen to their explanation, then compare your measures of center.

Student Response

1. The distribution is symmetric and bell-shaped.



2. Mean: 10. Median: 10.
3. Correct answers will have these properties:
   a. A mean greater than 10. Median: 10.
   b. A mean less than 10. Median: 10.
   c. The mean will depend on the values chosen. Median: 10.
   d. Mean: 16.5. Median: 10.
4. Sample response: The median remains 10 in each situation, but the mean changes based on the values.
5. Sample responses:
   a. Uniform distribution: 4, 4, 5, 5, 6, 6, 7, 7, 8, 8, 9, 9, 10, 10, 11, 11, 12, 12 Mean: 8. Median: 8.
   b. Skewed right: 9, 9, 9, 10, 10, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 14, 15, 16 Mean: 11.39. Median: 11.
   c. Skewed left: 11 ,11, 11, 10, 10, 10, 10, 10, 9, 9, 9, 8, 8, 7, 7, 6, 5, 4 Mean: 8.61. Median: 9.
   d. Symmetric: 4, 4, 4, 4, 4, 6, 6, 8, 10, 10, 12, 14, 14, 16, 16, 16, 16, 16 Mean: 10. Median: 10.
6. Sample response: When the distribution is symmetric or uniform, the mean and median are the same values. The mean seems to be more influenced by the extreme values in the tail than the median is.

Building on Student Thinking

Students may have difficulty using technology to create dot plots so you may need to demonstrate how to use the technology. Students may confuse mean and median. Ask them to refer to previous work in which they calculated each measure of center. After students input the additional values as directed, they may use the wrong n when calculating the new mean. Remind students to complete detailed calculations.

Activity Synthesis

The goal is to make sure that students understand that the median is the preferred measure of center when a distribution is skewed or if there are extreme values, and that the mean is the preferred measure of center when a distribution is symmetric and there are no extreme values. Here are some questions for discussion.

- "What do you notice and wonder about the mean and median for each of these distributions?" (I noticed that sometimes the median did not change and the mean did. I wondered what would happen if I added a value of 1,000 to the data set.)

- “For which distributions does it look like the mean best represents what is typical in the data?” (The symmetric distributions.)
- “When is the median a better statistic to describe typical values?” (The skewed distributions.)
- “Why is the median a better statistic for skewed distributions?” (When you add extreme values to a data set, they tend to have a greater effect on the mean than on the median.)

10.3 Plots Matching Measures

10 mins

Activity Narrative

In this activity students recognize the relationship between measures of center and the shape of the distribution by creating and describing distributions with given measures of center. Listen for students using the terms "symmetric," "uniform," and "skewed." By using technology to create a dot plot with a given mean and median, students are using their understanding of the structure of the distribution (MP7) to adjust individual data values to change the measures of center. Making a spreadsheet available gives students an opportunity to choose appropriate tools strategically (MP5).

Standards

Addressing **HSS-ID.A.1, HSS-ID.A.3**

Instructional Routines

- Analyze It
- MLR8: Discussion Supports

Sec D

Launch

Keep students in their groups.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support organization, provide students with grid or graph paper to organize their work with the 3 dot plots.
Supports accessibility for: Language, Organization

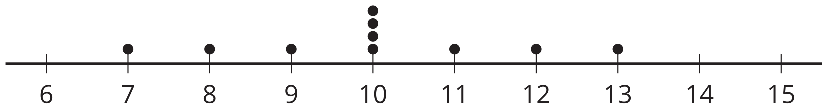
Student Task Statement

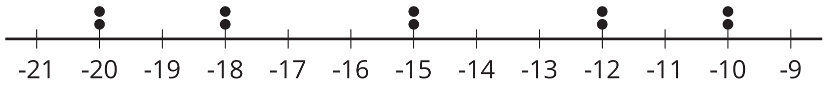
Create a possible dot plot with at least 10 values for each of the conditions listed. Each dot plot must have at least 3 values that are different.

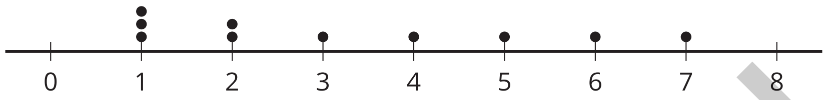
1. a distribution that has both mean and median of 10
2. a distribution that has both mean and median of -15
3. a distribution that has a median of 2.5 and a mean greater than the median
4. a distribution that has a median of 5 and a median greater than the mean

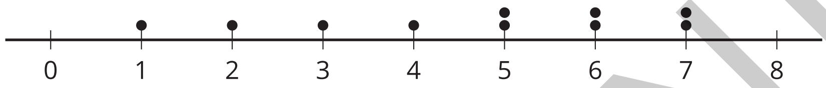
Student Response

Sample responses:

1. 

2. 

3. 

4. 

Are You Ready for More?

The mean and the median are by far the most common measures of center for numerical data. There are other measures of center, though, that are sometimes used. For each measure of center, list some possible advantages and disadvantages. Be sure to consider how it is affected by extremes.

1. *Interquartile mean*: The mean of only those points between the first quartile and the third quartile.
2. *Midhinge*: The mean of the first quartile and the third quartile.
3. *Midrange*: The mean of the minimum and maximum value.
4. *Trimean:* The mean of the first quartile, the median, the median again, and the third quartile. So we are averaging four numbers because the median is counted twice.

Extension Student Response

Sample responses:

1. An advantage is that, like the median, it is not affected much by extremes. A disadvantage is that we first need to trim the data and then compute a mean.
2. An advantage is that, like the median, it is not affected much by extremes. This number may not be that different from the median, though, which has a more direct, easy-to-understand meaning.
3. An advantage is that this can be quickly and easily computed. A disadvantage is that it would be greatly affected by extremes.
4. Like the median, this is not affected much by extremes, and it takes into consideration a little more of what is happening in the data than just the median. A disadvantage is that it is a little more difficult to compute and explain than is the median.

Activity Synthesis

The purpose of this discussion is for students to understand why the median is the preferred measure of center when a distribution is skewed or if there are extreme values, and the mean is the preferred measure of center when a distribution is symmetric and there are no extreme values.

For each description, select 2–3 groups to share their dot plots.

Here are some questions for discussion.

- "For the first and second dot plot, what do the distribution shapes have in common? Why do we choose the mean as the more appropriate measure of center?" (Symmetric. The mean of a set of data gives equal importance to each value to find the center, so it is a preferred measure of center when it accurately represents typical values for data.)
- "What do the shapes of the dot plots have in common when the mean is greater than the median?" (Skewed right.)
- "What information does the shape of the skewed distributions tell you about the median and mean?" (When distributions are skewed right, they will likely have a mean that is greater than the median because the values to the right disproportionately affect the mean. When distributions are skewed left, they will likely have a mean that is less than the median because the values to the left disproportionately affect the mean.)

Access for English Language Learners

Speaking: MLR 8 Discussion Supports. As groups share their dot plots with the whole class, revoice student ideas for determining the appropriate measure of center based on the shape of the distribution. Be sure to amplify mathematical uses of language by restating a statement as a question in order to clarify, apply appropriate language, and involve more students. For example, "Can someone else explain why the median is used for skewed data?" Ask for details in students' explanations by requesting students to elaborate on an idea or give an example from their data representation. This will help students define sequences with equations and answer questions about the context.
Design Principle(s): Support Sense-Making

Sec D

Lesson Synthesis

Here are some questions to draw out the relationship between measures of center and the shape of the distribution.

- "Why is the median preferred to the mean for a skewed data distribution?" (The values way to the right (or left) in skewed data have a greater effect on the mean, so the median is preferred to better reflect the typical values.)
- "When an extreme value is present, why is the median preferred to the mean?" (Extreme values have a greater effect on the mean than on the median, so the median is preferred.)
- "When the data distribution is symmetric or approximately symmetric, why is the mean preferred to the median?" (The mean takes into account every data value, so it is the preferred measure when it is representative of what is typical for the data.)

10.4 Shape and Statistics

Cool-down

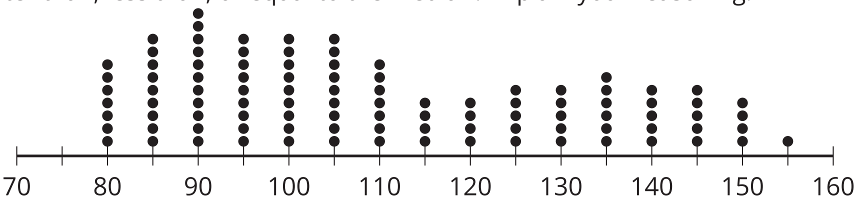
5 mins

Standards

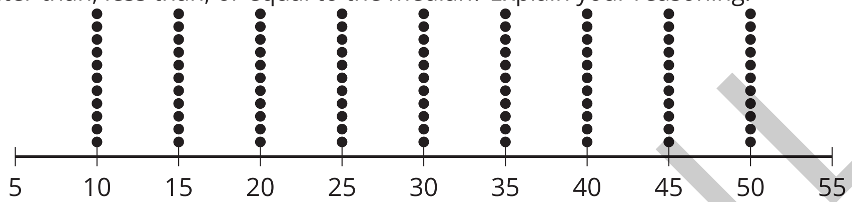
Building On	**HSS-ID.A.1**
Addressing	**HSS-ID.A.3**

Student Task Statement

1. Is the mean greater than, less than, or equal to the median? Explain your reasoning.



2. Is the mean greater than, less than, or equal to the median? Explain your reasoning.



Student Response

1. Sample response: The mean is greater than the median because the larger values to the right make the mean higher than it would be if the distribution were uniform.
2. Sample response. The mean is equal to the median because the data is symmetric.

Responding To Student Thinking

Points to Emphasize
If students struggle to recognize how mean and median are affected by extreme values, reinforce the idea as opportunities arise. For example, the matching activity referred to here provides many examples to highlight whether the mean or the median is greater in a skewed distribution and why.

Algebra 1, Unit 1, Lesson 11, Activity 2 Card Sort: Describing Data Distributions

Lesson 10 Summary

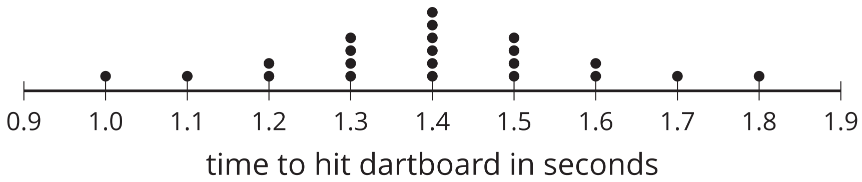
Is it better to use the mean or median to describe the center of a data set?

The mean gives equal importance to each value when finding the center. The mean usually represents the typical values well when the data have a symmetric distribution. On the other hand, the mean can be greatly affected by changes to even a single value.

The median tells you the middle value in the data set, so changes to a single value usually do not affect the median much. So, the median is more appropriate for data that are not very symmetrically distributed.

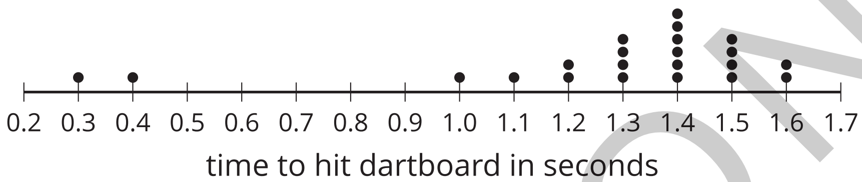
We can look at the distribution of a data set and draw conclusions about the mean and the median.

Here is a dot plot showing the amount of time a dart takes to hit a target in seconds. The data produce a symmetric distribution.



When a distribution is symmetric, the median and mean are both found in the middle of the distribution. Since the median is the middle value (or the mean of the two middle values) of a data set, you can use the symmetry around the center of a symmetric distribution to find it easily. For the mean, you need to know that the sum of the distances away from the mean of the values greater than the mean is equal to the sum of the distances away from the mean of the values less than the mean. Using the symmetry of the symmetric distribution you can see that there are four values 0.1 second above the mean, two values 0.2 seconds above the mean, one value 0.3 seconds above the mean, and one value 0.4 seconds above the mean. Likewise, you can see that there are the same number of values the same distances below the mean.

Here is a dot plot using the same data, but with two of the values changed, resulting in a skewed distribution.



When you have a skewed distribution, the distribution is not symmetric, so you are not able to use the symmetry to find the median and the mean. The median is still 1.4 seconds since it is still the middle value. The mean, on the other hand, is now about 1.273 seconds. The mean is less than the median because the lower values (0.3 and 0.4) result in a smaller value for the mean.

The median is usually more resistant to extreme values than is the mean. For this reason, the median is the preferred measure of center when a distribution is skewed or if there are extreme values. When using the median, you would also use the IQR as the preferred measure of variability. In a more symmetric distribution, the mean is the preferred measure of center, and the MAD is the preferred measure of variability.

Practice Problems

1 Student Task Statement

Select **all** the distribution shapes for which it is most often appropriate to use the mean.

A. bell-shaped

B. bimodal

C. skewed

D. symmetric

E. uniform

Solution

A, D, E

2 Student Task Statement

For which distribution shape is it usually appropriate to use the median when summarizing the data?

A. bell-shaped

B. skewed

C. symmetric

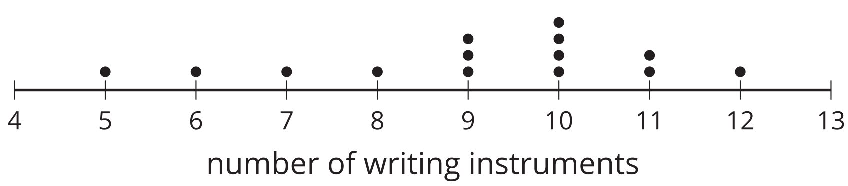
D. uniform

Solution

B

3 Student Task Statement

The number of writing instruments in some teachers' desks is displayed in the dot plot. Which is greater, the mean or the median? Explain your reasoning, using the shape of the distribution.



Solution

The median is greater than the mean. Sample reasoning: Since the distribution is skewed left, the mean will be less than the median.

4 from Unit 1, Lesson 9

Student Task Statement

A student has these scores on ten assignments. The teacher is considering dropping a lowest score. What effect does eliminating the lowest value, 0, from the data set have on the mean and median?

0 40 60 70 75 80 85 95 95 100

Solution

The mean increases from 70 to approximately 77.78. The median increases from 77.5 to 80.

5 from Unit 1, Lesson 9

Student Task Statement

a. What is the five-number summary for the data?

2 2 4 4 5 5 6 7 9 15

b. When the maximum, 15, is removed from the data set, what is the five-number summary?

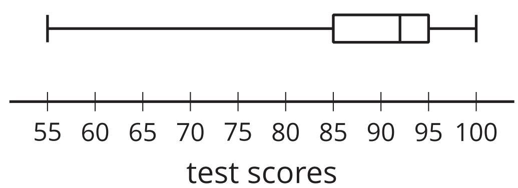
Solution

a. 2, 4, 5, 7, 15 (Minimum, Q1, Median, Q3, Maximum)
b. 2, 3, 5, 6.5, 9 (Minimum, Q1, Median, Q3, Maximum)

6 from Unit 1, Lesson 4

Student Task Statement

The box plot summarizes the test scores for 100 students:



Which term best describes the shape of the distribution?

Sec D

A. bell-shaped

B. uniform

C. skewed

D. symmetric

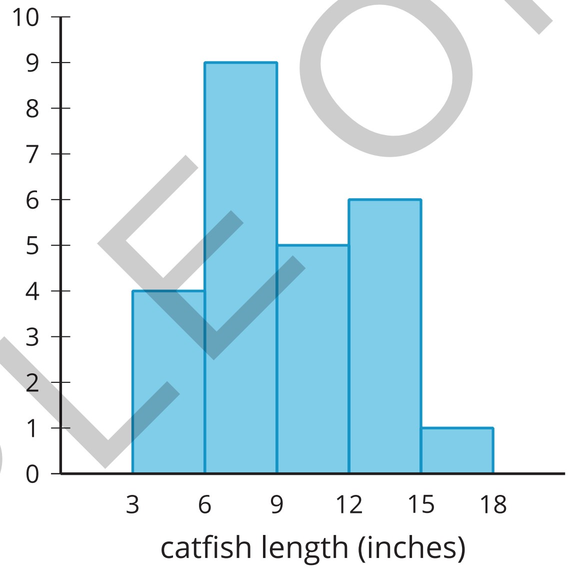
Solution

C

7 from Unit 1, Lesson 2

Student Task Statement

The histogram represents the distribution of lengths, in inches, of 25 catfish caught in a lake.



a. If possible, find the mean. If not possible, explain why not.

b. If possible, find the median. If not possible, explain why not.

c. Were any of the fish caught 12 inches long?

d. Were any of the fish caught 19 inches long?

Solution

a. Sample response: It is not possible to find the mean, because the exact values are not shown in a histogram.

b. Sample response: It is not possible to find the median, but you do know what interval the middle number is in. It is in the 6 to 9 inch interval, because the median is the thirteenth value in the data set.

c. Sample response: There is no way to tell whether a fish caught was 12 inches long using the histogram. You do know some fish were in the 12 to 15 inch interval, so it is possible.

d. No

Comparing and Contrasting Data Distributions

Goals

- Interpret (orally) a data set with greater MAD or IQR as having greater variability.

Learning Targets

- I can arrange data sets in order of variability, given graphic representations.

Lesson Narrative

Students make connections between different data displays and measures of center and measures of variability. In particular, they interpret data with greater MADs or IQRs as having greater variability. When students match data displays with descriptions of shape, measures of center, or measures of variability they make sense of the data display and the corresponding description or measure (MP2). When students participate in a *Math Talk* about finding the mean, they have an opportunity to notice and make use of the symmetric structure (MP7) of the values to determine the mean. Additionally, students need to be precise in their language (MP6). In the *Take Turns* routine students trade roles explaining their thinking and listening, providing opportunities to explain their reasoning and to critique the reasoning of others (MP3).

Math Community

Today, students use sticky notes to document actions in the "Doing Math" sections of the Math Community Chart that they see or hear throughout the lesson. During the *Lesson Synthesis*, students share what they noticed, and then they suggest additions for the chart as part of the *Cool-down*. The work today continues to build a foundation for developing math community norms in a later exercise and is the start of students identifying strengths in the actions of their peers.

Standards

Building On	HSS-ID.A.1
Addressing	HSS-ID.A.2
Building Towards	HSS-ID.A.3

Instructional Routines

- Card Sort
- Math Talk
- MLR2: Collect and Display
- MLR8: Discussion Supports
- Take Turns

Required Materials

Materials To Gather

- Math Community Chart: Lesson
- Math Community Chart: Activity 1
- Sticky notes: Activity 1

Materials To Copy

- Describing Data Distributions Cards (1 copy for every 2 students): Activity 2

Student Facing Learning Goals

Let's investigate variability using data displays and summary statistics.

11.1 Math Talk: Mean

Warm-up

10 mins

Activity Narrative

This *Math Talk* focuses on calculating the mean of a set of numbers. It encourages students to think about symmetry and to rely on the structure of the distribution to mentally solve problems. The strategies elicited here will be helpful later in the lesson when students examine and describe distributions.

To mentally calculate the mean, students need to look for and make use of structure (MP7).

This is the first *Math Talk* activity in the course. See the *Launch* for extended instructions for facilitating this activity successfully.

Standards

Addressing **HSS-ID.A.2**

Instructional Routines

- Math Talk
- MLR8: Discussion Supports

Launch

This is the first time students do the *Math Talk* instructional routine, so it is important to explain how it works before starting.

Explain the *Math Talk* routine: One problem is displayed at a time. For each problem, students are given a few minutes to quietly think and give a signal when they have an answer and a strategy. The teacher selects students to share different strategies for each problem, and might ask questions like "Who thought about it a different way?" The teacher records students' explanations for all to see. Students might be asked to provide more details about why they decided to approach a problem a certain way. It may not be possible to share every possible strategy for the given limited time, so the teacher may gather only two or three distinctive strategies per problem.

Consider establishing a small, discreet hand signal that students can display to indicate that they have an answer they can support with reasoning. This signal could be a thumbs-up, a certain number of fingers that tells the number of responses they have, or another subtle signal. This is a quick way to see if the students have had enough time to think about the problem. It also keeps students from being distracted or rushed by hands being raised around the class.

Display one problem at a time. Give students quiet think time for each problem, and ask them to give a signal when they have an answer and a strategy. Keep all problems displayed throughout the talk. Follow with a whole-class discussion.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support working memory, provide students with sticky notes or mini whiteboards.
Supports accessibility for: Memory, Organization

Student Task Statement

Evaluate the mean of each data set mentally.

- 27 30 33
- 61 71 81 91 101
- 0 100 100 100 100
- 0 5 7 12

Student Response

- 30. Sample reasoning: 27 is 3 below 30, 33 is 3 above 30, and 30 is right at 30, so 30 is the average because it is right in the center of all 3 numbers.
- 81. Sample reasoning: 61 and 101 balance each other out by being 20 away on each side of 81. Similarly, 71 and 91 balance each other out by being 10 away on either side.
- 80. Sample reasoning: The numbers add up to 400, and $400 \div 5 = 80$.
- 6. Sample reasoning: 0 and 12 balance each other out by being 6 away on each side of 6. 5 and 7 also balance each other out by being 1 away on each side of 6.

Building on Student Thinking

If students struggle to use symmetry as a method for finding the mean, consider asking them to find the mean for the values: 1, 2, 3, 4, 5.

Activity Synthesis

Ask students to share their strategies for each problem. Record and display their responses for all to see. To involve more students in the conversation, consider asking:

- "Who can restate ___'s reasoning in a different way?"
- "Did anyone have the same strategy but would explain it differently?"
- "Did anyone solve the problem in a different way?"
- "Does anyone want to add on to __'s strategy?"
- "Do you agree or disagree? Why?"

Although all correct methods for solving for the mean are valid, highlight the use of symmetry in the data. In previous lessons, students learned that symmetric distributions have a mean in the center of the data. When symmetry is present, it can be used to quickly discover the mean.

Math Community

After the *Warm-up*, display the revised Math Community Chart created from student responses in Exercise 3. Tell students that today they are going to monitor for two things:

- "Doing Math" actions from the chart that they see or hear happening.
- "Doing Math" actions that they see or hear that they think should be added to the chart.

Provide sticky notes for students to record what they see and hear during the lesson.

Access for English Language Learners

MLR8 Discussion Supports. Display sentence frames to support students when they explain their strategy. For example, “First, I _____ because . . . .” or “I noticed _____ so I . . . .” Some students may benefit from the opportunity to rehearse what they will say with a partner before they share with the whole class.
Advances: Speaking, Representing

11.2 Card Sort: Describing Data Distributions

25 mins

Activity Narrative

Students sort different representations of data during this activity. A sorting task gives students opportunities to analyze representations, statements, and structures closely and make connections (MP2, MP7).

Access for English Language Learners

This activity uses the *Collect and Display* math language routine to support students who are developing their mathematical language.

Standards

Building On	**HSS-ID.A.1**
Addressing	**HSS-ID.A.2**
Building Towards	**HSS-ID.A.3**

Instructional Routines

- Card Sort
- MLR2: Collect and Display
- Take Turns

Launch

Arrange students in groups of 2, and distribute the pre-cut cards. Allow students to familiarize themselves with the representations on the cards:

- Give students 1 minute to sort the cards into categories of their choosing.
- Pause the class after students have sorted the cards.
- Select groups to share their categories and how they sorted their cards.
- Discuss as many different types of categories as time allows.

Use *Collect and Display* to create a shared reference that captures students’ developing mathematical language. Collect the language that students use to describe categories and distributions. Display words and phrases such as “mean," "median," and "variability.”

After a brief discussion, invite students to complete the remaining questions.

Access for Students with Disabilities

Engagement: Internalize Self-Regulation. Provide students an opportunity to self-assess and reflect on their own progress. For example, “Which matches were easier to make?” and “Which matches were more difficult to explain?”
Supports accessibility for: Social-Emotional Functioning, Conceptual Processing

Student Task Statement

1. Your teacher will give you a set of cards. Take turns with your partner to match a data display with a written statement.
    a. For each match that you find, explain to your partner how you know it’s a match.
    b. For each match that your partner finds, listen carefully to their explanation. If you disagree, discuss your thinking and work to reach an agreement.
2. After matching, determine if the mean or median is more appropriate for describing the center of the data set based on the distribution shape. Discuss your reasoning with your partner. If it is not given, calculate (if possible) or estimate the appropriate measure of center. Be prepared to explain your reasoning.

Student Response

1. A. I
    B. L
    C. N
    D. M
    E. J
    F. K
    G. H
2. A. Row 1: mean, symmetric, 6
    B. Row 2: mean, symmetric, -14
    C. Row 3: mean, symmetric, -14
    D. Row 4: median, skewed, 11
    E. Row 5: median, skewed, 14.5
    F. Row 6: mean, symmetric, -14
    G. Row 7: median, skewed, 20

Activity Synthesis

Much discussion takes place between partners.

Direct students' attention to the reference created using *Collect and Display*. Ask students to share how they found their matches. Invite students to borrow language from the display as needed, and update the reference to include additional phrases as they respond. (For example, “The distribution is skewed right because there is a longer tail on that side.”)

After all groups have completed the matching, discuss the following:

- “Which matches were tricky? Explain why.” (The box plot labeled F was tricky because I had to use a process of elimination to figure out that it was the one that was uniform.)

- "Did you need to make adjustments in your matches? What might have caused an error? What adjustments were made?" (Yes. I realized that I thought incorrectly that skewed left meant that most of the data were on the left. However, I learned that skewed left means that there are data to the left of where most of the data are located.)
- "Can you determine the median using only a histogram? Why or why not?" (No, but you can determine the interval that contains the median.)
- "Can you determine if a distribution is uniform from a box plot? Why or why not?" (No. You can determine that the data distribution could possibly be symmetric based on the distribution of the five-number summary, but beyond that you would not be able to know, using only a box plot, that the data is uniform.)

The purpose of the second part of the activity is to discuss the relationship between mean and median based on the shape of the distribution and to make the connection to measures of variability. Ask:

- "If the mean is the appropriate measure of center, should we use the MAD or the IQR to measure variability?" (MAD)
- "If the median is the appropriate measure of center, should we use the MAD or the IQR to measure variability?" (IQR)

11.3 Visual Variability and Statistics

10 mins

Activity Narrative

This activity prompts students to compare variability in several data sets by analyzing the distributions shown on box plots and dot plots. Some students may reason about variability by observing the shapes and features of the data displays. Others may try to quantify the variability by finding the IQR from each box plot, or by estimating the MAD from each dot plot. Look for students who approach the task quantitatively.

Access for English Language Learners

This activity uses the *Collect and Display* math language routine to support students who are developing their mathematical language.

Standards

Building On HSS-ID.A.1

Addressing HSS-ID.A.2

Instructional Routines

- MLR2: Collect and Display

Launch

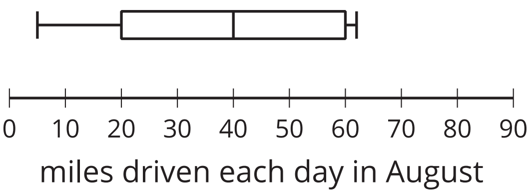
Arrange students in groups of 2. Give students five minutes to work through the questions, and then pause for a whole-group discussion.

Use *Collect and Display* to direct attention to words collected and displayed from an earlier activity. Invite students to borrow language from the display as needed, and update it throughout the lesson.

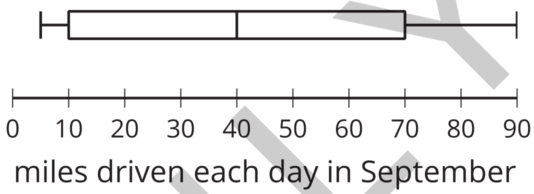
Student Task Statement

1. Each box plot summarizes the number of miles driven each day for 30 days in each month. The box plots represent, in order, the months of August, September, October, November, and December.
    a. The five box plots have the same median. Explain why the median is more appropriate for describing the center of the data set than the mean for these distributions.
    b. List the box plots in order of least variability to greatest variability. Check with another group to see if they agree.

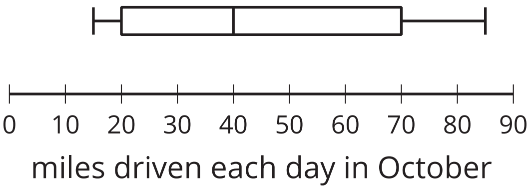
A



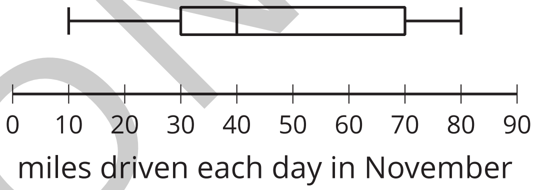
B



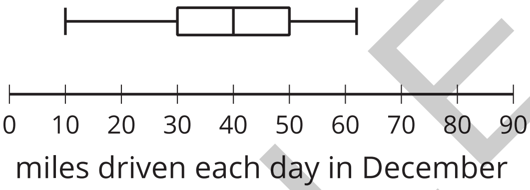
C



D

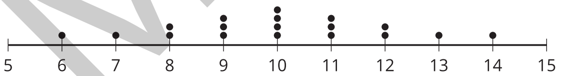


E

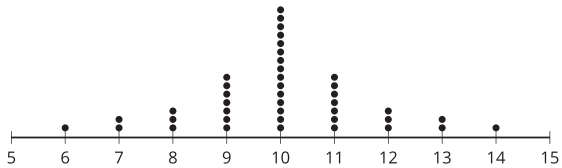


2. These five dot plots represent the number of books on different shelves of the same section of the library.
    a. The five dot plots have the same mean. Explain why the mean is more appropriate for describing the center of the data set than the median.
    b. List the dot plots in order of least variability to greatest variability. Check with another group to see if they agree.

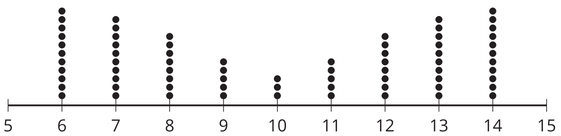
A



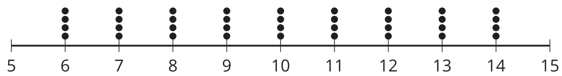
B



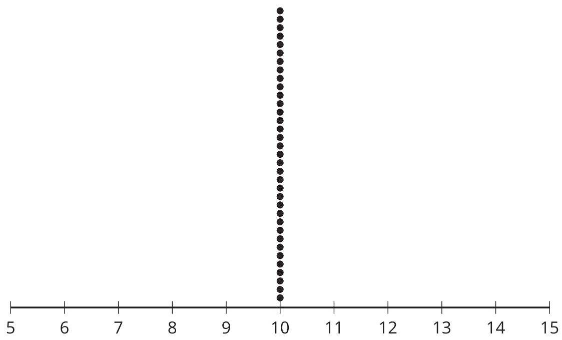
C



D



E



Student Response

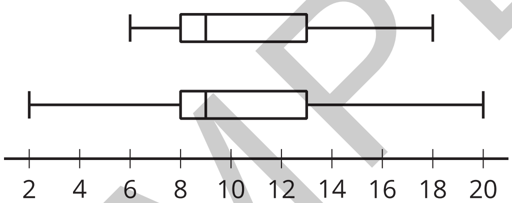
1. a. None of the distributions is symmetric.
   b. E, D and A, C, B, or E, A, D, C, B (if A and D are distinguished by range)
2. a. Each of the distributions is symmetric and does not include obvious outliers.
   b. E, B, A, D, C

Building on Student Thinking

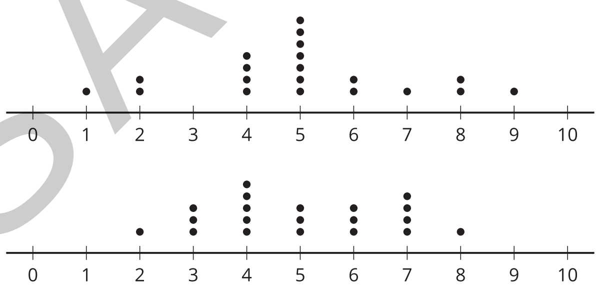
Students may have forgotten what variability means or which statistic to use to determine the variability in a data set. Refer them to previous work, or ask them what measure is useful in determining a data set's tendency to have different values.

Are You Ready for More?

1. These two box plots have the same median and the same IQR. How could we compare the variability of the two distributions?



2. These two dot plots have the same mean and the same MAD. How could we compare the variability of the two distributions?



Extension Student Response

Sample responses: We could look at the range and shape of the distribution to determine where the values are in comparison to one another.

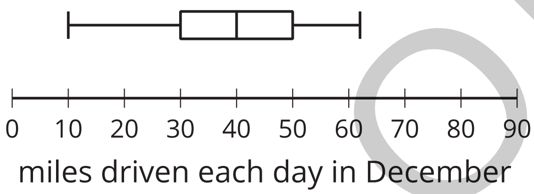
Activity Synthesis

The purpose of this discussion is to make the connection between the shape of the distribution and the use of either IQR or MAD to quantify variability. Another goal is to make sure students understand that a greater value for IQR or MAD means greater variability.

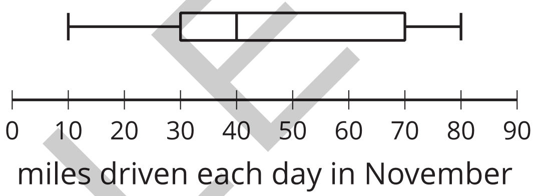
Direct students' attention to the reference created using *Collect and Display*. Ask students to share how they arranged the distributions. Invite students to borrow language from the display as needed, and update the reference to include additional phrases as they respond.

Display the box plots in order of variability with the IQR included, and then display the dots plots in order of variability with the MAD included.

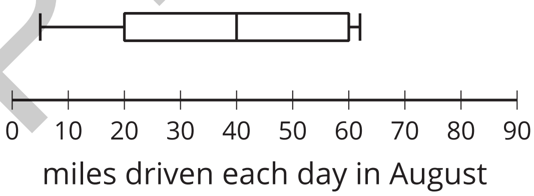
E. IQR: 20



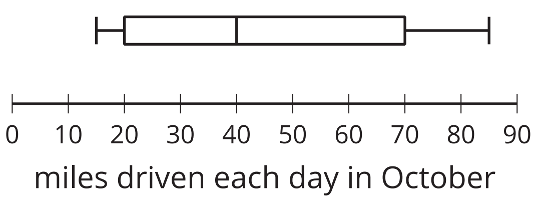
D. IQR: 40



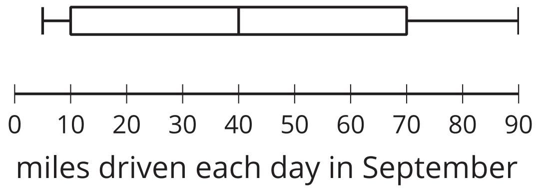
A. IQR: 40



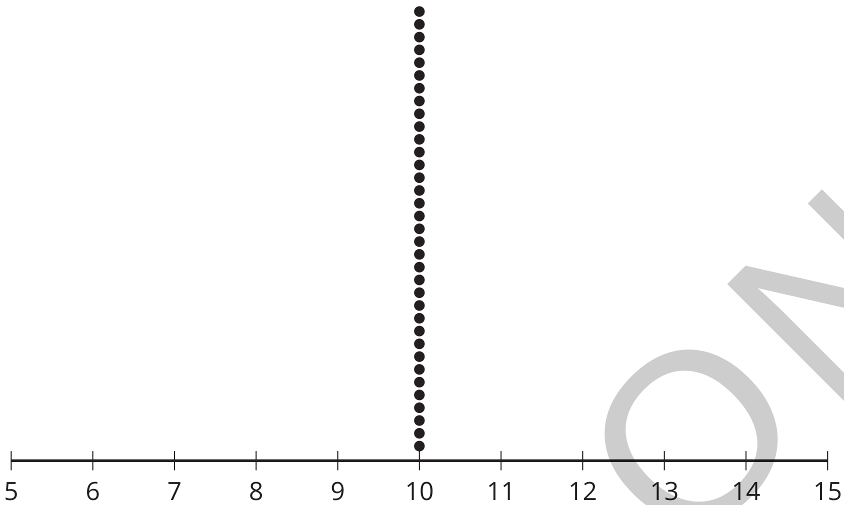
C. IQR: 50



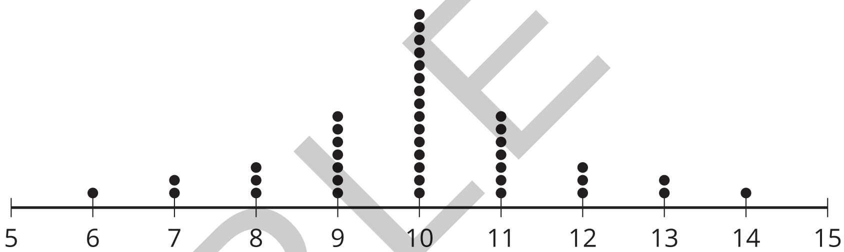
B. IQR: 60



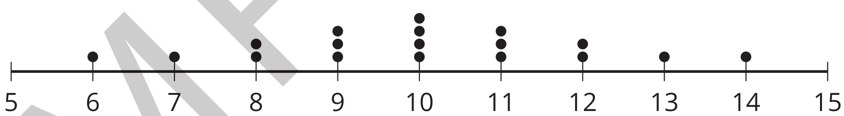
E. MAD: 0



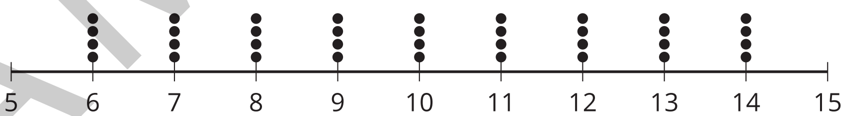
B. MAD: 1.12



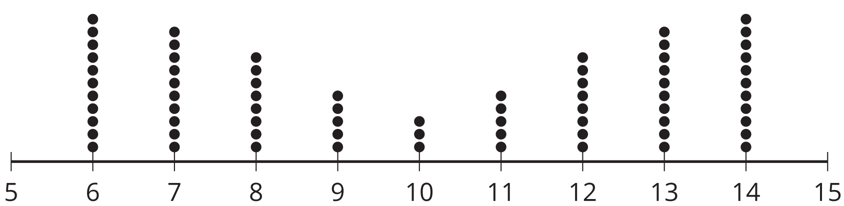
A. MAD: 1.56



D. MAD: 2.22



C. MAD: 2.68



The IQR for the data shown in box plots A through E are {40, 60, 50, 40, 20} and the MAD for the data shown in dot plots

A through E are approximately {1.56, 1.10, 2.68, 2.22, 0}. Here are some questions for discussion:

- “What are the IQR and MAD measuring?” (They are measuring the spread, or variability, of the data)
- “Which plots were the most difficult to arrange?” (The dot plots were more difficult because it was easy to find the IQR for the box plots.)
- “Does the order given when arranged by the IQR and MAD match your order?” (Yes, except for the box plots A and D which had the same IQR, and I didn’t know how to arrange them.)
- “What do you notice about the values for IQR and MAD?” (The values for the MAD were higher than I thought except for distribution J. I did not know that the MAD could be equal to zero.)
- “What advantages are offered by using IQR and MAD versus visual inspection?” (The IQR and MAD are values that can be easily sorted.)

If some students already arranged the plots using IQR or MAD, you should ask them, “Why did you choose to arrange the plots by IQR or MAD?” (I knew that IQR and MAD were measures of variability so I used them.)

Access for English Language Learners

Reading: MLR 2 Collect and Display. While students are working, circulate and listen to students talk about the connection between the shapes of the distribution and the use of either the IQR or MAD to quantify variability. After arranging the box plots and dot plots in order of variability, write down—directly on this visual display—common or important phrases that you hear students say about each representation. Capture words such as “greater variability,” “skewed,” or “spread of data.” This will help students read and use mathematical language during their paired and whole-class discussions.
Design Principle(s): Support Sense-Making

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support organization, provide students with a Venn diagram to compare how IQR and MAD are the same and different.
Supports accessibility for: Language, Organization

Lesson Synthesis

In this lesson, students investigated variability using data displays and summary statistics.

- “One data set’s measure of center is best represented by a median of 7 and another data set's by a median of 10. How would you determine which data set has greater variability?” (You calculate the IQR. Whichever one has a larger IQR is more variable.)
- “How do you determine which of two roughly symmetric distributions has less variability?” (You calculate the MAD. Whichever one has a lower MAD has less variability.)
- “What does it mean to say that one data set or distribution has more variability than another?” (The appropriate measure of variability for one data set is greater than the other. Using a data display, one distribution is more spread apart than the other.)

Math Community

Invite 2–3 students to share what “Doing Math” actions they noticed. Record and display their responses for all to see, such as by adding check marks to any already listed items or adding new items near the chart for the class to consider adding. Next, give students 1–2 minutes with a partner to discuss any changes or revisions they think the chart needs.

Tell students they can suggest revisions during the *Cool-down*.

11.4 Which Menu?

Cool-down

5 mins

Standards

Building On HSS-ID.A.1

Addressing HSS-ID.A.2

Launch

Math Community

Before distributing the *Cool-downs*, display the Math Community Chart and the community building question "What additions or revisions would you make to the Math Community Chart?" Ask students to respond to the question after completing the *Cool-down* on the same sheet.

After collecting the *Cool-downs*, identify themes from the community building question. Use them to add to or revise the Math Community Chart before Exercise 5.

Sec D

Student Task Statement

A restaurant owner believes that it is beneficial to have different menu items with a lot of variability so that people can have a choice of expensive and inexpensive food. Several chefs offer menus and suggested prices for the food they create. The owner creates dot plots for the prices of the menu items and finds some summary statistics. Which menu best matches what the restaurant is looking for? Explain your reasoning.

Italian:

mean: \$9.03

median: \$9

MAD: \$2.45

IQR: \$3.50

Diner:

mean: \$3.36

median: \$2

MAD: \$2.12

IQR: \$4

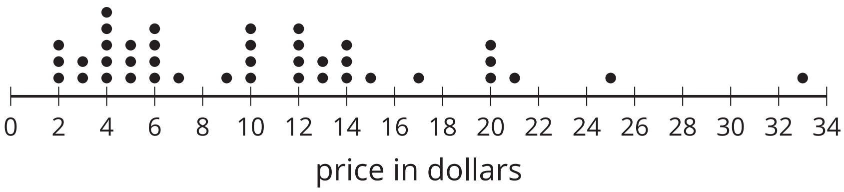
Japanese:

mean: $10.35

median: $10

MAD: $5.55

IQR: $9.50



Steakhouse:

mean: $11.51

median: $10.50

MAD: $3.69

IQR: $4.50

Student Response

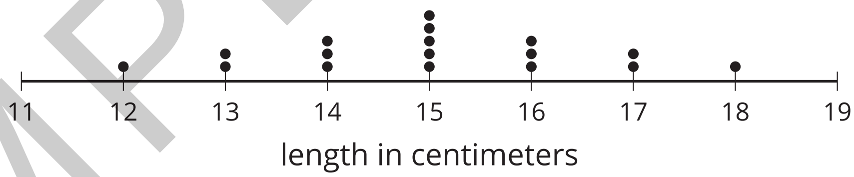
Japanese. The variability, whether measured with IQR or MAD, is greater than any of the other menus available.

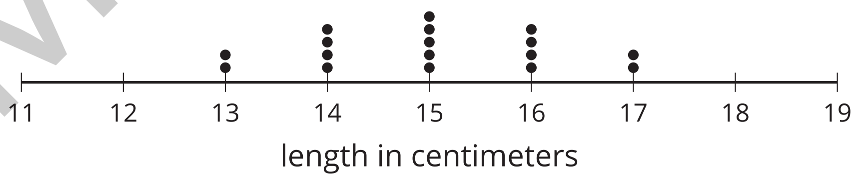
Responding To Student Thinking

More Chances
Students will have more opportunities to understand the mathematical ideas addressed here. There is no need to slow down or add additional work to the next lessons.

Lesson 11 Summary

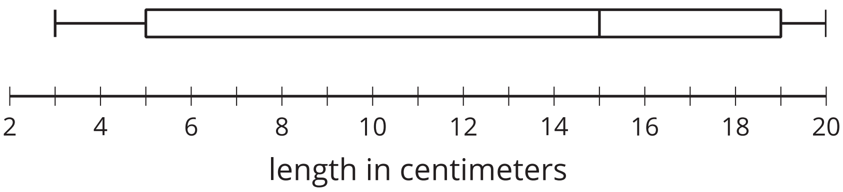
The mean absolute deviation, or MAD, is a measure of variability that is calculated by finding the mean distance from the mean of all the data points. Here are two dot plots, each with a mean of 15 centimeters, displaying the length of sea scallop shells in centimeters.

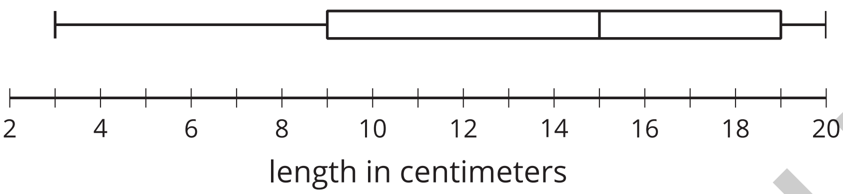




Notice that both dot plots show a symmetric distribution so the mean and the MAD are appropriate choices for describing center and variability. The data in the first dot plot appear to be more spread apart than the data in the second dot plot, so you can say that the first data set appears to have greater variability than does the second data set. This is confirmed by the MAD. The MAD of the first data set is 1.18 centimeters and the MAD of the second data set is approximately 0.94 cm. This means that the values in the first data set are, on average, about 1.18 cm away from the mean, and the values in the second data set are, on average, about 0.94 cm away from the mean. The greater the MAD of the data, the greater the variability of the data.

The interquartile range, IQR, is a measure of variability that is calculated by subtracting the value for the first quartile, Q1, from the value for the third quartile, Q3. These two box plots represent the distributions of the lengths in centimeters of a different group of sea scallop shells, each with a median of 15 centimeters.





Notice that neither of the box plots have a symmetric distribution. The median and the IQR are appropriate choices for describing center and variability for these data sets. The middle half of the data displayed in the first box plot appear to be more spread apart, or show greater variability, than the middle half of the data displayed in the second box plot. The IQR of the first distribution is 14 cm, and the IQR is 10 cm for the second data set. The IQR measures the difference between the median of the second half of the data, Q3, and the median of the first half, Q1, of the data, so it is not affected by the minimum or the maximum value in the data set. It is a measure of the spread of the middle 50% of the data.

The MAD is calculated using every value in the data set, and the IQR is calculated using only the values for Q1 and Q3.

Practice Problems

1 Student Task Statement

In science class, Clare and Lin estimate the mass of eight different objects that actually weigh 2,000 grams each. Some summary statistics:

Clare

- mean: 2,000 grams
- MAD: 275 grams
- median: 2,000 grams
- IQR: 500 grams

Lin

- mean: 2,000 grams
- MAD: 225 grams
- median: 1,950 grams
- IQR: 350 grams

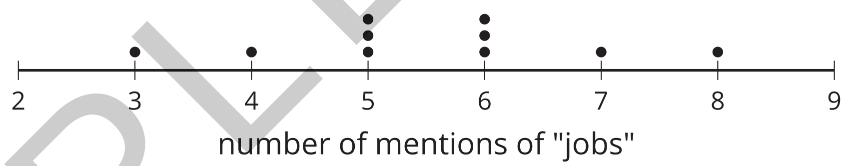
Which student was better at estimating the mass of the objects? Explain your reasoning.

Solution

Sample response: Lin was better at estimating than was Clare because although both students had measures of center very close to 2,000 grams, Lin's responses were less variable as determined by both the MAD and the IQR.

2 Student Task Statement

A reporter counts the number of times a politician talks about jobs in campaign speeches. What is the MAD of the data represented in the dot plot?



A. 1.1 mentions

B. 2 mentions

C. 2.5 mentions

D. 5.5 mentions

Solution

A

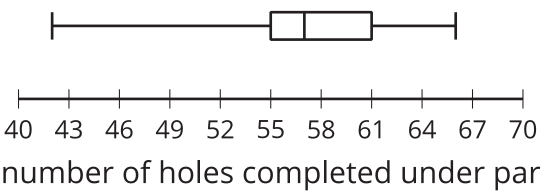
3 Student Task Statement

Four amateur miniature golfers attempt to finish 100 holes under par several times. After each round of

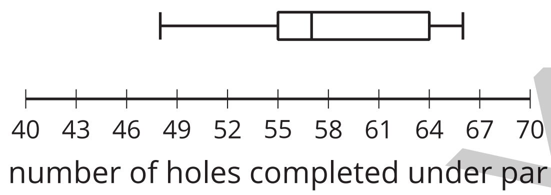
Sec D

100 holes, the number of holes they successfully complete under par is recorded. Due to the presence of extreme values, box plots were determined to be the best representation for the data. List the four box plots in order of variability from least to greatest.

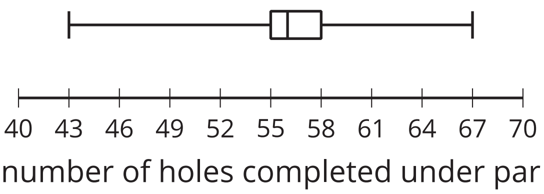
player a



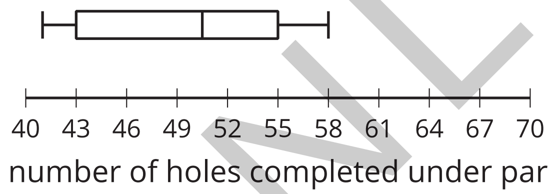
player b



player c



player d



Solution

c, a, b, d

Sec D

4 from Unit 1, Lesson 10

Student Task Statement

Select **all** the distribution shapes for which the median *could be* much less than the mean.

A. symmetric

B. bell-shaped

C. skewed left

D. skewed right

E. bimodal

Solution

D, E

5 from Unit 1, Lesson 9

Student Task Statement

a. What is the five-number summary for the data?

0 2 2 4 5 5 5 5 7 11

b. When the minimum, 0, is removed from the data set, what is the five-number summary?

Solution

a. 0, 2, 5, 5, 11 (Minimum, Q1, Median, Q3, Maximum)

b. 2, 3, 5, 6, 11 (Minimum, Q1, Median, Q3, Maximum)

6 from Unit 1, Lesson 9

Student Task Statement

What effect does eliminating the highest value, 180, from the data set have on the mean and median?

25 50 50 60 70 85 85 90 90 180

Solution

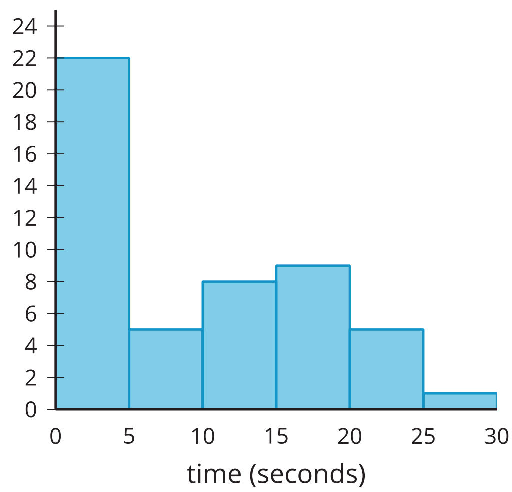
The mean decreases from 78.5 to approximately 67.22. The median decreases from 77.5 to 70.

Sec D

7 from Unit 1, Lesson 3

Student Task Statement

The histogram represents the distribution of the number of seconds it took for each of 50 students to find the answer to a trivia question using the internet. Which interval contains the median?



A. 0 to 5 seconds

B. 5 to 10 seconds

C. 10 to 15 seconds

D. 15 to 20 seconds

Solution

B

Unit 1, Lesson 12

Standard Deviation

Goals

- Comprehend (in spoken and written language) standard deviation as a measure of variability.
- Use technology to compute standard deviation.

Learning Targets

- I can describe standard deviation as a measure of variability.
- I can use technology to compute standard deviation.

Lesson Narrative

In this lesson, students are introduced to *standard deviation* as a measure of variation that is similar to the mean absolute deviation (MAD).

Standard deviation is a measure of variability that is similar to the mean absolute deviation. It is not important that students calculate the value by hand, although an extension is available for students who are curious.

When students manipulate data to achieve various specified measures of center or variability, they make use of the structure underlying standard deviation as a measure of variability (MP7).

Math Community

Today's math community building time has two goals. The first is for students to make a personal connection to the math actions chart and to share on their *Cool-down* the math action that is most important to them. The second is to introduce the idea that the math actions that students have identified will be used to create norms for their mathematical community in upcoming lessons.

Standards

Building On	**HSS-ID.A.1, HSS-ID.A.2**
Addressing	**HSS-ID.A.2, HSS-ID.A.3**

Instructional Routines

- 5 Practices
- MLR1: Stronger and Clearer Each Time
- Notice and Wonder

Required Materials

Materials To Gather

- Math Community Chart: Activity 1
- Statistical technology: Activity 2, Activity 3

Materials To Copy

- Algebra 1 Unit 1 Useful Terms and Displays (1 copy for every 30 students): Activity 3

Required Preparation

Activity 2:

Students should have access to technology for calculating standard deviation and mean.

Activity 3:

Students should have access to technology for calculating standard deviation and mean.

Student Facing Learning Goals

Let's learn about standard deviation, another measure of variability.

12.1 Notice and Wonder: Measuring Variability

5 mins

Warm-up

Activity Narrative

The purpose of this warm-up is to elicit the idea that calculating the **standard deviation** is very similar to calculating the MAD, which will be useful when students explore standard deviation in a later activity. While students may notice and wonder many things about these dot plots, the similarities and differences between standard deviation and the MAD as measures of variability are the important discussion points.

Standards

Building On HSS-ID.A.1

Addressing HSS-ID.A.2

Instructional Routines

- Notice and Wonder

Sec D

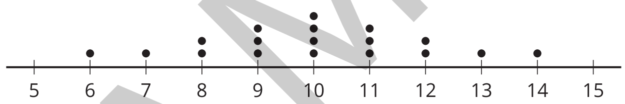
Launch

Display the dot plots and statistics for all to see. Ask students to think of at least one thing they notice and at least one thing they wonder. Give students 1 minute of quiet think time, and then 1 minute to discuss the things they notice with their partner. Follow that with a whole-class discussion.

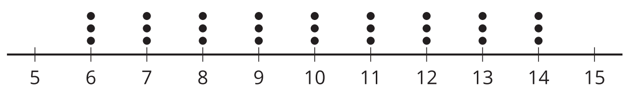
Student Task Statement

What do you notice? What do you wonder?

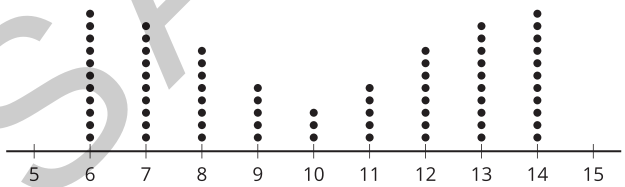
mean: 10, MAD: 1.56, standard deviation: 2



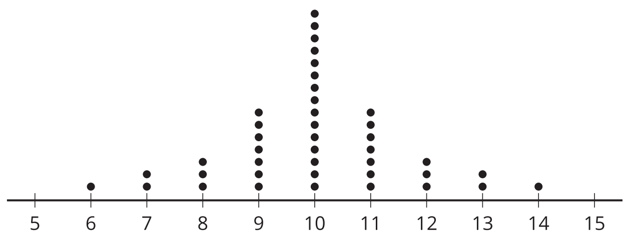
mean: 10, MAD: 2.22, standard deviation: 2.58



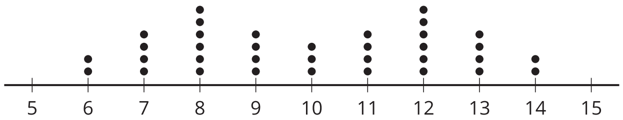
mean: 10, MAD: 2.68, standard deviation: 2.92



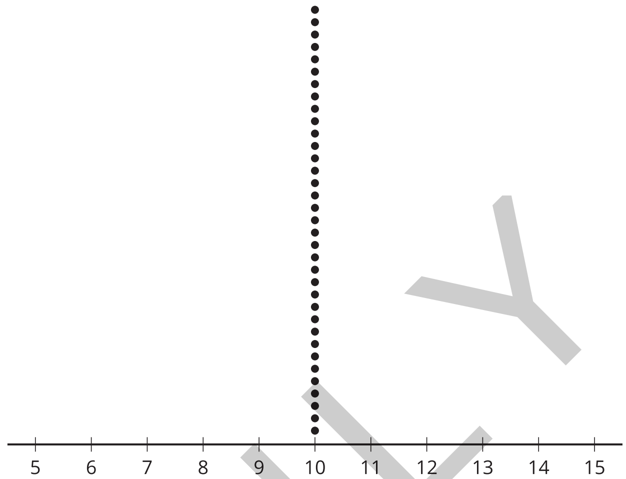
mean: 10, MAD: 1.12, standard deviation: 1.61



mean: 10, MAD: 2.06, standard deviation: 2.34



mean: 10, MAD: 0, standard deviation: 0



Student Response

Things students may notice:

- All of the sets have the same mean.
- The MAD and standard deviation were both large for the same dot plots and small for the same dot plots.
- The standard deviation is always greater than the MAD (except for the last plot where both are zero).

Things students may wonder:

- Are MAD and standard deviation both measures of variability?
- What is standard deviation?

Activity Synthesis

Ask students to share the things they noticed and wondered. Record and display their responses for all to see. If possible, record the relevant reasoning on or near the images. After all responses have been recorded without commentary or editing, ask students, “Is there anything on this list that you are wondering about now?” Encourage students to respectfully disagree, ask for clarification, or point out contradicting information. If the concept of variability does not come up during the conversation, ask students to discuss this idea.

Math Community

After the *Warm-up*, display the revisions to the class Math Community Chart that were made from student suggestions in an earlier exercise. Tell students that over the next few exercises, this chart will help the class decide on community norms—how they as a class hope to work and interact together over the year. To get ready for making those decisions, students are invited at the end of today’s lesson to share which “Doing Math” action on the chart is most important to them personally.

12.2 Investigating Standard Deviation

20 mins

Activity Narrative

The purpose of this activity is to let students investigate what happens to the standard deviation using different data sets. The goal is for students to make conjectures about what standard deviation measures and how relative size of the standard deviation can be estimated from the shape of the distribution. In particular, students should recognize that adding or subtracting the same value from each value in the data set will change the mean by the same amount, but the standard deviation remains unchanged. Multiplying or dividing each value in the data set by the same value scales both the mean and the standard deviation by the same value.

Monitor for students who:

- Check values without much thought until the mean or standard deviation is correct.
- Use a starting set of values, then modify a few of them up or down to match the statistics given.
- Use symmetry to get the mean correct.
- Adjust the data by multiplying all of the values by a number to change the spread of the distribution to get close to the standard deviation.

Plan to have students present in this order from least to more abstract ways of thinking about the data.

The routine of *Anticipate, Monitor, Select, Sequence, Connect (5 Practices)* requires a balance of planning and flexibility. The anticipated approaches might not surface in every class, and there may be reason to change the order in which strategies are presented. While monitoring, keep in mind the learning goal and adjust the order to ensure that all students have access to the first idea presented (whether that be a common misconception or a different approach).

Identify students who create data sets with different types of distributions for "10 different numbers that have a mean of 1" and "10 different numbers with a standard deviation as close to 2.5 as you can get in one minute."

Standards

Building On **HSS-ID.A.1**

Addressing **HSS-ID.A.2**

Instructional Routines

- 5 Practices

Launch

Demonstrate how to find standard deviation and mean using the technology available. In Algebra 1, students are dealing with the entire population, not sampling, so the population standard deviation is used.

Open the Spreadsheet & Statistics tool from Math Tools, or navigate to https://www.geogebra.org/classic/spreadsheet. Enter the values in column A. Select all of column A, and then choose the button that looks like a histogram and "One Variable Analysis." Click the button that says Σx. The population standard deviation is labeled σ. Additionally, the spreadsheet command =SD(A1:A10) will compute the population standard deviation for data in cells A1 through A10.

Arrange students in groups of 2. Give students time to work through the first two questions, followed by a whole-class discussion.

Select students with different strategies, such as those described in the *Activity Narrative*, to share later.

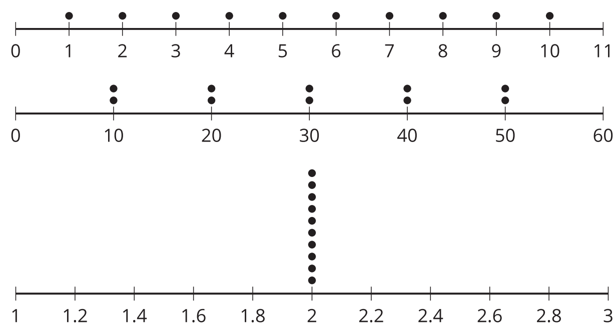
Student Task Statement

Use technology to find the mean and the standard deviation for the data in the dot plots.

1. What do you notice about the mean and standard deviation that you and your partner found for the three dot plots?
2. Invent some data that fit the conditions. Be prepared to share your data set and reasoning for choice of values.

Partner 1

Dot plots:

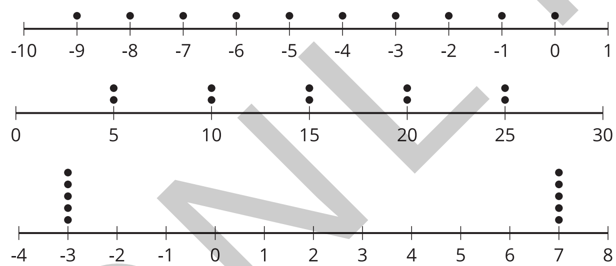


Conditions:

- 10 numbers with a standard deviation equal to the standard deviation of your first dot plot with a mean of 6.
- 10 numbers with a standard deviation three times greater than the data in the first row.
- 10 different numbers with a standard deviation as close to 2 as you can get in 1 minute.

Partner 2

Dot plots:



Conditions:

- 10 numbers with a standard deviation equal to the standard deviation of your first dot plot with a mean of 12.
- 10 numbers with a standard deviation four times greater than the data in the first row.
- 10 different numbers with a standard deviation as close to 2 as you can get in 1 minute.

Sec D

Student Response

1. Statistics for the dot plots.
   a. Partner 1:
      - mean: 5.5, standard deviation: 2.87
      - mean: 30, standard deviation: 14.14
      - mean: 2, standard deviation: 0
   b. Partner 2:
      - mean: -4.5, standard deviation: 2.87
      - mean: 15, standard deviation: 7.07
      - mean: 2, standard deviation: 5

   Sample response: For the first dot plot, the standard deviation is the same because the data values are distributed the same way around the mean. For the second dot plot, the standard deviation is greater for partner 1 because the data are spread further apart. For the third dot plot, the standard deviation is greater for partner 2 because there is some variation.

2. Sample responses for the conditions:
   a. Partner 1:
      - 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5. To keep the same standard deviation and change the mean, the data should shift up or down until it has the desired mean.
      - 3, 6, 9, 12, 15, 18, 21, 24, 27, 30. To triple the standard deviation, the data can be multiplied by 3.
      - 0.7, 1.4, 2.1, 2.8, 3.5, 4.2, 4.9, 5.6, 6.3, 7. Since the first data set had a standard deviation of around 2.87, dividing all the values by 2.87 would produce a data set with a standard deviation near 1. Multiplying those values by 2 would produce a data set with a standard deviation near 2.
   b. Partner 2:
      - 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, 13.5, 14.5, 15.5, 16.5. To keep the same standard deviation and change the mean, the data should shift up or down until it has the desired mean.
      - -36, -32, -28, -24, -20, -16, -12, -8, -4, 0. To quadruple the standard deviation, the data can be multiplied by 4.
      - -6.3, -5.6, -4.9, -4.2, -3.5, -2.8, -2.1, -1.4, -0.7, 0. Since the first data set had a standard deviation of around 2.87, dividing all the values by 2.87 would produce a data set with a standard deviation near 1. Multiplying those values by 2 would produce a data set with a standard deviation near 2.

Activity Synthesis

The purpose of this discussion is to understand that the standard deviation is a measure of variability related to the mean of the data set. The discussion also provides an opportunity for students to discuss what they notice and wonder about the mean and standard deviation.

Invite previously selected students to share their distributions and methods for finding values that worked. Sequence the discussion of the strategies by the order listed in the *Activity Narrative*. If possible, record and display their work for all to see.

For at least one student's data set that had a standard deviation close to 2.5, ask them how they might adapt the data set so that it has a mean of 1. (Shift the data values up or down so that the set has a mean of 1.)

Connect the different responses to the learning goals by asking questions such as:

- "What do you think that the standard deviation measures? Why do you think that?" (I think that it measures variability because it behaved like MAD. When all the values were the same it was zero.)
- "Why is the standard deviation the same for {1,2,3,4,5} and {-2,-1,0,1,2}?" (For each data set: The values to the left of the mean are a distance of 2 and 1 from the mean and the values to the right of the mean are a distance 2 and 1 from the mean, and also, the middle value is 0 away from the mean. This results in the standard deviation being the same.)
- "Why is the standard deviation different for {-4, -2, 0, 2, 4} and {-4, -3, -2, -1, 0}?" (The values for the first data set are twice the distance from the mean as are the values in the second data set. That makes the standard deviation of the first set greater than the standard deviation of the second data set.)
- "When is the standard deviation equal to zero?" (It is zero when all the values are the same as each other or when there is no variability.)
- "Was your mean the same as your partner's mean in the fifth match?" (My mean was much different from my partners. Mine was 5, and my partner's was 10.)

- “How did using technology help or hinder your mathematical thinking about standard deviation and mean?” (It really helped me because I did not get bogged down with the calculations, and I could look for patterns in the data, the data displays, and the statistics. In particular, having to come up with my own data and then see what happened to the statistics was really helpful.)

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support organization, provide students with a graphic organizer with standard deviation at the center, connecting to it all the related concepts mentioned during the discussion.
Supports accessibility for: Language, Organization

12.3 Investigating Variability

10 mins

Activity Narrative

The purpose of this mathematical activity is to let students investigate how the standard deviation and other measures of variability change when you add, change, or remove values in a data set. Monitor for students mentioning the concepts of shape, variability, and center. This activity works best when each student has access to statistical technology because it would take too long to do otherwise. If students don't have individual access, projecting the statistical technology would be helpful during the *Launch*.

Standards

Building On	**HSS-ID.A.1, HSS-ID.A.2**
Addressing	**HSS-ID.A.3**

Instructional Routines

- MLR1: Stronger and Clearer Each Time

Launch

Open the Spreadsheet & Statistics tool from Math Tools, or navigate to https://www.geogebra.org/classic/spreadsheet. Demonstrate how to add, change, and remove data points from a data set using technology. Provide access to statistical technology. Give students time to work through the questions, and then pause for a brief whole-class discussion.

Access for English Language Learners

MLR1 Stronger and Clearer Each Time. Before the whole-class discussion, give students time to meet with 2–3 partners to share and get feedback on their first draft response to the question: “What appears to change more easily, the standard deviation or the interquartile range?” Invite listeners to ask questions and give feedback that will help their partner clarify and strengthen their ideas and writing. Give students 3–5 minutes to revise their first draft based on the feedback they receive.
Advances: Writing, Speaking, Listening

Access for Students with Disabilities

Engagement: Develop Effort and Persistence. Encourage and support opportunities for peer interactions. Invite students to talk about their ideas with a partner before writing them down. Display sentence frames to support students when they explain their ideas. For example, "I notice that the mean is affected when . . . ." or "Standard deviation will always be affected because . . . ."
Supports accessibility for: Language, Organization

Student Task Statement

Begin with the data:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

1. Use technology to find the mean, standard deviation, median, and interquartile range.
2. How do the standard deviation and mean change when you remove the greatest value from the data set? How do they change if you add a value to the data set that is twice the greatest value?
3. What do you predict will happen to the standard deviation and mean when you remove the least value from the data set? Check to see if your prediction was correct.
4. What happens to the standard deviation and mean when you add a value to the data set equal to the mean? Add a second value equal to the mean. What happens?
5. Add, change, and remove values from the data set to answer the question: What appears to change more easily, the standard deviation or the interquartile range? Explain your reasoning.

Student Response

1. Mean: 10.5, standard deviation: 5.77, median: 10.5, IQR: 10.
2. Sample response: Both the mean and the standard deviation decreased when the greatest value was removed. They both increased when the outlier was added.
3. Sample response: The mean increased and the standard deviation decreased when the least value was removed.
4. Sample response: The mean remained the same and the standard deviation decreased slightly.
5. Sample response: The standard deviation tends to change more easily. Interquartile range is calculated by the position of the values within the ordered data set rather than the actual values themselves, so adding an additional extremely large value has the same effect as adding another value equal to the maximum value. On the other hand, standard deviation uses the distance from each value to the mean, so it is more easily affected by the inclusion or removal of values.

Building on Student Thinking

Students who compute a different standard deviation may be using the sample standard deviation statistic. Tell these students to use the value for σ rather than s for computations in this unit.

Are You Ready for More?

How is the standard deviation calculated? We have seen that the standard deviation behaves a lot like the mean absolute deviation, and that is because the key idea behind both is the same.

1. Using the original data set, calculate the deviation of each point from the mean by subtracting the mean from each data point.
2. If we just tried to take a mean of those deviations what would we get?
3. There are two common ways to turn negative values into more-useful positive values: take the absolute value or square the value. To find the MAD we find the absolute value of each deviation, then find the mean of those numbers. To find the standard deviation we square each of the deviations, then find the mean of those numbers. Then finally take the square root of that mean. Compute the MAD and the standard deviation of the original data set.

Extension Student Response

1. {-9.5,-8.5,-7.5,-6.5,-5.5,-4.5,-3.5,-2.5,-1.5,-0.5,0.5,1.5,2.5,3.5,4.5,5.5,6.5,7.5,8.5,9.5}
2. 0
3. MAD: 5, SD: 5.77

Activity Synthesis

The purpose of this discussion is to talk about standard deviation as a measure of variability. The goal of this discussion is to make sure that students understand that the standard deviation behaves similarly to the MAD and that it is a measure of variability that uses the mean as a measure of center. Discuss how the standard deviation is affected by the addition and removal of outliers in the data set. The standard deviation decreases when outliers are removed because the data in the distribution then display less variability, and the standard deviation increases when outliers are added because the data in the distribution then display more variability.

Add standard deviation to the display of measures of center and measures of variability created in an earlier lesson. The blackline master provides an example of what this display may look like after all items are added.

The MAD already included in the example display is approximately 1.09, and the standard deviation is 1.194.

Here are some discussion questions.

- “What does the standard deviation measure? How do you know?” (It measures variability. Like the MAD, the standard deviation increases when outliers are included in the data set.)
- “The standard deviation is calculated using the mean. Do you think it is more appropriate to use with symmetric or skewed data sets?” (Symmetric, because it is calculated using the mean.)

Lesson Synthesis

Here are some questions for discussion.

- “How does standard deviation compare and contrast with MAD and IQR?” (Like the MAD and IQR, the standard deviation is a measure of variability. Unlike the IQR, standard deviation is based on the mean not the median. It is very similar to the MAD because it involves finding the difference between each value and the mean, but then you perform different operations with those differences.)
- “One data set has a standard deviation of 5 and another data set has a standard deviation of 10. What does this tell you about the distribution of each data set?” (Standard deviation is a measure of variability, so it tells you how spread apart the data are. The second data set shows greater variability than the first data set shows.)

12.4 True or False: Reasoning with Standard Deviation

Cool-down

5 mins

Standards

Addressing **HSS-ID.A.2**

Launch

Math Community

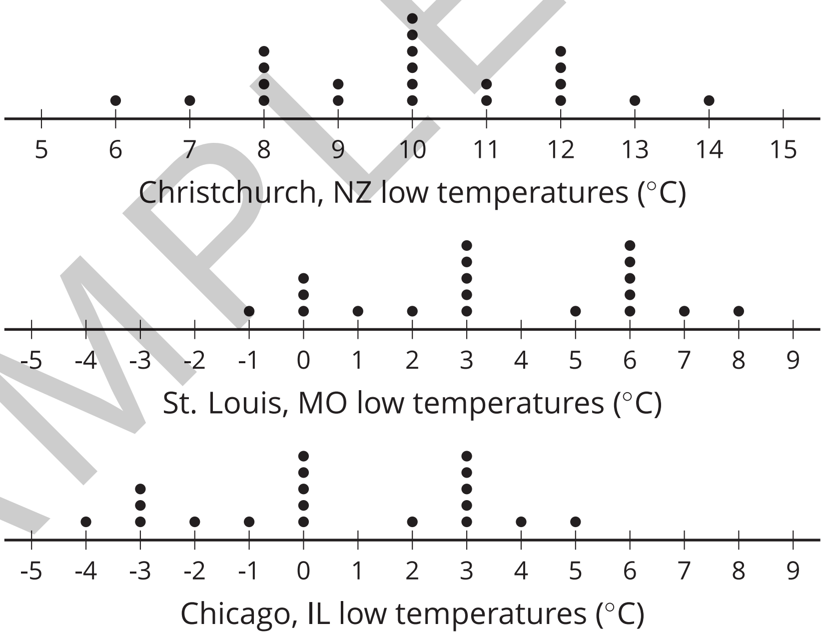
Before distributing the *Cool-downs*, display the Math Community Chart and the community building question "Which 'Doing Math' action is most important to you, and why?" Ask students to respond to the question after completing the *Cool-down* on the same sheet.

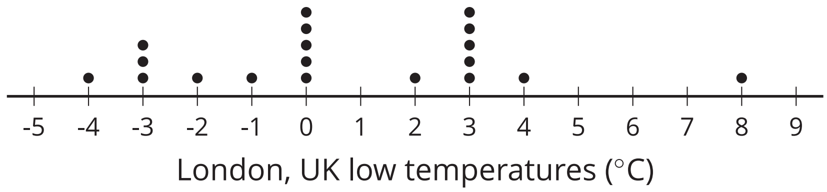
After collecting the *Cool-downs*, review student responses to the community building question. Use the responses to draft a student norm and a teacher norm to use as an example in Exercise 6. For example, if "sharing ideas" is a common choice for students, a possible norm is "We listen as others share their ideas."

For the teacher norms section, if "questioning vs. telling" from the "Doing Math" section is key for your teaching practice, then one way to express that as a norm is "Ask questions first to make sure I understand how someone is thinking."

Student Task Statement

The low temperature in degrees Celsius for some cities on the same days in March are recorded in the dot plots.





Decide if each statement is true or false. Explain your reasoning.

1. The standard deviation of Christchurch's temperatures is zero because the data distribution is symmetric.
2. The standard deviation of St. Louis's temperatures is equal to the standard deviation of Chicago's temperatures.
3. The standard deviation of Chicago's temperatures is less than the standard deviation of London's temperatures.

Student Response

1. False. Sample response: The standard deviation is a measure of variability and there is some variability in the data set.
2. True. Sample response: Chicago's distribution of temperatures is the same as St. Louis's, but 3 degrees cooler. The two cities have the same variability in temperature, and so they have the same standard deviation.
3. True. Sample response: London has the same low temperatures as does Chicago except on the hottest day, London's temperature is 3 degrees warmer than Chicago's. Therefore, the temperatures in London have more variability than Chicago's temperatures on these days.

Responding To Student Thinking

More Chances

Students will have more opportunities to understand the mathematical ideas addressed here. There is no need to slow down or add additional work to the next lessons.

Lesson 12 Summary

We can describe the variability of a distribution using the **standard deviation**. The standard deviation is a measure of variability that is calculated using a method that is similar to the one used to calculate the MAD, or mean absolute deviation.

A deeper understanding of the importance of standard deviation as a measure of variability will come with a deeper study of statistics. For now, know that the standard deviation is mathematically important and will be used as the appropriate measure of variability when the mean is an appropriate measure of center.

Like the MAD, the standard deviation is large when the data set is more spread out, and the standard deviation is small when the variability is small. The intuition you gained about MAD will also work for the standard deviation.

Glossary

- standard deviation

Practice Problems

1 Student Task Statement

The shoe size for all the pairs of shoes in a person's closet are recorded.

7 7 7 7 7 7 7 7 7 7

a. What is the mean?

b. What is the standard deviation?

Solution

a. 7

b. 0

2 Student Task Statement

Here is a data set: 1 2 3 3 4 4 4 4 5 5 6 7

a. What happens to the mean and standard deviation of the data set when the 7 is changed to a 70?

b. For the data set with the value of 70, why would the median be a better choice for the measure of center than the mean?

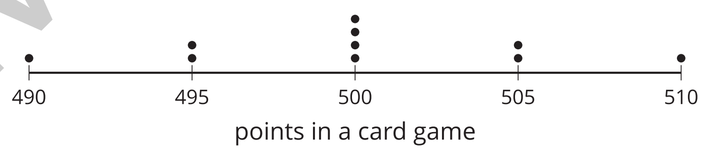
Solution

a. The mean increases from 4 to 9.25 and the standard deviation increases from about 1.58 to 18.36.

b. The median would be a better choice, because the data set with the 70 is a skewed distribution.

3 Student Task Statement

Which of these best estimates the standard deviation of points in a card game?



A. 5 points

B. 20 points

C. 50 points

D. 500 points

Solution

A

4 from Unit 1, Lesson 11

Student Task Statement

The mean of data set A is 43.5 and the MAD is 3.7. The mean of data set B is 12.8 and the MAD is 4.1.

a. Which data set shows greater variability? Explain your reasoning.

b. What differences would you expect to see when comparing the dot plots of the two data sets?

Solution

a. B shows greater variability because the MAD is greater than the MAD for A.

b. Sample response: A's dot plot will have most of the data centered around 43.5 with the data, on average, 3.7 units above or below 43.5. B's dot plot will have most of the data centered around 12.8 with the data, on average, 4.1 units above or below 12.8.

5 from Unit 1, Lesson 10

Student Task Statement

Select **all** the distribution shapes for which the mean and median *must be* about the same.

A. bell-shaped

B. bimodal

C. skewed

D. symmetric

E. uniform

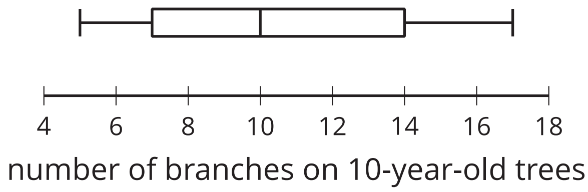
Solution

A, D, E

6 from Unit 1, Lesson 11

Student Task Statement

What is the IQR?



A. 5 branches

B. 7 branches

C. 10 branches

D. 12 branches

Solution

B

7 from Unit 1, Lesson 9

Student Task Statement

The data represent the number of cans collected by different classes for a service project.

12 14 22 14 18 23 42 13 9 19 22 14

a. Find the mean.

b. Find the median.

c. Eliminate the greatest value, 42, from the data set. Explain how the measures of center change.

Solution

a. 18.5 cans

b. 16 cans

c. The mean decreases to about 16.36 cans and the median decreases to 14 cans.

More Standard Deviation

Goals

- Describe (orally and in writing) the meaning of standard deviation in context.

Learning Targets

- I can use standard deviation to say something about a situation.

Lesson Narrative

One purpose of this lesson is to introduce students to the *Information Gap* routine. In addition to the routine, students work to deepen their understanding of standard deviation by interpreting it in various contexts. If time allows, the optional activity provides additional practice with interpreting dot plots.

Interpreting standard deviation in context is evidence that students are reasoning abstractly and quantitatively (MP2). The *Information Gap* activity may take students several rounds before students understand how to ask questions to gain the information they need to solve the problem (MP1). The information gap structure also allows them to refine the language that they use and to ask increasingly more precise questions until they get the information they need (MP6).

Math Community

In today's activities, students are introduced to the idea of math norms as expectations that help everyone in the room feel safe, comfortable, and productive doing math together. Students then consider what norms would connect and support the math actions the class recorded so far in the Math Community Chart.

Standards

Building On	**HSS-ID.A.1**
Addressing	**HSS-ID.A.2**
Building Towards	**HSS-ID.A.2**

Instructional Routines

- MLR4: Information Gap Cards
- MLR8: Discussion Supports

Required Materials

Materials To Gather

- Chart paper: Activity 1
- Math Community Chart: Activity 1

Materials To Copy

- African and Asian Elephants Cards (1 copy for every 2 students): Activity 2

Student Facing Learning Goals

Let's continue to interpret standard deviation.

13.1

What Do You Want to Know?

Warm-up

10 mins

Activity Narrative

The purpose of this *Warm-up* is to prepare students for the *Information Gap* activity that follows. First, students are given a problem with incomplete information. They are prompted to brainstorm what they need to know to solve a problem that involves the masses of elephants. Next, they practice asking for information, explaining the rationale for their request, and persevering if their initial questions are unproductive (MP1). Once students have enough information, they solve the problem.

Standards

Building Towards HSS-ID.A.2

Launch

Display the first paragraph of the activity statement for all to see. Ask students to solve the problem. When they recognize that not enough information is given, display the second prompt and ask what they need to know to be able to solve the problem. Display the sentence frame, "Can you tell me . . .," for all to see, and invite students to use it to frame their information request. Give students 2 minutes of quiet think time.

Student Task Statement

100 captive Asian and 100 wild Asian elephants are weighed. There is a meaningful difference between the masses of the 2 groups if the measures of center are at least twice as far apart as the measure of variability. Is there a meaningful difference between the masses of these 2 groups of elephants? Explain your reasoning.

Student Response

Sample responses:

- Can you tell me the mean and median for each group?
- Can you tell me the standard deviation and interquartile range for each group?
- Can you tell me whether the distributions are symmetric or not?

Activity Synthesis

Tell students that the problem is a part of an *Information Gap* routine. In the routine, one person has a problem with incomplete information, and another person has data that can help with solving it. Explain that it is the job of the person with the problem to think about what is needed to answer the question, and then to request it from the person with information.

Tell students that they will try to solve the problem this way as a class to learn the routine. In this round, the students have the problem, and the teacher has the information.

- Ask students, "What specific information do you need to find out whether there is a significant difference?"
- Select students to ask their questions. Respond to each question with, "Why do you need to know _____?"
- After students justify their question, answer questions only if the questions can be answered using these data:

- The distribution of masses for each set of elephants is approximately symmetric.
- Captive Asian elephants:
  - Mean mass of captive elephants: 3,073 kg
  - Standard deviation of captive elephant masses: 282 kg
  - Median mass of captive elephants: 3,055 kg
  - IQR of captive elephant masses: 399 kg
- Wild Asian elephants:
  - Mean mass of wild elephants: 2,373 kg
  - Standard deviation of wild elephant masses: 121 kg
  - Median mass of wild elephants: 2,386 kg
  - IQR of wild elephant masses: 163 kg

- If students ask for information that is not on the data card, respond with, "I don't have that information."

When students think they have enough information, give them 2 minutes to solve the problem. (There is a significant difference in mass between captive and wild Asian elephants. Because the distributions are symmetric, it makes sense to use the mean and standard deviation masses for the two groups. The difference in means of 700 kilograms is very different, even in light of the standard deviations for the groups.)

Tell students that they will work in small groups and use the routine to solve problems in the next activity.

Math Community

At the end of the *Warm-up*, display the Math Community Chart. Tell students that norms are expectations that help everyone in the room feel safe, comfortable, and productive doing math together. Using the Math Community Chart, offer an example of how the "Doing Math" actions can be used to create norms. For example, "In the last exercise, many of you said that our math community sounds like 'sharing ideas.' A norm that supports that is 'We listen as others share their ideas.' For a teacher norm, 'questioning vs telling' is very important to me, so a norm to support that is 'Ask questions first to make sure I understand how someone is thinking.'"

Invite students to reflect on both individual and group actions. Ask, "As we work together in our mathematical community, what norms, or expectations, should we keep in mind?" Give 1–2 minutes of quiet think time and then invite as many students as time allows to share either their own norm suggestion or to "+1" another student's suggestion. Record student thinking in the student and teacher "Norms" sections on the Math Community Chart.

Conclude the discussion by telling students that what they made today is only a first draft of math community norms and that they can suggest other additions during the *Cool-down*. Throughout the year, students will revise, add, or remove norms based on those that are and are not supporting the community.

13.2 Information Gap: African and Asian Elephants

30 mins

Activity Narrative

This is the first *Information Gap* activity in the course. In this activity, students compare distributions but do not initially have enough information to do so. To bridge the gap, they need to exchange questions and ideas.

The *Information Gap* structure requires students to make sense of problems by determining what information is necessary, and then to ask for information that they need to solve it. This may take several rounds of discussion if their first requests do not yield the information that they need (MP1). It also allows them to refine the language that they use and to ask increasingly more precise questions until they get the information they need (MP6).

Access for English Language Learners

This is the first time *Math Language Routine 4: Information Gap* is suggested in this course. This routine facilitates meaningful interactions by positioning some students as holders of information that is needed by other students, thereby creating a need for communication. This routine supports language development by providing students with opportunities to ask for and share information, and to justify their reasoning within conversation.

Standards

Building On **HSS-ID.A.1**

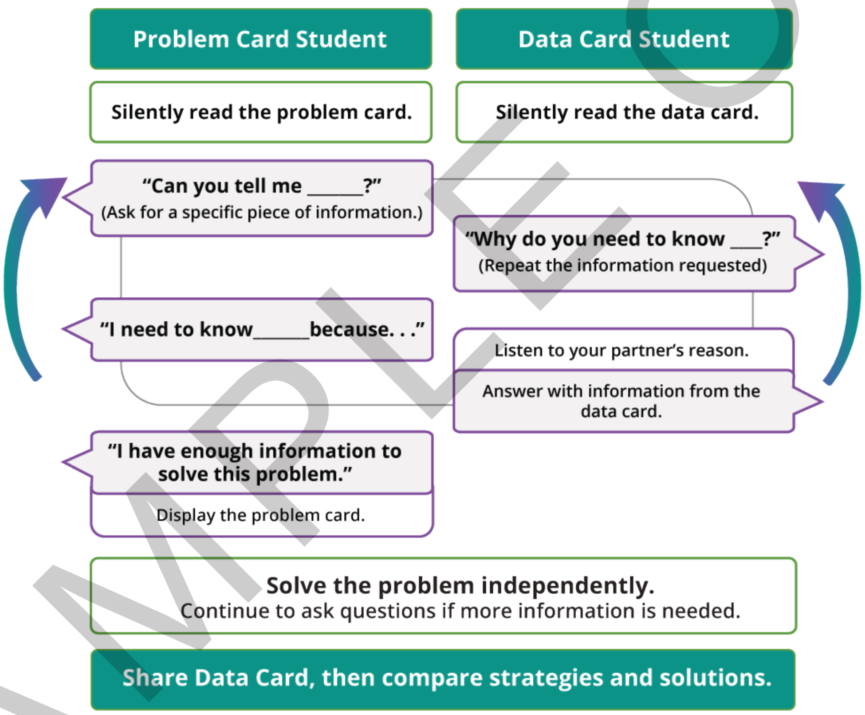
Addressing **HSS-ID.A.2**

Instructional Routines

- MLR4: Information Gap Cards

Launch

Display for all to see the graphic that illustrates a framework for the *Information Gap* routine.



Explain that in an *Information Gap* routine students work with a partner. One partner gets a problem card with a question that doesn't have enough given information, and the other partner gets a data card with information relevant to the problem card.

The person with the problem card asks questions like "Can you tell me ____?" and is expected to explain what they will do with the information. If that person asks for information that is not on the data card (including the answer!) and gives their reason, then the person with the data card must respond with, "I don't have that information." The person with the data card should only be providing information, not making assumptions. Note that it is okay to help a stuck partner by saying something like "I don't have the median mass. I only have information about the mean."

Once the partner with the problem card has enough information, both partners look at the problem card and solve the problem independently.

Arrange students in groups of 2 or 4. If students are new to the *Information Gap* routine, allowing them to work in groups of 2 for each role supports communication and understanding. In each group, distribute a problem card to one student (or group) and a data card to the other student (or group). After reviewing their work on the first problem, give them the cards for a second problem, and instruct them to switch roles.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. Check for understanding by inviting students to rephrase directions in their own words. Keep a display of the *Information Gap* graphic visible throughout the activity or provide students with a physical copy.
Supports accessibility for: Memory, Organization

Student Task Statement

Your teacher will give you either a *problem card* or a *data card*. Do not show or read your card to your partner.

If your teacher gives you the *problem card*:

1. Silently read your card and think about what information you need to answer the question.
2. Ask your partner for the specific information that you need. “Can you tell me _____?”
3. Explain to your partner how you are using the information to solve the problem. “I need to know _____ because . . . .”

   Continue to ask questions until you have enough information to solve the problem.
4. Once you have enough information, share the problem card with your partner, and solve the problem independently.
5. Read the data card, and discuss your reasoning.

If your teacher gives you the *data card*:

1. Silently read your card. Wait for your partner to ask for information.
2. Before telling your partner any information, ask, “Why do you need to know _____?”
3. Listen to your partner’s reasoning and ask clarifying questions. Only give information that is on your card. Do not figure out anything for your partner!

   These steps may be repeated.
4. Once your partner says they have enough information to solve the problem, read the problem card, and solve the problem independently.
5. Share the data card, and discuss your reasoning.

Student Response

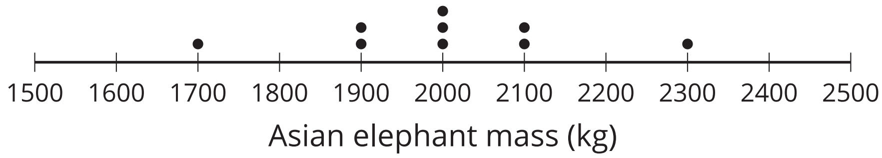
Problem Card 1:

Sample response:

1. The typical masses of the elephants in population A are about 130 kilograms more massive than those in population B based on the mean. I used the mean to determine this since the distributions are symmetric.
2. The elephants in population B have greater variation in mass because their standard deviation is about 76 kilograms greater. The symmetry of the distributions means that standard deviation is a more appropriate measure of variability.

Problem Card 2:

Sample response:



Activity Synthesis

After students have completed their work, share the correct answers, and ask students to discuss the process of solving the problems. Here are some questions for discussion:

- "What was interesting about Problem Card 1?" (It was interesting to use a real context to think about variability.)
- "Was it more appropriate to use the mean or the median to compare the typical masses for Problem Card 1? Why?" (The mean because of the symmetry of the data distribution.)
- "Which measure of variability did you choose to compare for Problem Card 1? Why?" (The standard deviation, because I used the mean, and the data has a symmetric distribution.)
- "What was challenging about Problem Card 2?" (It was difficult to figure out how to make the dot plot without the actual data).
- "How do you know your dot plot is correct?" (There is no way to know. It can only be approximated based on the mean, the standard deviation, and the approximate shape.)

Highlight for students how they used the information about the center, shape, and variability of the distribution to solve the problems.

13.3 Interpreting Measures of Center and Variability

Optional

10 mins

Activity Narrative

This task is provided for optional, additional practice for understanding variability in context.

The mathematical purpose of this activity is to compare data sets, and to interpret the measures of center and measures of variability in the context of the problem.

Standards

Building On HSS-ID.A.1

Addressing HSS-ID.A.2

Instructional Routines

- MLR8: Discussion Supports

Launch

Keep students in the same groups. Give students 5 minutes to work on the questions, and follow with a whole-class discussion.

Access for Students with Disabilities

Action and Expression: Develop Expression and Communication. To help get students started, display sentence frames such as “In this situation, a larger/smaller value for the standard deviation means _____.” or “Both dot plots are different because . . . .”
Supports accessibility for: Language, Organization

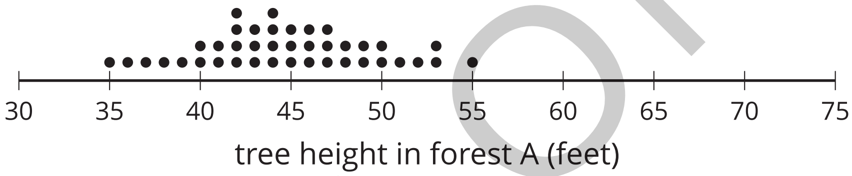
Student Task Statement

For each situation, you are given two graphs of data, a measure of center for each, and a measure of variability for each.

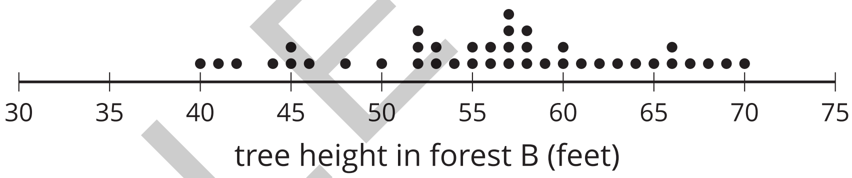
- Interpret the measure of center in terms of the situation.
- Interpret the measure of variability in terms of the situation.
- Compare the two data sets.

1. The heights of the 40 trees in each of two forests are collected.

   mean: 44.8 feet, standard deviation: 4.72 feet

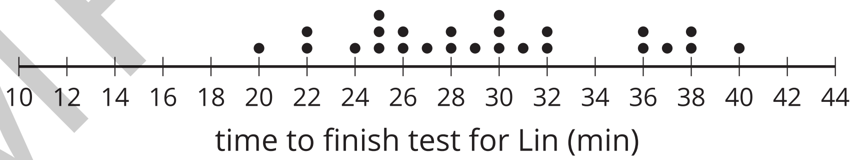


   mean: 56.03 feet, standard deviation: 7.87 feet

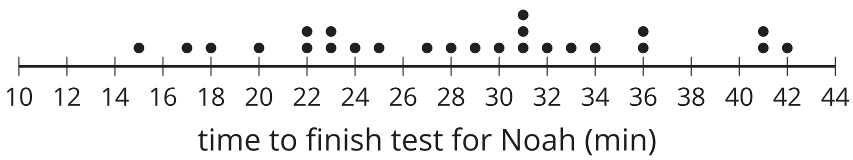


2. The number of minutes it takes Lin and Noah to finish their tests in German class is collected for the year.

   mean: 29.48 minutes, standard deviation: 5.44 minutes



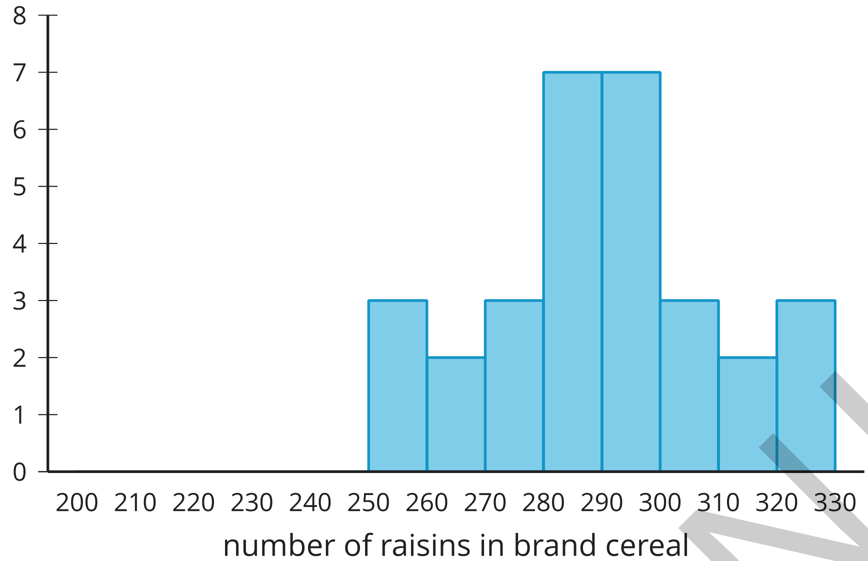
   mean: 28.44 minutes, standard deviation: 7.40 minutes



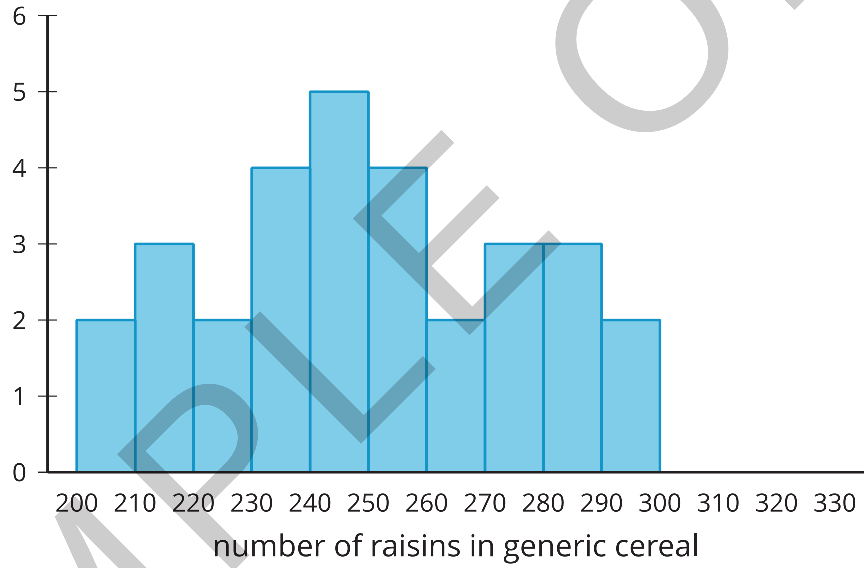
3. The number of raisins in a cereal with a name brand and the generic version of the same cereal are collected for several boxes.

Sec D

mean: 289.1 raisins, standard deviation: 19.8 raisins



mean: 249.17 raisins, standard deviation: 26.35 raisins



Student Response

Sample responses:

1. a. The mean tells us the average height of trees in each forest. Some will be taller and some shorter.
   b. The standard deviation is a measure of how variable the tree heights are. A larger value means there are much taller and much shorter trees than average while a smaller value means they are mostly the same height.
   c. The trees in forest B tend to be taller based on the mean, but they are also more variable, so trees in forest A will be a little more uniform than in forest B.
2. a. The mean tells us the average number of minutes it takes each person to finish the tests for German.
   b. The standard deviation represents how variable the times are. A greater value means sometimes the person finishes much slower or faster than the mean while a smaller value means the person finishes around the

same time for each test.

c. Noah is faster to finish the tests on average, but Lin is more consistent in her finishing times.

3. a. The mean tells us the average number of raisins in a box of cereal.

   b. The standard deviation shows how variable the number of raisins are. A smaller value means the boxes of that type of cereal tend to be more consistent with the number of raisins in the box while a larger value means there is more variety in different boxes of the same type of cereal.

   c. The name brand cereal has more raisins on average and is more consistent with the number of raisins in a box. The generic cereal tends to have fewer raisins, but is also less consistent with the number of raisins in a box.

Building on Student Thinking

Identify students who struggle making the connections between the summary statistics and the problem context. Ask them what information the standard deviation conveys. Tell them to focus less on the actual value of the standard deviation and more on which of the two distributions have a greater standard deviation and what that might mean.

Are You Ready for More?

One use of standard deviation is that it gives a natural scale as to how far above or below the mean a data point is. This is incredibly useful for comparing points from two different distributions.

For example, there is a saying that you cannot compare apples and oranges, but here is a way. The average weight of a granny smith apple is 128 grams with a standard deviation of about 10 grams. The average weight of a navel orange is 140 grams with a standard deviation of about 14 grams. If we have a 148 gram granny smith apple and a 161 gram navel orange, we might wonder which is larger for its species even though they are both about 20 grams above their respective mean. We could say that the apple, which is 2 standard deviations above its mean, is larger for its species than the orange, which is only 1.5 standard deviations above its mean.

1. How many standard deviations above the mean height of a tree in forest A is its tallest tree?
2. How many standard deviations above the mean height of a tree in forest B is its tallest tree?
3. Which tree is relatively taller in its forest?

Extension Student Response

1. Approximately 2.16 standard deviations.
2. Approximately 1.78 standard deviations.
3. The tree in forest A.

Activity Synthesis

The purpose of this discussion is for students to explain the mean and the standard deviation in context.

For each set of graphs, select students read their answers for all three prompts. When necessary, prompt students to revise their language to include the terms shape, measure of center, and variability.

Access for English Language Learners

MLR8 Discussion Supports. For each observation that is shared, invite students to turn to a partner and restate

what they heard, using precise mathematical language.
Advances: Listening, Speaking

Lesson Synthesis

Here are some questions for discussion.

- "What information does standard deviation tell you about a data set?" (It tells you how variable the data are relative to the mean.)
- "Two professional race car drivers have the same average lap times after fifty laps. What does it mean to say that the first driver's lap times have a greater standard deviation than the second driver's lap times?" (It means that even though the drivers had the same average lap time, the first driver showed greater variability in the lap times of the 50 laps. The first driver has lap times that were slower and faster and less clustered near the mean than the second driver's lap times.)
- "In one class, students had an average height of 68 inches and a standard deviation of 5 inches. In a second class, the students had an average height of 67 inches and a standard deviation of 3.5 inches. What differences would you expect to see if you looked at the dot plots that represent the distributions of data for each class?" (I would expect that the data in the dot plot for the first class would be more spread apart and centered slightly to the right when compared to the data in the dot plot for second class. I might expect that the data values for the second class would be more tightly clustered around 67 inches than the data values for the first class would be clustered around 68 inches.)

13.4 Majors and Salaries

Cool-down

5 mins

Standards

Building On HSS-ID.A.1
Addressing HSS-ID.A.2

Launch

Math Community

Before distributing the *Cool-downs*, display the Math Community Chart and the norms question "Which norm has not already been listed that you'd like to add to our chart?" Ask students to respond to the question after completing the *Cool-down* on the same sheet.

After collecting the *Cool-downs*, identify themes from the norms question. Use that information to add to the initial draft of the "Norms" sections of the Math Community Chart.

Student Task Statement

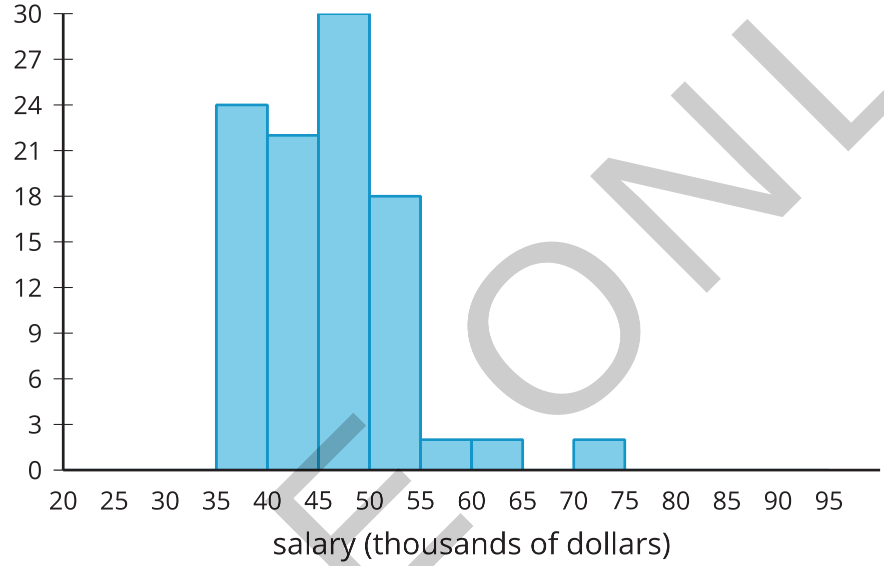
A college is looking at the data for its most recent college graduates based on their major.

- The mean salary of 100 recent college graduates who majored in engineering is $63,750 with a standard deviation of $10,020.

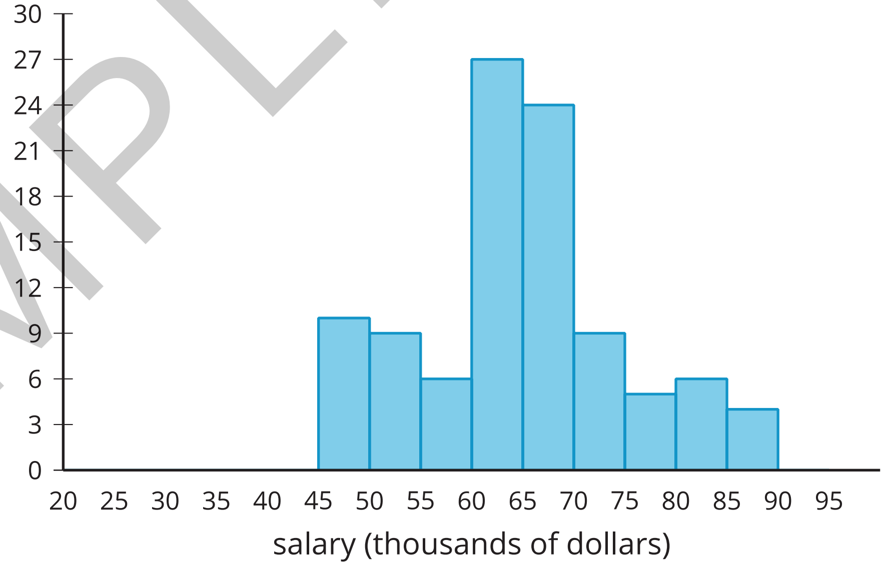
- The mean salary of 100 recent college graduates who majored in business is $52,200 with a standard deviation of $19,400.
- The mean salary of 100 recent college graduates who majored in the social sciences is $45,230 with a standard deviation of $6,750.

Match each histogram to the majors based on the description.

- engineering
- business
- social sciences



1.



2.

3.

Student Response

1. social sciences
2. engineering
3. business

Sec D

Responding To Student Thinking

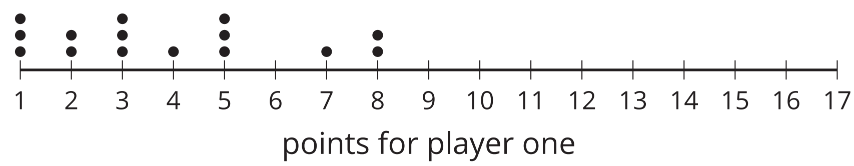
Points to Emphasize
If students struggle to recognize mean as the center of the distribution and standard deviation as the spread, look for opportunities for students to practice estimating these values from histograms. For example, at the start of the *Warm-up* referred to here, display the histogram for all to see, and invite students to estimate the mean and standard deviation.

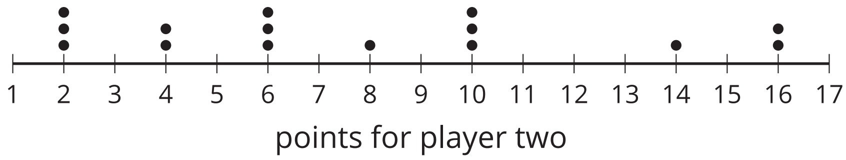
Algebra 1, Unit 1, Lesson 14, Activity 1 Health Care Spending

Lesson 13 Summary

The more variation a distribution has, the greater the standard deviation. A more compact distribution will have a lesser standard deviation.

The first dot plot shows the number of points that a player on a basketball team made during each of 15 games. The second dot plot shows the number of points scored by another player during the same 15 games.





The data in the first plot have a mean of approximately 3.87 points and standard deviation of about 2.33 points. The data in the second plot have a mean of approximately 7.73 points and a standard deviation of approximately 4.67 points. The second distribution has greater variability than the first distribution because the data are more spread out. This is shown in the standard deviation for the second distribution being greater than the standard deviation for the first distribution.

Standard deviation is calculated using the mean, so it makes sense to use it as a measure of variability when the mean is appropriate to use for the measure of center. In cases where the median is a more appropriate measure of center, the interquartile range is still a better measure of variability than standard deviation.

Practice Problems

1 Student Task Statement

Three drivers compete in the same fifteen races. The mean and standard deviation for the race times of each of the drivers are given.

Driver A has a mean race time of 4.01 seconds and a standard deviation of 0.05 seconds.

Driver B has a mean race time of 3.96 seconds and a standard deviation of 0.12 seconds.

Driver C has a mean race time of 3.99 seconds and a standard deviation of 0.19 seconds.

a. Which driver has the fastest typical race time?
b. Which driver's race times are the most variable?
c. Which driver do you predict will win the next race? Support your prediction using the mean and standard deviation.

Solution

a. Driver B
b. Driver C
c. Sample response: I predict driver C will win. Even though driver C's mean race time is in between the other two mean race times, driver C shows more variability. This means that driver C sometimes is relatively slow, but sometimes is relatively fast. (Note that one could also reasonably predict driver B would win, because Driver B has the smallest mean and the standard deviation of Driver B's race times is smaller than that of driver C.)

2 Student Task Statement

The widths, in millimeters, of fabric produced at a ribbon factory are collected. The mean is approximately 23 millimeters, and the standard deviation is approximately 0.6 millimeters.

Interpret the mean and standard deviation in the context of the problem.

Solution

Sample response: The width of the fabric is typically 23 millimeters. The standard deviation of 0.6 millimeters means that there was very little variability. The width of most of the fabric is between 22.4 and 23.6 millimeters.

3 Student Task Statement

Select **all** the statements that are true about standard deviation.

A. It is a measure of center.

B. It is a measure of variability.

C. It is the same as the MAD.

D. It is calculated using the mean.

E. It is calculated using the median.

Solution

B, D

4 from Unit 1, Lesson 12

Student Task Statement

The number of different species of plants in some gardens is recorded.

1 2 3 4 4 5 5 6 7 8

a. What is the mean?

b. What is the standard deviation?

Solution

a. 4.5 species

b. Approximately 2.06 species

5 from Unit 1, Lesson 12

Student Task Statement

A set of data has ten numbers. The mean of the data is 12 and the standard deviation is 0. What values could make up a data set with these statistics?

Solution

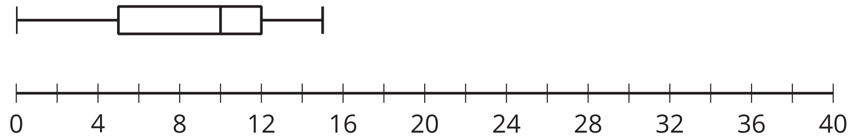
12, 12, 12, 12, 12, 12, 12, 12, 12, 12

6 from Unit 1, Lesson 11

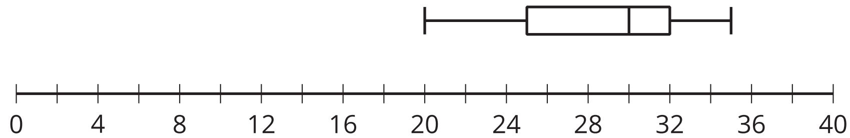
Student Task Statement

Which box plot has the largest interquartile range?

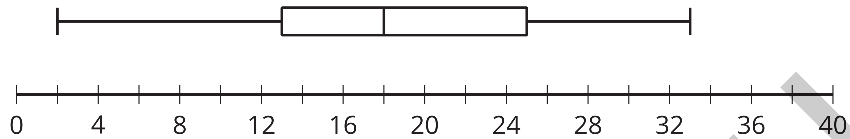
A.



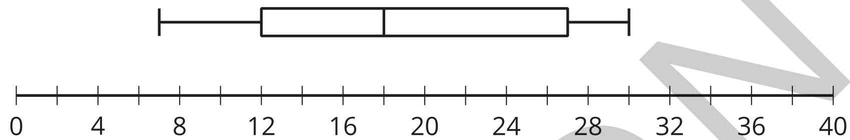
B.



C.



D.



Solution

D

7 from Unit 1, Lesson 9

Student Task Statement

a. What is the five-number summary for?

1 3 3 3 4 8 9 10 10 17

b. When the maximum, 17, is removed from the data set, what is the five-number summary?

Solution

a. 1, 3, 6, 10, 17 (Minimum, Q1, Median, Q3, Maximum)

b. 1, 3, 4, 9.5, 10 (Minimum, Q1, Median, Q3, Maximum)

Unit 1, Lesson 14

Outliers

Goals

- Describe (orally and in writing) how outliers impact measure of center and measures of variability.
- Determine (in writing) when values are considered outliers, investigate their source, and determine if they should be excluded from the data.

Learning Targets

- I can find values that are outliers, investigate their source, and figure out what to do with them.
- I can tell how an outlier affects mean, median, IQR, or standard deviation.

Lesson Narrative

Students are introduced to **outliers**, investigate their source, make decisions about excluding them from the data set, and examine how the presence of outliers impacts measures of center and measures of variability.

When students analyze data in the context of a problem to determine whether or not to exclude an outlier, they are reasoning abstractly and quantitatively (MP2). This reasoning process is also an aspect of mathematical modeling (MP4).

Standards

Building On	**HSS-ID.A.1**
Addressing	**HSS-ID.A.1, HSS-ID.A.2, HSS-ID.A.3**
Building Towards	**HSS-ID.A.3**

Instructional Routines

- Analyze It
- MLR2: Collect and Display
- Notice and Wonder

Required Materials

Materials To Gather

- Statistical technology: Lesson
- Statistical technology: Activity 2

Materials To Copy

- Algebra 1 Unit 1 Useful Terms and Displays (1 copy for every 30 students): Activity 1

Student Facing Learning Goals

Let's investigate outliers and how to deal with them.

14.1 Health Care Spending

Warm-up

10 mins

Activity Narrative

The purpose of this *Warm-up* is to elicit the idea that outliers are often present in data, which will be useful when students investigate the source of outliers and what to do with them in a later activity.

As students work, monitor for students who

- Estimate the IQR from values in the box plot.
- Use a measurement tool to determine the IQR from the box plot.

Students are given the formulas for **outliers**: A value is considered an outlier for a data set if it is greater than Q3 + 1.5 • IQR or less than Q1 - 1.5 • IQR. To find extreme values, we are comparing very large or small values to the bulk of the data. This means using the quartiles and interquartile range to compare the value to typical distances to the center of the data.

Standards

Addressing HSS-ID.A.1

Building Towards HSS-ID.A.3

Instructional Routines

- Notice and Wonder

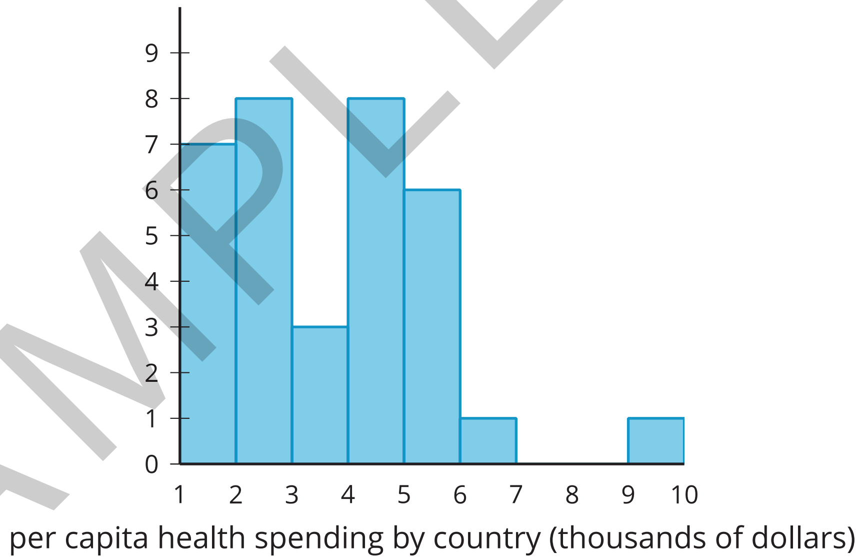
Launch

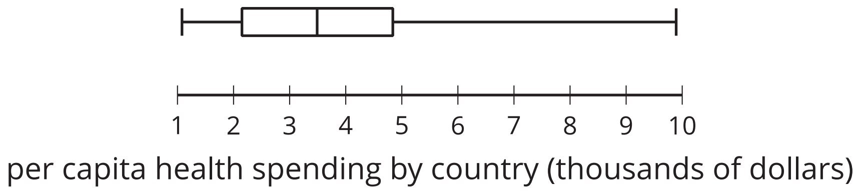
Display the histogram and box plot for all to see. Tell students to think of one thing they notice and one thing they wonder about the images. Give students 1 minute of quiet think time, and then 1 minute to discuss with a partner the things they notice. Listen for students who notice that there is a value that seems greatly different from the rest of the data. Select a few students to share things they notice and wonder, making sure to select identified students who notice an extreme value.

Sec D

Student Task Statement

The histogram and box plot show the average amount of money, in thousands of dollars, spent on each person in the country (per capita spending) for health care in 34 countries.



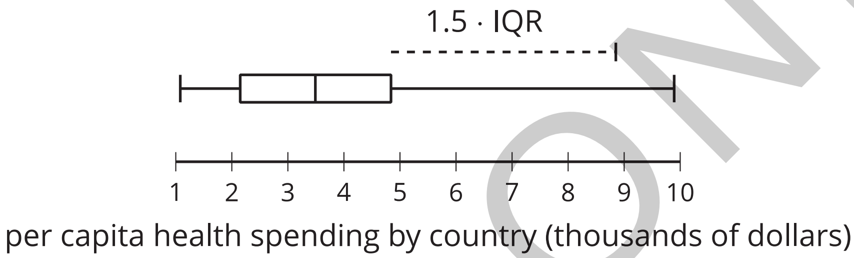


1. One value in the set is an **outlier**. Which one is it? What is its approximate value?
2. By one rule for deciding, a value is an outlier if it is more than 1.5 times the IQR greater than Q3. Show on the box plot whether or not your value meets this definition of outlier.

Student Response

Sample response:

1. The value between 9 and 10 thousand dollars is the outlier.



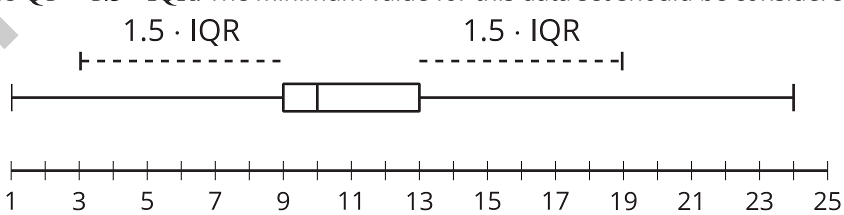
2.

Activity Synthesis

Select previously identified students in the order listed in the lesson narrative to share their method for creating this visualization of outliers in the box plot.

Tell students:

- Values in a data set that are greatly different from the rest of the data are called outliers. The precise meaning of *greatly different* will be different for different situations. For example, a possible $4,000 difference in this graph does seem like a lot, but if the data represented the entire budgets of these countries in the billions or trillions of dollars (rather than spending on each member of the population for healthcare), it would not be a great difference.
- Using the IQR to determine outliers helps to adjust the difference to the variability of the bulk of the middle data. Using 1.5 times the IQR allows for some variability on the ends of the distribution to be considered usual.
- It is also possible for there to be values that are unusually low compared to the rest of the data set. Consider this box plot that displays $Q1 - 1.5 \cdot IQR$. The minimum value for this data set should be considered an outlier.



- For the purposes of this unit, a value will be considered an outlier for a data set if it is greater than Q3 + 1.5 • IQR or less than Q1 - 1.5 • IQR. These formulas compare extreme values to the middle half of the data to determine if the value should be considered an outlier.

Add "outlier" to the classroom display created in earlier lessons. The blackline master provides an example of what this

Sec D

display may look like after all items are added.

14.2 Investigating Outliers

15 mins

Activity Narrative

The mathematical purpose of this activity is for students to investigate the effect of outliers on measures of center and variability, and to make decisions about whether or not to include outliers in a data set.

Standards

Building On **HSS-ID.A.1**

Addressing **HSS-ID.A.2, HSS-ID.A.3**

Instructional Routines

- Analyze It
- MLR2: Collect and Display

Launch

Arrange students in groups of 2. Provide access to devices that can run GeoGebra or other statistical technology.

Display the data showing Per Capita Health Spending by Country in 2016 for all to see. Orient students to the data, and explain that the distribution of this data set is represented in the histogram used in the *Warm-up*. Ensure that students understand what per capita means. “Per capita health spending” means the average health spending per person. For example, the United States spends approximately $9,892 on healthcare for each person in the population.

Remind students that we will classify a value in a data set as an outlier if it is greater than Q3 + 1.5 • IQR or less than Q1 - 1.5 • IQR.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support development of organizational skills in problem-solving, chunk this task into more manageable parts. For example, present one question at a time. *Supports accessibility for: Organization, Attention*

Student Task Statement

Here is the data set used to create the histogram and box plot from the warm-up.

1.0803	1.0875	1.4663	1.7978	1.9702	1.9770	1.9890	2.1011	2.1495	2.2230	2.5443
2.7288	2.7344	2.8223	2.8348	3.2484	3.3912	3.5896	4.0334	4.1925	4.3763	4.5193
4.6004	4.7081	4.7528	4.8398	5.2050	5.2273	5.3854	5.4875	5.5284	5.5506	6.6475
9.8923										

1. Use technology to find the mean, standard deviation, and five-number summary.
2. The maximum value in this data set represents the spending for the United States. Should the per capita health spending for the United States be considered an outlier? Explain your reasoning.

3. Although outliers should not be removed without considering their cause, it is important to see how influential outliers can be for various statistics. Remove the value for the United States from the data set.
    a. Use technology to calculate the new mean, standard deviation, and five-number summary.
    b. How do the mean, standard deviation, median, and interquartile range of the data set with the outlier removed compare to the same summary statistics of the original data set?

Student Response

1. Mean: 3.7259, standard deviation: 1.827, minimum: 1.0803, Q1: 2.1495, median: 3.4904, Q3: 4.8398, maximum: 9.8923
2. Yes, since $4.84 + 1.5 \cdot 2.69 = 8.875$ and $9.89 > 8.875$.
3.
    a. Mean: 3.5391, standard deviation: 1.501, minimum: 1.0803, Q1: 2.1253, median: 3.3912, Q3: 4.7963, maximum: 6.6475
    b. Mean decreased by 0.1868, standard deviation decreased by 0.326, median decreased by 0.0992, and IQR decreased by 0.0193 (IQR before removing is 2.6903 and IQR after removing is 2.671).

Building on Student Thinking

Students may incorrectly compute the expression for outliers. Remind them to use the correct order of operations, using the *Math Talk Warm-up* in a previous lesson.

Activity Synthesis

The goal is to make sure that students understand that outliers can significantly affect measures of center and variability. Discuss the effect of the outlier on the median, mean, and standard deviation and the student responses to "Do you think that 9.8923 should be eliminated from the data set? Why or why not?"

If time permits, discuss questions such as:

- "Which measure of center is more greatly affected by the inclusion of extreme values, the mean or median? Explain your reasoning." (The mean because it uses the actual numerical value rather than the position of the values like the median does.)
- "Which measure of variability is more greatly affected by the inclusion of extreme values, the standard deviation or the interquartile range? Explain your reasoning." (The standard deviation because it uses the mean as well as the numerical value of each number in the data set, whereas the IQR uses only the position of the middle half of the data.)

Access for English Language Learners

MLR2 Collect and Display. Circulate, listen for, and collect the language that students use as they use statistical tools to calculate and display numerical statistics. On a visible display, record words and phrases such as "mean," "median," "outliers," or "standard deviation." Invite students to borrow language from the display as needed, and update it throughout the lesson.
Advances: Conversing, Reading

14.3 Origins of Outliers

10 mins

Activity Narrative

The mathematical purpose of this activity is to get students thinking about the source of outliers and whether or not it is appropriate to include them when analyzing data. It is important to stress that data should not be removed simply because it is an outlier. If there is any doubt about the reason for the outlier, the data should be included in any analysis done on the data set.

Standards

Building On **HSS-ID.A.1**

Addressing **HSS-ID.A.3**

Launch

Arrange students in groups of 2. Give students quiet think time to answer the first question. Ask partners to compare answers.

Student Task Statement

1. The number of passenger electric vehicles registered is collected for the 39 counties of Washington State.
    - mean: 3,589.3 cars
    - minimum: 3 cars
    - Q1: 170 cars
    - median: 506 cars
    - Q3: 1,560 cars
    - maximum: 73,996 cars

    16223 337 460 60 467 73996 8368 1556 222 806 1736 3 238 3444 165

    706 3497 10657 37 278 186 424 170 45 4425 4601 36 677 554 25 858

    12 796 773 1560 44 194 841 506

    a. Are any of the values outliers? Explain or show your reasoning.

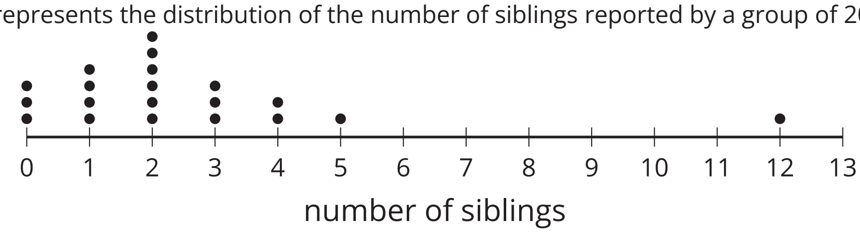
    b. If there are any outliers, why do you think they might exist? Should they be included in an analysis of the data?

2. The situations described here each have an outlier. For each situation, how would you determine if it is appropriate to keep or remove the outlier when analyzing the data? Discuss your reasoning with your partner.

    a. A number cube has sides labeled 1–6. After rolling 15 times, Tyler records his data:

    1 1 1 1 2 2 3 3 4 4 5 5 5 6 20

b. The dot plot represents the distribution of the number of siblings reported by a group of 20 people.



c. In a science class, 11 groups of students are synthesizing biodiesel. At the end of the experiment, each group recorded the mass in grams of the biodiesel they synthesized. The masses of biodiesel are

0 1.245 1.292 1.375 1.383 1.412 1.435 1.471 1.482 1.501 1.532

Student Response

Sample responses:

1. a. The values 4425, 4601, 8368, 10657, 16223, and 73996 are each outliers, since these values are each greater than 3645 ($Q3 + 1.5 \cdot IQR = 1560 + 1.5 \cdot 1390$ which is 3645).
   b. Washington State has counties with larger and smaller populations. The data should be included in an analysis, since it is not an error in recording or collecting information.
2. a. The outlier, 20, should be eliminated because it is not a plausible value. That is, a number cube cannot roll the number 20.
   b. The outlier, 12 siblings, should be investigated further to see if it was a data collection error before it is removed. While it is unusual for someone to have 12 siblings, it is still possible. If it is not an error or if it cannot be determined, it should be kept for any further analysis.
   c. The outlier, 0 grams, should be investigated further to see if the experiment was done correctly or if there was an error in following the directions or materials used. If the experiment was done correctly, the data should be kept, but if there is an error in following the directions or in the materials used, the 0 could be excluded.

Are You Ready for More?

Look back at some of the numerical data that you and your classmates collected in the first lesson of this unit.

1. Are any of the values outliers? Explain or show your reasoning.
2. If there are any outliers, why do you think they might exist? Should they be included in an analysis of the data?

Extension Student Response

Answers vary.

Activity Synthesis

The purpose of this discussion is to highlight different reasons that outliers appear in data. For example, they could be data-entry or data collection errors, or they could be representative of the sample. The goal is to make sure that students understand that the inclusion of outliers in a data set needs to be evaluated in the context of the data. For the number cube rolls, it is clear that the data should not be used since it is impossible to achieve in the right circumstances. For the other two scenarios, students should understand that a deeper investigation should be done to determine whether the outlier should be included and be able to state circumstances for including or excluding the outlier in each context.

Here are some questions for discussion:

- "Why is it important to analyze the source of outliers?" (To determine if all of the data fit conditions for the question being asked so that the analysis is valid. It can also give insights into different situations that may not be common.)
- "What are reasons to keep an outlier in a data set?" (Just because a value is extremely large or small does not discount its reality. To be honest in the analysis, all valid data should be included.)
- "What are reasons to remove an outlier from a data set?" (If there is an error in data collection or recording, the data may be faulty, and it would not be honest to include data that do not fit the question asked.)
- "What could be done about the 6 outliers for the electric vehicle data to account for county size as the source of the outliers?" (The question could be changed to examine the number of electric vehicles in counties with a population below some level. It may make better sense to look at the percentage of cars that are electric or something similar that takes into account the county size rather than just the number of electric cars.)
- "How do you know that a value is an outlier?" (If it is greater than the third quartile by more than one and a half times the interquartile range or less than the first quartile by the same amount.)

Lesson Synthesis

The purpose of this discussion is to make sure that students know what outliers are, what to do with them, and how they impact measures of center and measures of variability. Here are some questions for discussion:

- "What is an outlier?" (A data value that differs from the other values in the data set. It can be defined in terms of the IQR and quartiles. If a data value is 1.5 times the IQR greater than the 3rd quartile or 1.5 times the IQR less than the 1st quartile, then it is considered an outlier.)
- "Why are outliers important to notice in a data set?" (They can indicate an error in the data collection process or an interesting case to more closely study. They are not always representative of the whole data set. Their presence can disproportionally affect the values of the mean, MAD, and standard deviation.)
- "How do outliers affect measures of center?" (Outliers can cause the mean to be much higher or lower than what appears to be typical depending on if the outlier is much greater or much less than the mean. They have less effect on the median.)
- "How do outliers affect measures of variability?" (They cause the variability to be higher than it would be if the outliers were not present.)
- "Why would you eliminate an outlier?" (An outlier would be eliminated if it is an error or if it is not representative of the sample as a whole. It depends on the context of the problem and the data collection process.)

14.4 Expecting Outliers

Cool-down

5 mins

Standards

Addressing **HSS-ID.A.2, HSS-ID.A.3**

Student Task Statement

A group of 20 students are asked to report the number of pets they keep in their house. The results are:

0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 3, 4, 4, 4, 21

- mean: 2.4 pets
- standard deviation: 4.47 pets
- Q1: 0.5 pets
- median: 1 pet
- Q3: 2.5 pets

1. Would any of these values be considered outliers? Explain your reasoning.
2. After being told that they should not count any fish in the report, the value of 3 becomes a 2 and the value of 21 becomes 1. Would these changes affect the median, mean, standard deviation, or interquartile range? If so, would each measure decrease or increase from their original values?

Student Response

1. Yes, 21 pets is an outlier since it is greater than $5.5 = 2.5 + 1.5 \cdot 2$.
2. The mean and standard deviation would decrease with the changes. The median would stay the same and the IQR would decrease slightly.

Responding To Student Thinking

Points to Emphasize
If most students struggle to explain why 21 is an outlier using calculations, revisit how to do so as opportunities arise. For example, in the *Activity Synthesis* of the activity referred to here, discuss the importance of calculating outliers in removing bias rather than pointing out the greatest or least values.

Algebra 1, Unit 1, Lesson 15, Activity 2 Comparing Marathon Times

Lesson 14 Summary

In statistics, an **outlier** is a data value that is unusual in that it differs quite a bit from the other values in the data set.

Outliers occur in data sets for a variety of reasons including, but not limited to:

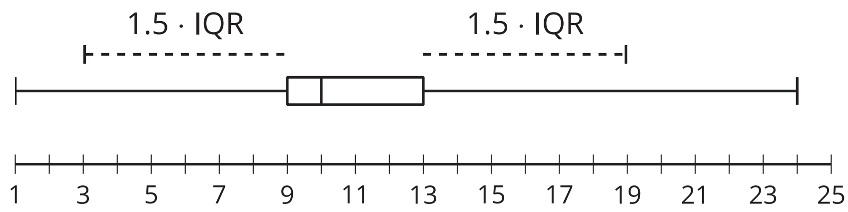
- Errors in the data that result from the data collection or data entry process.
- Results in the data that represent unusual values that occur in the population.

Outliers can reveal cases worth studying in detail or errors in the data collection process. In general, they should be included in any analysis done with the data.

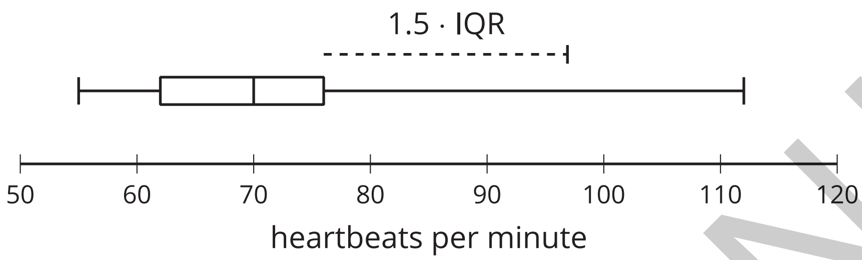
A value is an outlier if it is

- More than 1.5 times the interquartile range greater than Q3 (if $x > \text{Q3} + 1.5 \cdot \text{IQR}$).
- More than 1.5 times the interquartile range less than Q1 (if $x < \text{Q1} - 1.5 \cdot \text{IQR}$).

In this box plot, the minimum and maximum are at least two outliers.



It is important to identify the source of outliers because outliers can affect measures of center and variability in significant ways. The box plot displays the resting heart rate, in beats per minute (bpm), of 50 athletes taken five minutes after a workout.



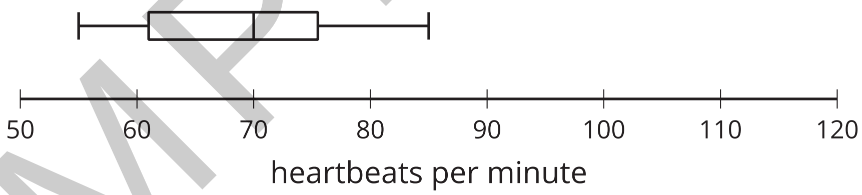
Some summary statistics include:

- mean: 69.78 bpm
- standard deviation: 10.71 bpm
- minimum: 55 bpm
- Q1: 62 bpm
- median: 70 bpm
- Q3: 76 bpm
- maximum: 112 bpm

It appears that the maximum value of 112 bpm may be an outlier. Beacuse the interquartile range is 14 bpm ($76 - 62 = 14$) and $Q3 + 1.5 \cdot IQR = 97$, we should label the maximum value as an outlier. Searching through the actual data set, it could be confirmed that this is the only outlier.

After reviewing the data collection process, it is discovered that the athlete with the heart rate measurement of 112 bpm was taken one minute after a workout instead of five minutes after. The outlier should be deleted from the data set because it was not obtained under the right conditions.

Once the outlier is removed, the box plot and summary statistics are:



- mean: 68.92 bpm
- standard deviation: 8.9 bpm
- minimum: 55 bpm
- Q1: 61 bpm
- median: 70 bpm
- Q3: 75.5 bpm
- maximum: 85 bpm

The mean decreased by 0.86 bpm and the median remained the same. The standard deviation decreased by 1.81 bpm which is about 17% of its previous value. Based on the standard deviation, the data set with the outlier removed shows much less variability than the original data set containing the outlier. Because the mean and standard deviation use all of the numerical values, removing one very large data point can affect these statistics in important ways.

The median remained the same after the removal of the outlier and the IQR increased slightly. These measures of center and variability are much more resistant to change than the mean and standard deviation are. The median and IQR measure the middle of the data based on the number of values rather than the actual numerical values themselves, so the loss of a single value will not often have a great effect on these statistics.

The source of any possible errors should always be investigated. If the measurement of 112 beats per minute was found to be taken under the right conditions and merely included an athlete whose heart rate did not slow as much as the other athletes' heart rate, it should not be deleted so that the data reflect the actual measurements. If the situation cannot be revisited to determine the source of the outlier, it should not be removed. To avoid tampering with the data and to report accurate results, data values should not be deleted unless they can be confirmed to be an error in the data collection or data entry process.

Glossary

- outlier

Practice Problems

1 Student Task Statement

The number of letters received in the mail over the past week is recorded.

2 3 5 5 5 15

Which value appears to be an outlier?

A. 2

B. 3

C. 5

D. 15

Solution

D

2 Student Task Statement

Elena collects 112 specimens of beetle and records their lengths for an ecology research project. When she returns to the laboratory, Elena finds that she incorrectly recorded one of lengths of the beetles as 122 centimeters (about 4 feet). What should she do with the outlier, 122 centimeters, when she analyzes her data?

Solution

Sample response: Elena should go back to the 112 specimens and measure them again to find the error. If she cannot do this, then she should eliminate the outlier from her analysis, because this is an error in recording.

3 Student Task Statement

Mai took a survey of students in her class to find out how many hours they spend reading each week. Here are some summary statistics for the data that Mai gathered:

mean: 8.5 hours　　standard deviation: 5.3 hours

median: 7 hours　　Q1: 5 hours　　Q3: 11 hours

a. Give an example of a number of hours larger than the median which would be an outlier. Explain your reasoning.

b. Are there any outliers below the median? Explain your reasoning.

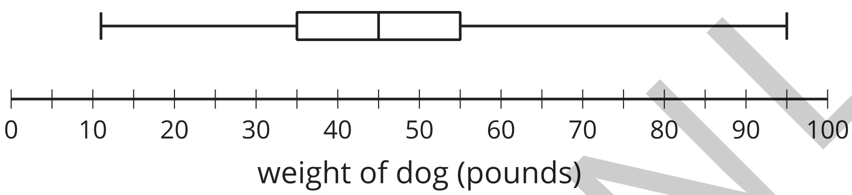
Solution

a. Sample response: 22 hours, since Q3 together with one and a half times the interquartile range is 20.

b. No, because outliers below the median must be less than -4 ($5 - 1.5 \cdot (11 - 5) = -4$), and it is not possible to read for a negative number of hours.

4 Student Task Statement

The box plot shows the statistics for the weight, in pounds, of some dogs.



Are there any outliers? Explain how you know.

Solution

Yes, there is at least one outlier that is too high. The dog that weighs about 95 pounds is an outlier. The IQR is about 20 and $1.5 \cdot 20 = 30$. The dog that weighs about 95 pounds is about 40 pounds above the third quartile.

Sec D

5

from Unit 1, Lesson 13

Student Task Statement

The mean exam score for the first group of twenty examinees applying for a security job is 35.3 with a standard deviation of 3.6.

The mean exam score for the second group of twenty examinees is 34.1 with a standard deviation of 0.5. Both distributions are close to symmetric in shape.

a. Use the mean and standard deviation to compare the scores of the two groups.

b. The minimum score required to get an in-person interview is 33. Which group do you think has more people get in-person interviews?

Solution

a. Sample response: The first group scored higher on average based on the mean, but showed much greater variability than the second group, because the first group had a higher standard deviation.

b. Sample response: I think more people in the second group get in-person interviews. The minimum score is 33 and they have a mean of 34.1 with a standard deviation of 0.5, so the bulk of their scores had to be above 33. If not, they would have had a standard deviation greater than 1.1.

6 from Unit 1, Lesson 13

Student Task Statement

A group of pennies made in 2018 are weighed. The mean is approximately 2.5 grams with a standard deviation of 0.02 grams.

Interpret the mean and standard deviation in terms of the context.

Solution

The average penny weighs 2.5 grams. There is a little bit of variation in the weights, but most of the pennies weigh very close to 2.5, within 0.02 grams on average as measured by the standard deviation.

7 from Unit 1, Lesson 12

Student Task Statement

These values represent the expected number of paintings a person will produce over the next 10 days.

0 0 0 1 1 1 2 2 3 5

a. What are the mean and standard deviation of the data?

b. The artist is not pleased with these statistics. If the 5 is increased to a larger value, how does this affect the median, mean, and standard deviation?

Solution

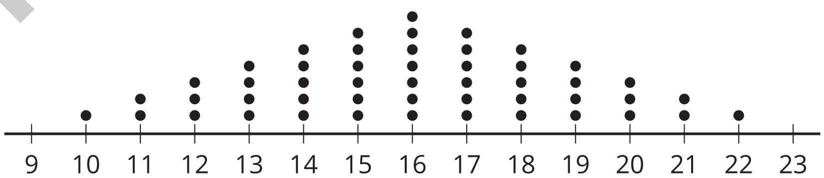
a. Both the mean and the standard deviation are 1.5 paintings.

b. The median will still be 1 painting. The mean will increase, and so will the standard deviation, because the data are more spread out.

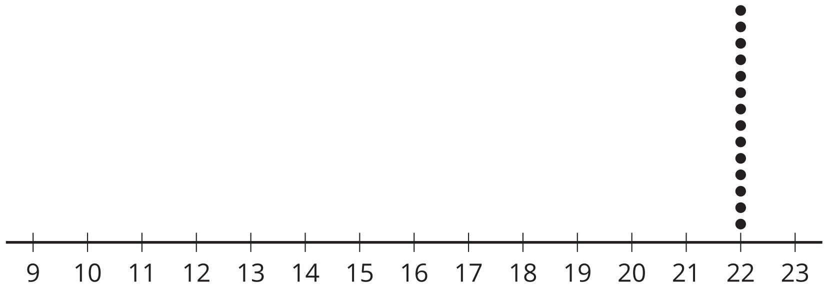
8 from Unit 1, Lesson 11

Student Task Statement

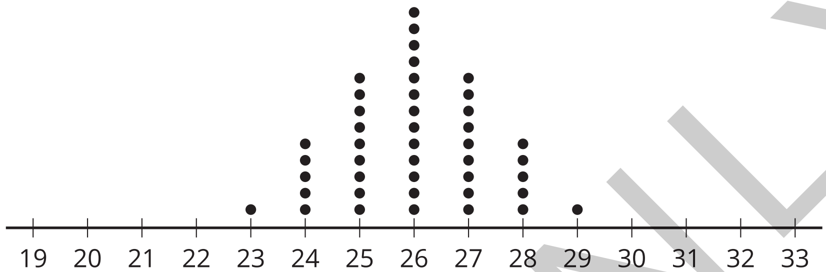
List the four dot plots in order of variability from least to greatest.

a. 

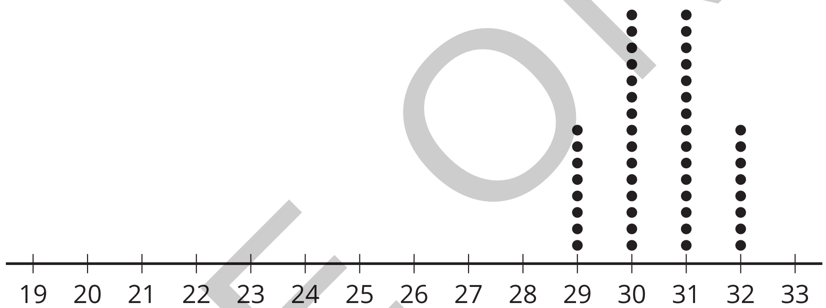
b.



c.



d.



Solution

b, d, c, a

Unit 1, Lesson 15

Comparing Data Sets

Goals

- Compare and contrast (orally and in writing) situations using measures of center and measures of variability.

Learning Targets

- I can compare and contrast situations using measures of center and measures of variability.

Lesson Narrative

Students compare measures of center, standard deviation, and the IQR for different data sets. This lesson provides opportunities for students to collaborate, share mathematical ideas, and reflect on their mathematical thinking about measures of center and measures of variability. This lesson also gives students another opportunity to compare measures of center and variability, but now using standard deviation and outliers in the comparisons. When students are describing measures of center and measures of variability in the context of marathon time they are reasoning abstractly and quantitatively (MP2) because they are interpreting the meaning of their answer in context.

Standards

Building On	**HSS-ID.A.1**
Addressing	**HSS-ID.A.1, HSS-ID.A.2**
Building Towards	**HSS-ID.A.2**

Instructional Routines

- 5 Practices
- MLR5: Co-Craft Questions
- Take Turns

Student Facing Learning Goals

Let's compare statistics for data sets.

15.1 Bowling Partners

Warm-up

10 mins

Activity Narrative

The mathematical purpose of this activity is for students to compare different distributions using shape, measures of center, and measures of variability. This *Warm-up* prompts students to compare four distributions representing recent bowling scores for potential teammates. It gives students a reason to use language precisely (MP6) and gives you the opportunity to hear how they use terminology and talk about characteristics of the images in comparison to one another.

Standards

Addressing	**HSS-ID.A.1**
Building Towards	**HSS-ID.A.2**

Launch

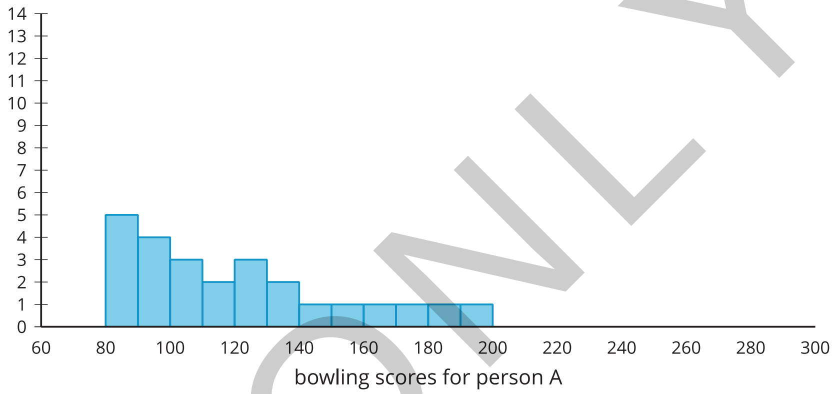
Arrange students in groups of 2–4. Tell students that bowling is a game in which a higher score is better and the maximum score is 300. It is typical for non-professional bowlers to score a little over 100.

Student Task Statement

Each histogram shows the bowling scores for the last 25 games played by each person. Choose 2 of these people to join your bowling team. Explain your reasoning.

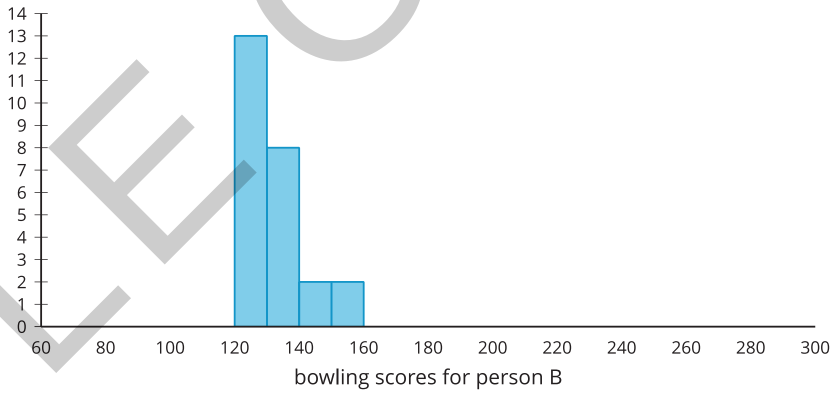
Person A

- mean: 118.96
- median: 111
- standard deviation: 32.96
- interquartile range: 44



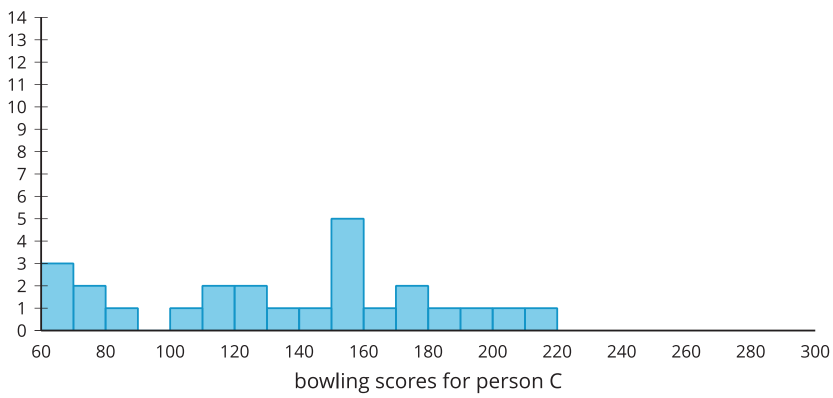
Person B

- mean: 131.08
- median: 129
- standard deviation: 8.64
- interquartile range: 8



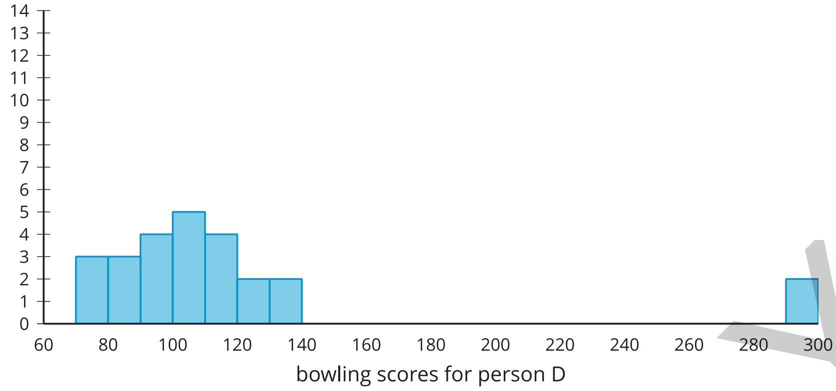
Person C

- mean: 133.92
- median: 145
- standard deviation: 45.04
- interquartile range: 74



Person D

- mean: 116.56
- median: 103
- standard deviation: 56.22
- interquartile range: 31.5



Student Response

Reasons for each bowler:

1. Although Person A has a low mean score, the large variability of scores indicates that this person can score well sometimes. This person's scores are spread out, and it is not rare for this person to score over 140.
2. Person B is very consistent. Although this person does not have great games like the other players, Person B's mean score is near the top, and this person will reliably get a score between 120 and 160.
3. Person C has a high mean score but has scores that are also really variable. This person had the lowest score in this group but has the greatest median score. I think we could coach this person to be more consistent, so this wild card could be on my team.
4. Person D has the lowest mean, but is also capable of bowling a near perfect game (twice). This person had two games with scores that were much greater than anyone else's on the list, so maybe Person D does well when it counts.

Activity Synthesis

Ask each group to share one bowler they would choose and their reasoning. If none of the groups select a certain player, ask why this player was not chosen, or give reasons why another team may want this player on their team. If time allows, ask if there is any additional information that might make that player more desirable. For example, knowing the conditions behind each score might be helpful. Player C might be a new bowler, and the lower scores might have been when Player C was learning, but the newer scores may all be closer to 200. Player D might not take practice seriously, but can bowl a perfect game in competition.

Because there is no single correct answer, attend to students' explanations, and ensure that the reasons given are correct. During the discussion, ask students to explain how they used the statistics given as well as the histograms.

15.2 Comparing Marathon Times

15 mins

Activity Narrative

The mathematical purpose of this activity is for students to compare measures of center and measures of variability in context. Monitor for students who

- Determine the slower age group by using an informal description of the shift in data.
- Determine the slower age group by using a numerical estimate for the mean or median for measures of center.
- Determine variability from the range of values.
- Use a numerical estimate for IQR or standard deviation as measures of variability.

Plan to have students present strategies in this order—from less to more formal descriptions of the distribution.

Access for English Language Learners

This activity uses the *Co-Craft Questions* math language routine to allow students to make sense of a context and to practice generating mathematical questions.

Standards

Building On **HSS-ID.A.1**

Addressing **HSS-ID.A.2**

Instructional Routines

- 5 Practices
- MLR5: Co-Craft Questions

Launch

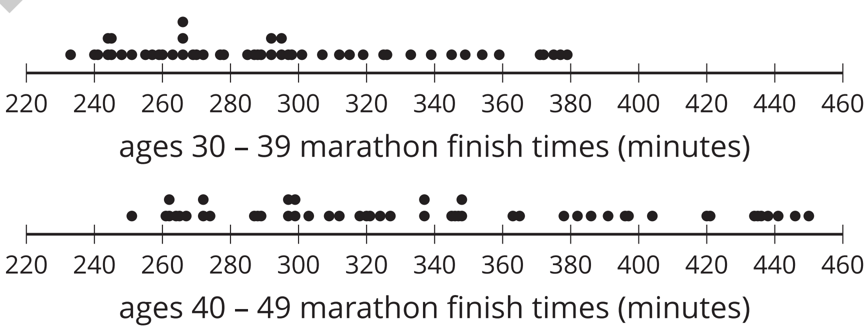
Arrange students in groups of 2. Introduce the context of the marathon runners. Use *Co-Craft Questions* to orient students to the context and to elicit possible mathematical questions.

- Display only the problem stem and related image, without revealing the questions. Give students 1–2 minutes to write a list of mathematical questions that could be asked about the situation. Then ask them to compare their questions with a partner.
- Invite several partners to share one question with the class, and record responses. Ask the class to make comparisons among the shared questions and their own. Ask, "What do these questions have in common? How are they different?" Listen for and amplify language related to the learning goal, such as how the measures of center and variability can be used to compare the groups.
- Use this time to also answer any questions that students may have about understanding marathons that would help with these questions.

Select students who used each strategy described in the activity narrative, and ask them to share later. Aim to elicit both key mathematical ideas and a variety of student voices, especially from students who haven't shared recently.

Student Task Statement

All of the marathon runners from each of two different age groups have their finishing times represented in the dot plot.



1. Which age group tends to take longer to run the marathon? Explain your reasoning.
2. Which age group has more variable finish times? Explain your reasoning.

Student Response

Sample responses:

1. The runners ages 40–49 are slower because they have a greater median finish time. The center of the data is farther to the right on this distribution.
2. The runners ages 40–49 have times that are more variable because this data are more spread out along the number line than the data in the dot plot for ages 30–39 are. They have a greater IQR because the middle half of the data look more spread out.

Are You Ready for More?

1. How do you think finish times for a 20–29 age range will compare to these two distributions?
2. Find some actual marathon finish times for this group and make a dot or box plot of your data to help compare.

Extension Student Response

Answers vary.

Activity Synthesis

The purpose of this discussion is for students to understand how to compare data sets using measures of center and measures of variability.

Invite previously selected students to share their answers and reasoning. Sequence the discussion of the strategies by the order listed in the *Activity Narrative*. If possible, record and display their work for all to see.

After several estimates for measure of center and measure of variation are mentioned, display the actual values for these data sets.

Ages 30–39

- Mean: 294.5 minutes
- Standard deviation: 41.84 minutes
- Median: 288.5 minutes
- IQR: 65 minutes
- Q1: 260 minutes
- Q3: 325 minutes

Ages 40–49

- Mean: 340.6 minutes
- Standard deviation: 59.93 minutes
- Median: 332 minutes
- IQR: 92 minutes

- Q1: 289 minutes
- Q3: 381 minutes

Connect the different responses to the learning goals by asking questions such as:

- "How did you use measures of center and variability to compare the distributions?" (I used median and IQR to compare the distributions because both distributions appeared skewed. The median time for the older group was slower by a little than an IQR.)
- "Using your method, would you say there is a lot of overlap between the data distributions for the two groups?" (Yes, the measures of center are about 1 measure of variability away from one another, so they are not significantly different.)
- "If a runner took 300 minutes to complete the race, which group would you think that runner is more likely to belong to? Explain your reasoning." (300 minutes is closer to the median time for the 30–39 age group. So, I might guess that the runner is in that group, but due to the overlap, it would not be surprising if the runner were in the 40–49 age group instead.)
- "Which points show values that are most likely to be outliers?" (The time values for the slowest runners in each group, on the right end of the dot plot, might be outliers.)
- "Based on the displayed information, are there any outliers in these data sets?" (No. For the 30–39 age group, outliers would need to be values greater than 422.5 minutes, but the slowest times are around 380 minutes. For the 40–49 age group, outliers would need to be greater than 534 minutes, but the longest times are around 450 minutes.)

15.3 Comparing Measures

15 mins

Activity Narrative

In this activity, students take turns with a partner determining the best measure of center and the best measure of variability for several data sets. Students trade roles explaining their thinking and listening, providing opportunities to explain their reasoning and critique the reasoning of others (MP3). Students also determine which data set has a greater measure of center and which has a greater measure of variability.

Standards

Building On **HSS-ID.A.1**

Addressing **HSS-ID.A.2**

Instructional Routines

- Take Turns

Launch

Arrange students in groups of 2. Tell students that for each data display or description of a data set in column A, one partner determines the appropriate measure of center and measure of variability and explains why it is appropriate. The partner's job is to listen and make sure they agree. If they don't agree, the partners discuss until they come to an agreement. For the next data display or description of a data set in column B, the students swap roles. If necessary, demonstrate this protocol before students start working. The last item has a column C. Students can work together to determine the best measures for set C. When an agreement is reached for each group of data sets, students will determine which data set has the greatest measure of center, and which data set has the greatest measure of variability.

If time allows, ask students to work through all 7 sets, otherwise have each group select one pair of dot plots, one pair of box plots, and one descriptive group.

Access for Students with Disabilities

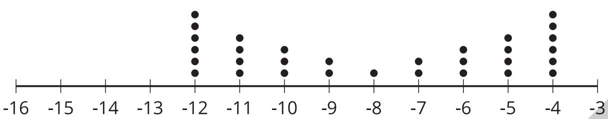
Engagement: Develop Effort and Persistence. Encourage and support opportunities for peer collaboration. When students share their work with a partner, display sentence frames to support conversation, such as "That measure of center could (or couldn't) be true because . . . ." or "Based on the shape of the distribution, a better choice would be _____ because . . . ."
Supports accessibility for: Language, Social-Emotional Functioning

Student Task Statement

For each group of data sets,

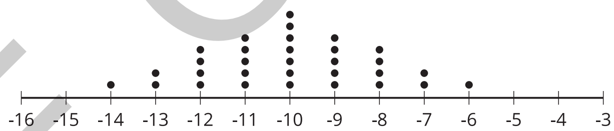
- Determine the best measure of center and measure of variability to use based on the shape of each distribution.
- Determine which set has the greatest measure of center.
- Determine which set has the greatest measure of variability.
- Be prepared to explain your reasoning.

1a



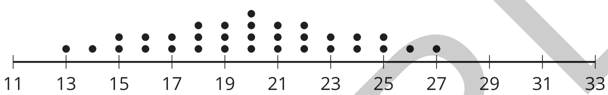
distribution 1a

1b



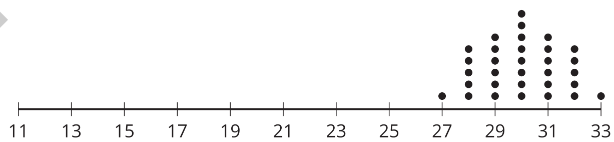
distribution 1b

2a



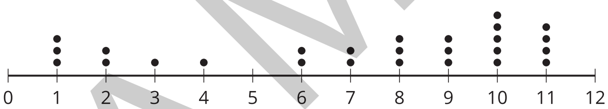
distribution 2a

2b



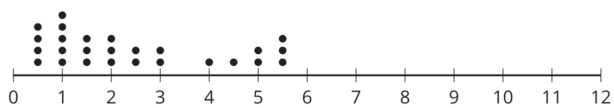
distribution 2b

3a



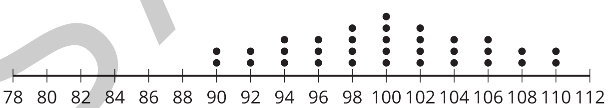
distribution 3a

3b



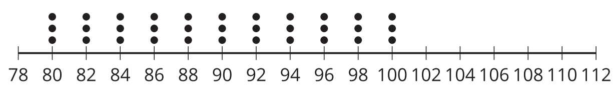
distribution 3b

4a



distribution 4a

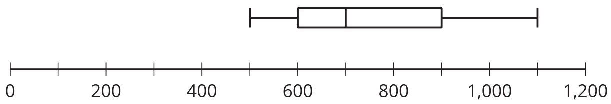
4b



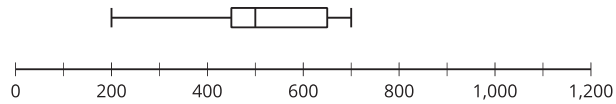
distribution 4b

5a

5b



distribution 5a



distribution 5b

6a

A political podcast has mostly reviews that either love the podcast or hate it.

6b

A cooking podcast has reviews that neither hate nor love the podcast.

7a

Stress testing concrete from site A has all 12 samples break at 450 pounds per square inch (psi).

7b

Stress testing concrete from site B has samples break every 10 psi starting at 450 psi until the last core is broken at 560 psi.

7c

Stress testing concrete from site C has 6 samples break at 430 psi and the other 6 break at 460 psi.

Student Response

1. The mean and standard deviation are appropriate for these distributions since they are symmetric. The mean for distribution 1a is greater since the center is to the right of the center for distribution 1b. The standard deviation for distribution 1a is also greater since most of the data are farther away from the mean than in distribution 1b.
2. The mean and standard deviation are appropriate for these distributions since they are symmetric. The mean for distribution 2b is greater since the center is to the right of the center for distribution 2a. The standard deviation for distribution 2a is greater since the data are more spread out in that distribution.
3. The median and interquartile range are appropriate for these distributions since they are not symmetric. The median for distribution 3a is greater since the center is to the right of the center for distribution 3b. The interquartile range is greater for distribution 3a since the data are more spread out in that distribution.
4. The mean and standard deviation are appropriate for these distributions since they are symmetric. The mean for distribution 4a is greater since the center is to the right of the center for distribution 4b. The standard deviation for distribution 4b is greater since the data are more spread out in that distribution.
5. The median and interquartile range are most appropriate for these distributions since they are not symmetric and the box plot makes these values clear. The median for distribution 5a is greater since the median is 700 units, but for distribution 5b it is 500 units. The IQR is greater for distribution 5a since it is 300 units, but only 200 units for distribution 5b.
6. Distribution 6a has greater variability since the values will be concentrated on the ends for the political podcast while the cooking podcast likely has values more clustered near the center. Answers vary for the other determinations.
7. Since all three distributions are symmetric, mean and standard deviation are the best measures to use. Site B will have the greatest mean since the least value for that distribution is the mean for site A (450 psi) and greater than the mean for site C (445 psi). The standard deviation for site B will be the greatest since site A has a standard deviation of 0 psi, site C has all of its data within 15 psi of the mean, and site B only has 4 of the values within 15 psi of the mean.

Building on Student Thinking

For the situations described in words, students may think there is not enough information to answer the question. Ask these students, "What do you think the distributions might look like for the situations described?" Tell them to use their distributions to answer the question and be prepared to explain their reasoning.

Activity Synthesis

Select students to share how they determined whether to use the mean or the median, and how they figured out which data set showed greater variability.

- "What were some ways you handled the last two problems?" (I reasoned about what a dot plot might look like to imagine where the center might be and how spread out the data might look.)
- "Describe any difficulties you experienced and how you resolved them." (I forgot what information was in a box plot, so I asked my partner.)
- "How did you decide whether to use the mean or the median?" (I recalled what I learned about shape from the previous lesson. When the distribution was symmetric or close to it, I used the mean. When it was skewed, I used the median.)
- "How did you decide which data set showed greater variability?" (It was easy for the box plot, I calculated the IQR. For the dot plots, I looked at which one was more spread apart. For the problem contexts, I tried to figure out which one would have data that would be more varied.)

Lesson Synthesis

The purpose of this discussion is to ensure that students know how to compare data sets using measures of variability, including standard deviation, and measures of center. Here are some questions for discussion.

- "How do you compare the measures of variability for a data set?" (You either calculate them or estimate them from a data display. The data set with the higher measure of variability is more variable.)
- "How do you estimate variability when looking at data displays?" (You try to estimate the center and then estimate how spread apart the data are.)
- "How do you determine which measure of center to use for a data set?" (You look at the shape and use the mean when it is symmetric or really close and the median when it is skewed or if there are outliers.)
- "Why is the median the preferred measure of center for skewed distributions?" (The median is preferred because it more accurately represents the center of the data. Data values farther from the center affect the median less than the mean, so the median remains near the typical values.)
- "Why is the mean the preferred measure of center for symmetric distributions?" (In a symmetric distribution, the mean is equal to the median. The mean is preferred because it takes into account all of the values in the data set when it is calculated.)

15.4 Comparing Mascots

Cool-down

5 mins

Standards

Addressing **HSS-ID.A.2**

Student Task Statement

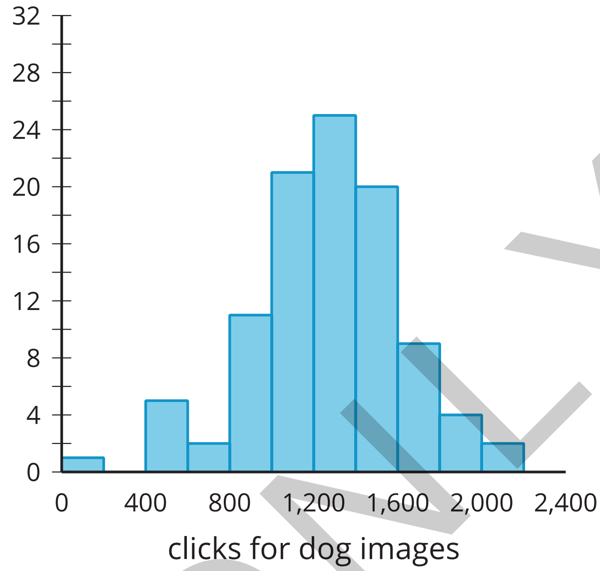
A new pet food company wants to sell their product online and use social media to promote themselves. To determine whether to use a dog or a cat as their mascot, they research the number of clicks on links with an image of a dog or a cat.

mean: 1,263.5 clicks

median: 1,282 clicks

standard deviation: 357.4 clicks

IQR: 409 clicks

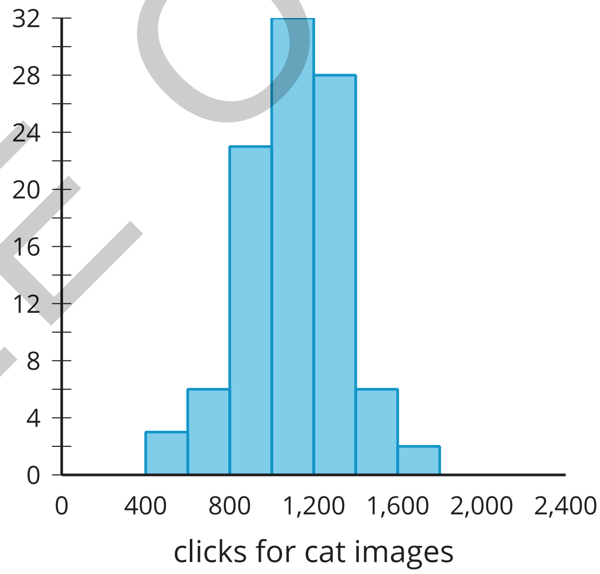


mean: 1,105.4 clicks

median: 1,125.5 clicks

standard deviation: 239.3 clicks

IQR: 312.5 clicks



1. Based on the shape of the distributions, what measure of center and measure of variability would you use to compare the distributions? Explain your reasoning.
2. Based on the data shown here, should the company use a dog or cat mascot? Explain your reasoning.

Student Response

1. Mean and standard deviation. Since the distributions are approximately symmetric, the mean and standard deviation are the best choice to represent the data.
2. Sample responses:
    - The company should use a dog mascot since the mean is greater.
    - The company should use a cat mascot since the standard deviation shows that the images are more consistently clicked over 1,000 times while the dog images sometimes get fewer than 200 clicks.

Responding To Student Thinking

Points to Emphasize

If most students struggle with either identifying the best measure of center to use based on the shape of data, or connecting mean to standard deviation and median to IQR, invite students to share their data in the activity referred to here, and discuss how they chose the appropriate measure of center and variability to describe their data set.

Algebra 1, Unit 1, Lesson 16, Activity 2 Dropping the Ruler

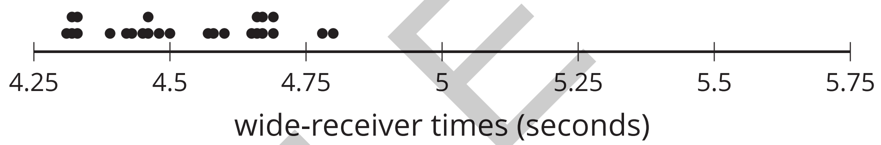
Lesson 15 Summary

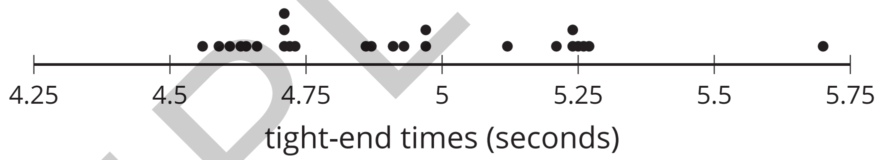
To compare data sets, it is helpful to look at the measures of center and measures of variability. The shape of the distribution can help choose the most useful measure of center and measure of variability.

When distributions are symmetric or approximately symmetric, the mean is the preferred measure of center and should be paired with the standard deviation as the preferred measure of variability. When distributions are skewed or when outliers are present, the median is usually a better measure of center and should be paired with the interquartile range (IQR) as the preferred measure of variability.

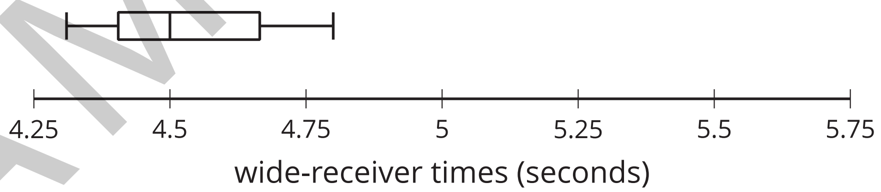
Once the appropriate measure of center and measure of variability are selected, these measures can be compared for data sets with similar shapes.

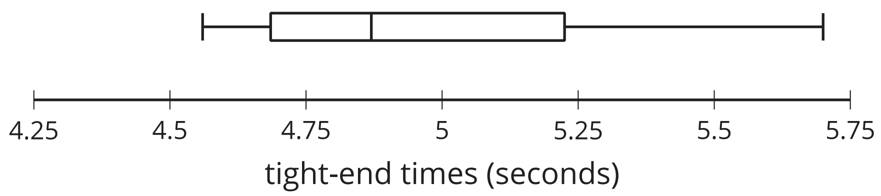
For example, let's compare the number of seconds it takes football players to complete a 40-yard dash at two different positions. First, we can look at a dot plot of the data to see that the tight-end times do not seem distributed symmetrically, so we should probably find the median and IQR for both sets of data to compare information.





The median and IQR could be computed from the values, but can also be determined from a box plot.





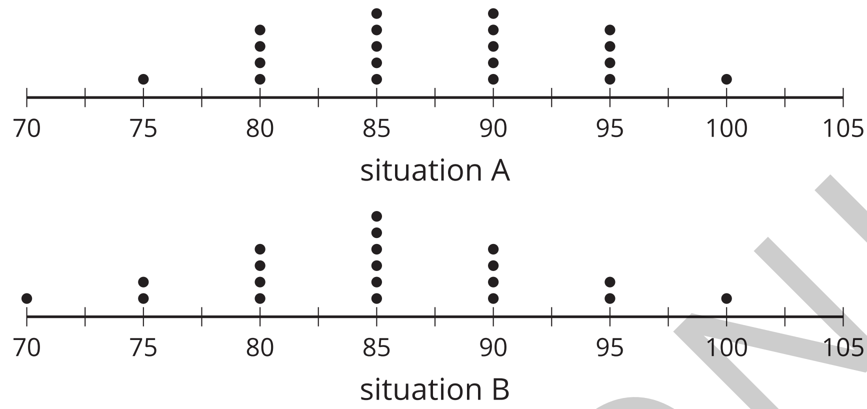
This shows that the tight-end times have a greater median (about 4.9 seconds) compared to the median of wide-receiver times (about 4.5 seconds). The IQR is also greater for the tight-end times (about 0.5 seconds) compared to the IQR for the wide-receiver times (about 0.25 seconds).

This means that the tight ends tend to be slower in the 40-yard dash when compared to the wide receivers. The tight ends also have greater variability in their times. Together, this can be taken to mean that, in general, a typical wide receiver is faster than a typical tight end is, and the wide receivers tend to have more similar times to one another than the tight ends do to one another.

Practice Problems

1 Student Task Statement

Twenty students participated in a psychology experiment that measured their heart rates in two different situations.



a. What are the appropriate measures of center and variability to use with the data? Explain your reasoning.

b. Which situation shows a greater typical heart rate?

c. Which situation shows greater variability?

Sec D

Solution

a. The mean and standard deviation are appropriate because the distributions are both symmetric.

b. Situation A

c. Situation B has a greater standard deviation.

2 Student Task Statement

a. Invent two situations that you think would result in distributions with similar measures of variability. Explain your reasoning.

b. Invent two situations that you think would result in distributions with different measures of variability. Explain your reasoning.

Solution

a. Sample response: Rolling number cubes with your eyes closed and rolling number cubes with your eyes open. They would be similar because closing your eyes should have no effect on the sum of the two number cubes.

b. Sample response: The daily high temperatures in July and the daily high temperatures for the year. The temperatures for the year should vary more than the temperatures for the month because of the change of seasons.

3 Student Task Statement

The data set and some summary statistics are listed.

11.5 12.3 13.5 15.6 16.7 17.2 18.4 19 19.5 21.5

- mean: 16.52
- median: 16.95
- standard deviation: 3.11
- IQR: 5.5

a. How does adding 5 to each of the values in the data set affect the shape of the distribution?

b. How does adding 5 to each of the values in the data set affect the measures of center?

c. How does adding 5 to each of the values in the data set affect the measures of variability?

Solution

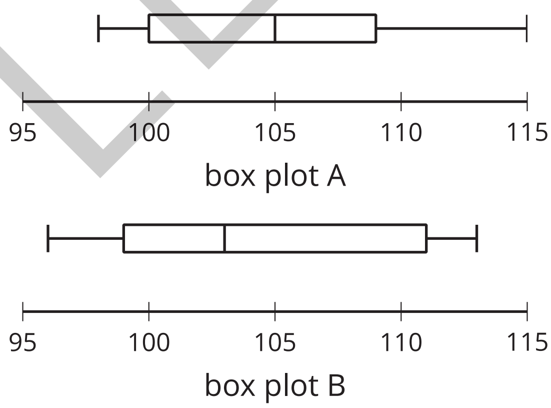
a. Sample response: The shape does not change; the values are all shifted 5 units to the right.

b. Sample response: Both of the measures of center will go up by 5.

c. Sample response: Adding 5 has no affect on the measures of variability.

4 Student Task Statement

Here are two box plots:



a. Which box plot has a greater median?

b. Which box plot has a greater measure of variability?

Solution

a. A

b. B

5 from Unit 1, Lesson 13

Student Task Statement

The depth of two lakes is measured at multiple spots. For the first lake, the mean depth is about 45 feet with a standard deviation of 8 feet. For the second lake, the mean depth is about 60 feet with a standard deviation of 27 feet.

Noah says the second lake is generally deeper than the first lake. Do you agree with Noah?

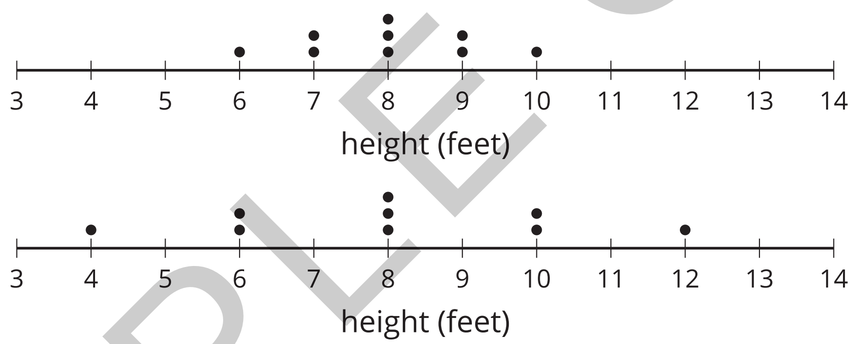
Solution

Sample response: Noah is correct. The average depth of the second lake is substantially greater than the average depth of the first lake, as measured by the mean values. On the other hand, there is a lot more variation in the depth of the second lake, so there may be substantial parts of the second lake that are shallow compared to the first lake.

6 from Unit 1, Lesson 12

Student Task Statement

The dot plots display the height, rounded to the nearest foot, of maple trees from two different tree farms.



a. Compare the mean and standard deviation of the two data sets.
b. What does the standard deviation tell you about the trees at these farms?

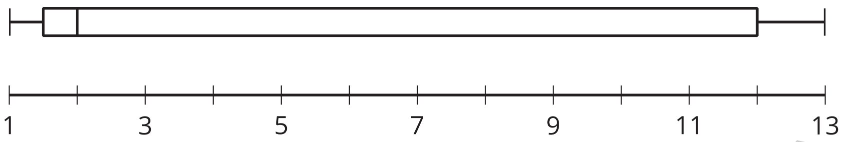
Solution

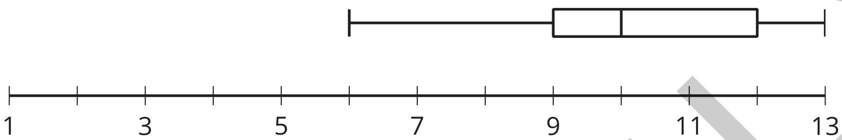
a. The means of the two data sets both equal 8 feet. The standard deviation of the first data set is less than the standard deviation of the second data set.
b. At the second farm, there is a wider range of heights than at the first farm, where the tree heights are closer to one another.

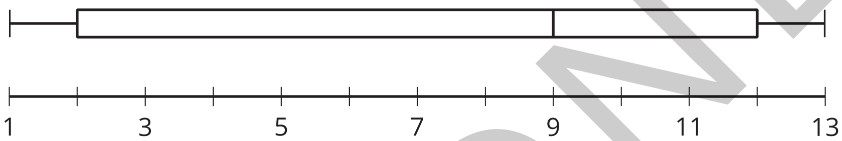
7 from Unit 1, Lesson 11

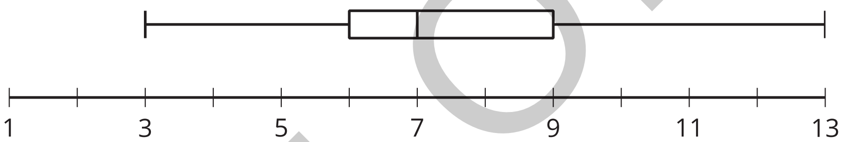
Student Task Statement

Which box plot has an IQR of 10?

A. 

B. 

C. 

D. 

Solution

C

8 from Unit 1, Lesson 9

Student Task Statement

What effect does eliminating the lowest value, -6, from the data set have on the mean and median?

-6 3 3 3 3 5 6 6 8 10

Solution

The mean increases from 4.1 to approximately 5.22. The median increases from 4 to 5.

Sec D

Section E: Let's Put it to Work

Section Narrative

The final section of this unit is a lesson where students have the opportunity to apply their thinking from throughout the unit. As this is a short section followed by an End-of-Unit Assessment, there are no section goals or checkpoint questions.

Teacher Reflection Questions

- **Math Content and Student Thinking:** The work of this unit provided students with a variety of tools for data display and analysis. Which did you see your students using? Were any over- or underused?

Analyzing Data

Goals

- Collect and compare (using words and other representations) data sets with different conditions for an experiment based on the measures of center and measures of variability.

Learning Targets

- I can collect data from an experiment and compare the results using measures of center and measures of variability.

Lesson Narrative

In this lesson, students pose and answer a statistical question by designing an experiment, collecting data, and analyzing data. They determine how best to display data, select appropriate measures of center and variability, and answer a statistical question involving two different treatments.

Technology isn't required for this lesson, but there are opportunities for students to choose to use appropriate technology to solve problems. We recommend making technology available. When students choose an appropriate display for data and choose to use technology to calculate statistics, they are using appropriate tools strategically (MP5). Students make sense of problems and persevere in solving them (MP1) when they make sense of the situation in order to come up with a meaningful statistical question and then follow through to answer their question. Students also must select an appropriate variable to analyze for comparing student heights, which engages them in aspects of mathematical modeling (MP4).

Standards

Addressing	**HSS-ID.A, HSS-ID.A.1, HSS-ID.A.2, HSS-ID.A.3**
Building Towards	**HSS-ID.A.1, HSS-ID.A.2, HSS-ID.A.3**

Instructional Routines

- Analyze It
- Aspects of Mathematical Modeling
- MLR1: Stronger and Clearer Each Time

Required Materials

Materials To Gather

- Rulers: Activity 1

Materials To Copy

- Heights and Handedness Handout (1 copy for every 2 students): Activity 3

Required Preparation

Activity 2:

One ruler for every pair of students.

Student Facing Learning Goals

Let's answer statistical questions by analyzing data, and comparing and contrasting their distribution shape and measures of center and variability.

16.1 Experimental Conditions

Warm-up

5 mins

Activity Narrative

The mathematical purpose of this activity is for students to write a statistical question for dropping and catching a ruler under different conditions.

Standards

Building Towards **HSS-ID.A.1, HSS-ID.A.2, HSS-ID.A.3**

Launch

Arrange students in groups of 2.

Demonstrate how to drop the ruler and how to measure the distance dropped. Show this video if necessary.

Video 'Ruler Drop Demonstration' available here: https://player.vimeo.com/video/304121341.

Explain that the ruler is being held by one person at the 12 inch mark and is caught by another person just below the 7 inch mark. The distance the ruler fell is about 6 inches. For groups struggling to think of conditions that might be interesting, here are some examples to help them get started:

- Standing and sitting
- Listening to music and quiet
- Listening to a favorite song and one that is less interesting
- Releasing quickly after they are ready and waiting at least 3 seconds after they are ready before dropping

Student Task Statement

To test reaction time, the person running the test will hold a ruler at the 12 inch mark. The person whose reaction time is being tested will hold their thumb and forefinger open on either side of the flat side of the ruler at the 0 inch mark on the other side of the ruler. The person running the test will drop the ruler and the other person should close their fingers as soon as they notice the ruler moving to catch it. The distance that the ruler fell should be used as the data for this experiment.

With your partner, write a statistical question that can be answered by comparing data from two different conditions for the test.

Student Response

Sample response: Is my partner's reaction time faster if I drop the ruler while my partner is standing on two feet or while my partner is standing on one foot?

Activity Synthesis

The goal of this discussion is to make sure that everyone has a statistical question about reaction time that will require collecting data from two different conditions to test.

Check student questions and assist them in creating a question that meets the requirement. Here are some questions for discussion.

- “What are your two conditions?” (Standing on one foot and standing on both feet.)
- “How are you going to collect these data values?” (I will hold and drop the ruler while my partner stands on one foot. Then I will repeat this while my partner is standing on two feet.)
- “How many trials do you think you should do under each condition?” (At least ten)

16.2 Dropping the Ruler

25 mins

Activity Narrative

The mathematical purpose of this activity is for students to design an experiment to answer a statistical question, to collect data, to analyze data using statistics, and to communicate the answer to the statistical question using a display. It may be helpful to have multiple groups combine to allow students to experience the different conditions for many experiments. Making statistical technology available gives students an opportunity to choose appropriate tools strategically (MP5).

Notice groups that create displays that communicate their mathematical thinking clearly, contain an error that would be instructive to discuss, or organize the information in a way that is useful for all to see.

Standards

Addressing **HSS-ID.A.1, HSS-ID.A.2, HSS-ID.A.3**

Instructional Routines

- Analyze It
- MLR1: Stronger and Clearer Each Time

Launch

Keep students in groups of 2. Provide each group with tools for creating a visual display. If students have access to statistical technology, suggest that it might be a helpful tool in this activity.

Explain to students that they will collect and analyze data using statistics to answer their statistical question from the warm-up.

Students will create a display showing the statistical questions, the data, a data display, and an answer to the statistical question with any supporting mathematical work.

Access for Students with Disabilities

Action and Expression: Internalize Executive Functions. To support development of organizational skills in problem solving, chunk this task into more manageable parts. For example, show 1–2 features that the display must include instead of the entire list. Provide students with access to grid or graph paper to organize their supporting mathematical work for the visual display.

Supports accessibility for: Memory, Organization

Student Task Statement

Earlier, you and your partner agreed on a statistical question that can be answered using data collected in 2 different ruler-dropping conditions. With your partner, run the experiment to collect at least 20 results under each condition.

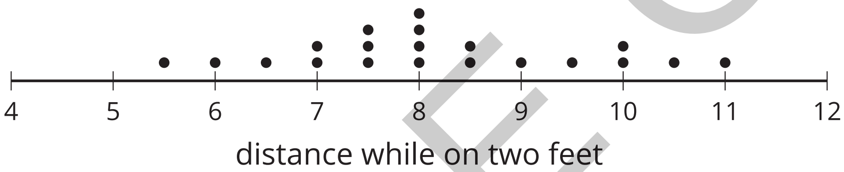
Analyze your 2 data sets to compare the distributions from the statistical questions. Next, create a visual display that includes:

- Your statistical question.
- The data that you collected.
- A data display.
- The measure of center and variability that you found that are appropriate for the data.
- An answer to the statistical question with any supporting mathematical work.

Student Response

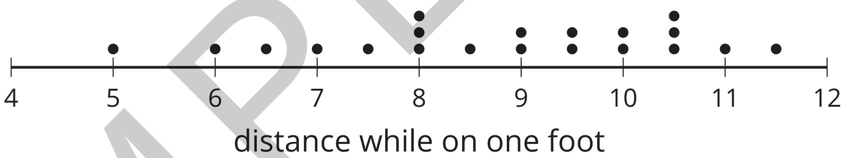
Sample response:

Condition A: standing on two feet 5.5, 6, 6.5, 7, 7, 7.5, 7.5, 7.5, 8, 8, 8, 8, 8.5, 8.5, 9, 9.5, 10, 10, 10.5, 11



mean: 8.18, standard deviation: 1.44

Condition B: standing on one foot 5, 6, 6.5, 7, 7.5, 8, 8, 8, 8.5, 9, 9, 9.5, 9.5, 10, 10, 10.5, 10.5, 10.5, 11, 11.5



mean: 8.78, standard deviation: 1.72

The reaction time was faster while standing on two feet. I know this because the mean distance dropped was 8.18 inches compared to 8.78 inches while standing on one foot. The greater distance means a slower reaction time. I also noticed that there was greater variability in the data collected while standing on one foot. I knew this by looking at the standard deviation. It was 1.72 inches for the one foot data and 1.44 inches for the two feet data. Interestingly enough, the shortest distance was for the one foot data, but so was the greatest distance. I chose to use the mean and the standard deviation because the shape of both distributions was approximately symmetric.

Activity Synthesis

Select groups to share their visual displays. Encourage students to ask questions about the mathematical thinking or design approach that went into creating the display. For students who had an error in their display, ask “What error do you see in the display and how would you resolve it?” (Sample responses:

- I noticed that they used the mean instead of the median. I would resolve it by using the median since there was an outlier
- I noticed that the IQR was calculated incorrectly. I would have used technology to verify my statistics.)

Here are questions for discussion, if not already mentioned by students:

- "Once you collected your data, how did you answer your statistical question?" (I used two different dot plots to display my data and determined that the mean was the most appropriate measure of center because my dot plots were approximately symmetric. I then calculated the standard deviation using technology because standard deviation is appropriate to use with the mean. I then compared the mean and the standard deviation for the two sets of data and determined that standing on two feet gave my partner a slightly faster reaction time and showed less variability than standing on one foot did.)
- "How did you choose which measure of center and which measure of variability to use?" (I used the distribution shape to determine the measure of center. My data distribution was skewed, so I used the median. I used the IQR as a measure of variability because it is based on the median. If I had used the mean, I would have used the standard deviation.)
- "Using the context of the two treatments, what did the measure of variability tell you?" (The measure of variability told me how spread apart the results were from the measure of center I chose. It let me know how consistent the reaction time was for each treatment.)
- "Imagine that you collected data for the same treatments from all the students in the class. How would this change how you displayed or analyzed your data?" (There would be a lot of data. I would probably have to use technology to find the statistics and to create the data display. I think that I might need to use a histogram to represent the distribution of the data.)

Access for English Language Learners

MLR1 Stronger and Clearer Each Time. Before the whole-class discussion, give students time to meet with 2–3 partners to share and get feedback on their first draft response to their statistical question. Invite listeners to ask questions and give feedback that will help their partner clarify and strengthen their ideas and writing. Give students 3–5 minutes to revise their first draft based on the feedback they receive.
Advances: Writing, Speaking, Listening

Sec E

16.3 Heights and Handedness

Optional

10 mins

Activity Narrative

In this activity, students use a large data set to compare the size of students with different dominant hands. Students must choose the appropriate tools (MP5) to analyze a large data set such as this. Additional variables are provided if students wish to extend their analysis to find any other connections, although bivariate data are the focus of a future unit in this course.

Standards

Addressing HSS-ID.A

Instructional Routines

- Aspects of Mathematical Modeling

Launch

Arrange students in groups of 2. Distribute a copy of the blackline master to each group.

Access for Students with Disabilities

Representation: Internalize Comprehension. Activate or supply background knowledge. Engage in a classroom discussion drawing attention to the relationship between height, foot length, and arm span of right- and left-handed students in the class. This will help students make connections between the large data set and specific students in the class.
Supports accessibility for: Conceptual Processing, Memory

Student Task Statement

Is there a connection between a student's dominant hand and their height? Use the table of information to compare the size of students with different dominant hands.

Student Response

Sample response:

Using the data as given, left-handed students have a mean height of 164.86 cm and a standard deviation of 25.18 cm, while right-handed students have a mean height of 153.84 cm and a standard deviation of 41.67 cm. While the mean height of left-handed students is greater, the large standard deviation indicates that there is not a great difference between the groups.

Looking closer at the data, though, there are some outliers that do not make sense. In particular, it is unlikely that there would be any students who are less than 100 centimeters (3 feet 3 inches) tall. These students likely misinterpreted the question as being in terms of inches or entered incorrect data. Removing any values less than 100 from the height information changes the mean and standard deviation. The mean height for left-handed students becomes 169.79 cm with a standard deviation of 13.86 cm. For right-handed students, the new mean height is 168.60 cm with a standard deviation of 10.63 cm.

Using the updated means, the two groups are much closer and still have a relatively large standard deviation, so there does not seem to be a great difference between the two groups of students.

Activity Synthesis

Select students to share their analysis of the data.

Lesson Synthesis

The purpose of this discussion is to help students reflect on what they learned about data collection, data analysis, and answering a statistical question. Here are some questions for discussion.

- "What did you find the most challenging about this lesson?" (It was really difficult to figure out what to do with the data that I collected because there were no directions. I had to look back to my statistical question and think about what tools I would need to use to answer it.)
- "What did you find interesting about this lesson?" (It was really interesting that I could actually use the statistics I

learned about to answer a question. It seemed a lot like when we do experiments in science class.)

- “What mathematics do you need to know more about?” (I am still struggling with figuring out the shape of the data distribution. Sometimes I think the data distribution is roughly symmetric but it actually is skewed. I would like more practice about describing the shape of distributions.)

Practice Problems

1

from Unit 1, Lesson 15

Student Task Statement

Here are the statistics for the high temperatures in a city during October:

- mean of 65.3 degrees Fahrenheit
- median of 63.5 degrees Fahrenheit
- standard deviation of 9.3 degrees Fahrenheit
- IQR of 7.1 degrees Fahrenheit

Recall that the temperature C, measured in degrees Celsius, is related to the temperature F, measured in degrees Fahrenheit, by $C = \frac{5}{9}(F - 32)$.

a. Describe how the value of each statistic changes when 32 is subtracted from the temperature in degrees Fahrenheit.

b. Describe how the value of each statistic further changes when the new values are multiplied by $\frac{5}{9}$.

c. Describe how to find the value of each statistic when the temperature is measured in degrees Celsius.

Solution

a. Subtracting 32 from the temperature in degrees Fahrenheit will decrease the mean and median by 32 but it will not change the standard deviation or the IQR.

b. Multiplying the new values (which have been shifted by 32) by $\frac{5}{9}$ will multiply the mean and the median by $\frac{5}{9}$. Both the standard deviation and the IQR are multiplied by $\frac{5}{9}$.

c. For the mean and median, subtract 32 from the given values and then multiply by $\frac{5}{9}$. For the standard deviation and IQR, multiply the given values by $\frac{5}{9}$.

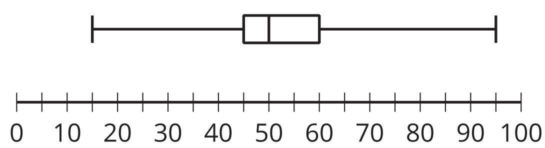
Sec E

2

from Unit 1, Lesson 15

Student Task Statement

Here is a box plot.

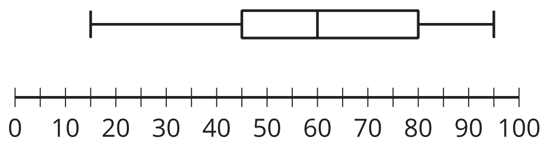


Give an example of a box plot that has a greater median and a greater measure of variability, but the same minimum and maximum values.

Solution

The minimum should be 15, the maximum 95, the median greater than 50, and the interquartile range larger than 15.

Sample response:



3 from Unit 1, Lesson 14

Student Task Statement

The mean vitamin C level for 20 dogs was 7.6 milligrams per liter, with a standard deviation of 2.1 milligrams per liter.

One dog's vitamin C level was not in the normal range. It was 0.9 milligrams per liter, which is a very low level of vitamin C.

a. If the value 0.9 is eliminated from the data set, does the mean increase or decrease?

b. If the value 0.9 is eliminated from the data set, does the standard deviation increase or decrease?

Solution

a. The mean will increase because 0.9 milligrams per liter is below the mean of 7.6 milligrams per liter.

b. The standard deviation will decrease because 0.9 milligrams per liter is very far from the mean, so the new data set will be more concentrated around the center when that value is removed.

Sec E

4 from Unit 1, Lesson 14

Student Task Statement

The data set represents the number of hours that fifteen students walked during a two-week period.

6 6 7 8 8 8 9 10 10 12 13 14 15 16 30

The median is 10 hours, Q1 is 8, Q3 is 14, and the IQR is 6 hours. Are there any outliers in the data? Explain or show your reasoning.

Solution

Yes, 30 is an outlier since $14 + 1.5 \cdot 6 = 23$ and $30 > 23$.

5 from Unit 1, Lesson 14

Student Task Statement

Here are some summary statistics about the number of accounts that follow some bands on social media.

- mean: 15,976 followers
- median: 16,432 followers
- standard deviation: 3,279 followers
- Q1: 13,796
- Q3: 19,070
- IQR: 5,274 followers

a. Give an example of a number of followers that a very popular band might have that would be considered an outlier for this data. Explain or show your reasoning.

b. Give an example of a number of followers that a relatively unknown band might have that would be considered an outlier for this data. Explain or show your reasoning.

Solution

a. Sample response: 30,000 followers since it is an outlier if the number of followers is greater than $19{,}070 + 1.5 \cdot 5{,}274$, or 26,981 followers.

b. Sample response: 5,000 followers since it is an outlier if the number of followers is less than $13{,}796 - 1.5 \cdot 5{,}274$, or 5,885 followers.

Sec E

6 from Unit 1, Lesson 13

Student Task Statement

The weights of one population of brown bears have a mean of 428 pounds and standard deviation of 36 pounds. The weights of another population of brown bears have a mean of 397 pounds and standard deviation of 25 pounds. Andre says the two populations are similar. Do you agree? Explain your reasoning.

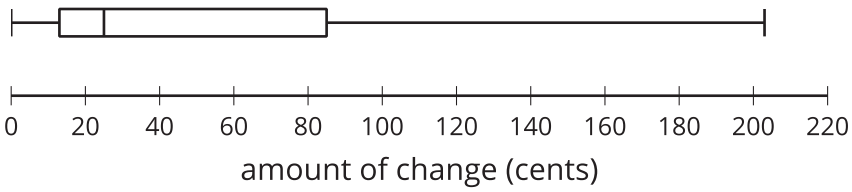
Solution

Sample response: I agree with Andre. Even though the weights of the two populations have different means, the standard deviation lets me know that many brown bears in the first population weigh less than 428 pounds, and many brown bears in the second population weigh more than 397 pounds. The populations have weights that overlap, so I think they are similar.

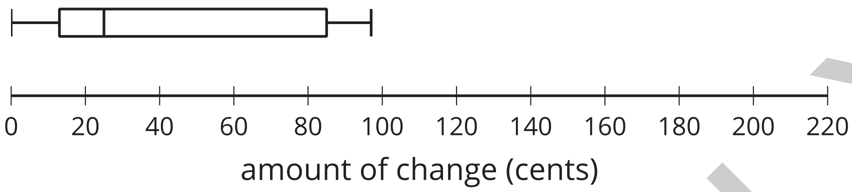
7 from Unit 1, Lesson 10

Student Task Statement

The box plot represents the distribution of the amount of change, in cents, that 50 people were carrying when surveyed.



The box plot represents the distribution of the same data set, but with the maximum, 203, removed.



The median is 25 cents for both plots. After examining the data, the value 203 is removed because it was an error in recording.

a. Explain why the median remains the same when 203 cents was removed from the data set.

b. When 203 cents is removed from the data set, does the mean remain the same? Explain your reasoning.

Solution

a. The median remains the same because removing an extreme value from a data set tends not to have much effect or no effect on the median. In this case, there may be multiple people carrying 25 cents.

b. The mean decreases because 203 cents is greater than the mean of the data set.

Learning Targets

Lesson 1 Getting to Know You

- I can tell statistical questions from non-statistical questions and can explain the difference.
- I can tell the difference between numerical and categorical data.

Lesson 2 Data Representations

- I can find the five-number summary for data.
- I can use a dot plot, histogram, or box plot to represent data.

Lesson 3 A Gallery of Data

- I can graphically represent the data that I collected and critique the representations of others' data.

Lesson 4 The Shape of Distributions

- I can use a graphical representation of data to suggest a situation that produced the data pictured.
- I can use the terms "symmetric," "skewed," "uniform," "bimodal," and "bell-shaped" to describe the shape of a distribution.

Lesson 5 Calculating Measures of Center and Variability

- I can calculate mean absolute deviation, interquartile range, mean, and median for a set of data.

Lesson 6 Mystery Computations

- I can determine basic relationships between cell values in a spreadsheet by changing the values and noticing what happens in another cell.

Lesson 7 Spreadsheet Computations

- I can use a spreadsheet as a calculator to find solutions to word problems.

Lesson 8 Spreadsheet Shortcuts

- I can use shortcuts to fill in cells on a spreadsheet.

Lesson 9 Technological Graphing

- I can use technology to create graphic representations of data and calculate statistics.

Lesson 10 The Effect of Extremes

- I can describe how an extreme value will affect the mean and median.
- I can use the shape of a distribution to compare the mean and median.

Lesson 11 Comparing and Contrasting Data Distributions

- I can arrange data sets in order of variability, given graphic representations.

Lesson 12 Standard Deviation

- I can describe standard deviation as a measure of variability.
- I can use technology to compute standard deviation.

Lesson 13 More Standard Deviation

- I can use standard deviation to say something about a situation.

Lesson 14 Outliers

- I can find values that are outliers, investigate their source, and figure out what to do with them.
- I can tell how an outlier affects mean, median, IQR, or standard deviation.

Lesson 15 Comparing Data Sets

- I can compare and contrast situations using measures of center and measures of variability.

Lesson 16 Analyzing Data

- I can collect data from an experiment and compare the results using measures of center and measures of variability.