Plan for Algebra 2 Unit 7: Statistical Inferences

Relevant Unit(s) to review: Algebra 1 Unit 1: One-variable Statistics

Essential prior concepts to engage with this unit	• data representations including dot plots and histograms • standard deviation
Brief narrative of approach	This unit compares methods of collecting and interpreting data. If the Check Your Readiness shows that students do not remember how to create a histogram it will be necessary to add activities from a previous course to teach that content. If the Check Your Readiness shows that students are not familiar with standard deviation it will be necessary to add activities from a previous course to teach that content as well.

Lessons to Add	Lessons to Remove or Modify
1. Algebra 1 Unit 1 Lesson 2 2. Algebra 1 Unit 1 Lesson 12	1. Remove Alg2.7.15: This lesson consists primarily of an Info Gap that reinforces analyzing data from an experiment. 2. Remove Alg2.7.16: Students complete the entire cycle of collecting and analyzing data for additional practice.
Lessons added: 2	Lessons removed: 2

Modified Plan for Algebra 2 Unit 7

Day	IM lesson	Notes
	Alg2.7 Check Your Readiness assessment	Note that the Check Your Readiness assessment includes item-by-item guidance to inform just-in-time adjustments to instruction within the lessons in Alg2.7.

1	Alg2.7.1	
2	Alg2.7.2	
3	Alg2.7.3	
4	Alg1.1.2	If the initial assessment shows that students are not familiar with histograms, include this lesson before continuing with grade-level content.
5	Alg1.1.12	If the initial assessment shows that students are not familiar with standard deviation, include this lesson before continuing with grade-level content.
6	Alg2.7.4	
7	Alg2.7.5	
8	Alg2.7.6	
9	Alg2.7.7	
10	Alg2.7.8	
11	Alg2.7.9	
12	Alg2.7 Mid Assessment	
13	Alg2.7.10	
14	Alg2.7.11	
15	Alg2.7.12	
16	Alg2.7.13	
17	Alg2.7.14	
18	Alg2.7 End Assessment	

Priority and Category List for Lessons

High priority (+), Medium priority (0), Low priority (-)

E: Explore, Play, and Discuss, D: Deep Dive, A: Synthesize and Apply

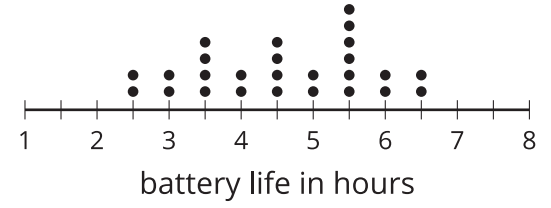
Lesson	Priority (+, 0, -)	Category (E, D, A)	Notes
Alg2.7.1	+	E	This lesson introduces sample surveys, experiments, and observational studies.
Alg2.7.2	0	E	This lesson expands on the previous lesson to include the difference between association and causation.
Alg2.7.3	+	D	This lesson defines random selection and studies the importance of random selection for samples.
Alg2.7.4	-	A	This lesson reviews distribution, mean, median, and standard deviation from prior grade levels.
Alg2.7.5	+	D	This lesson introduces the normal distribution.
Alg2.7.6	0	A	This lesson expands on the previous lesson to include using the area under a normal curve to estimate percentiles.
Alg2.7.7	0	D	This lesson expands on the previous lesson to include interpreting the area as a proportion of values in the interval.
Alg2.7.8	0	E	This lesson explores how to generate data and use it to support or oppose mathematical claims based on a model.
Alg2.7.9	+	D	This lesson introduces the margin of error.
Alg2.7.10	0	A	This lesson offers an opportunity to practice using data to understand the margin of error.

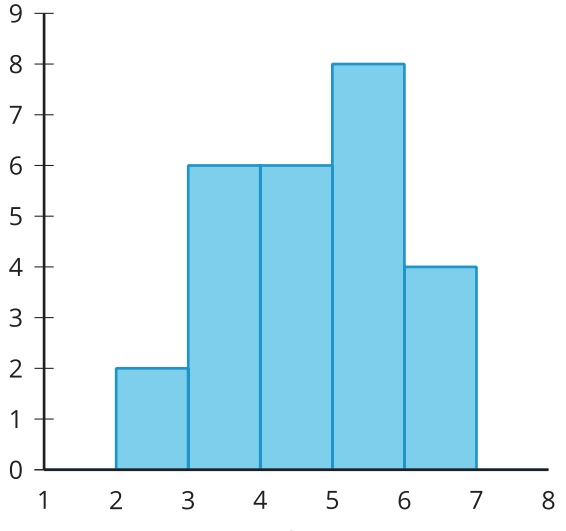
Alg2.7.11	0	D	This lesson analyzes the relationship between sample size and margin of error.
Alg2.7.12	0	A	This lesson offers an opportunity to practice estimating the population mean with an associated margin of error using sample means from random samples.
Alg2.7.13	+	E	This lesson introduces the terms treatment and randomization distribution.
Alg2.7.14	0	D	This lesson analyzes whether a difference in means for experimental groups is likely due to the way the subjects were grouped or likely due to the treatment.
Alg2.7.15	-	A	This lesson consists primarily of an info gap that reinforces analyzing data from an experiment.
Alg2.7.16	-	A	This lesson offers an opportunity to complete the entire cycle of collecting and analyzing data.

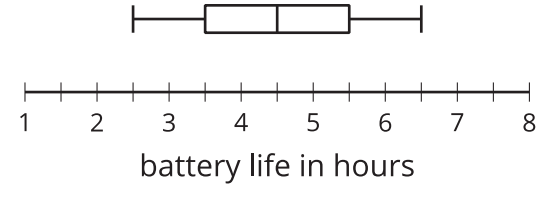
Lesson 2: Data Representations

2.1: Notice and Wonder: Battery Life

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







2.2: Tomato Plants: Histogram

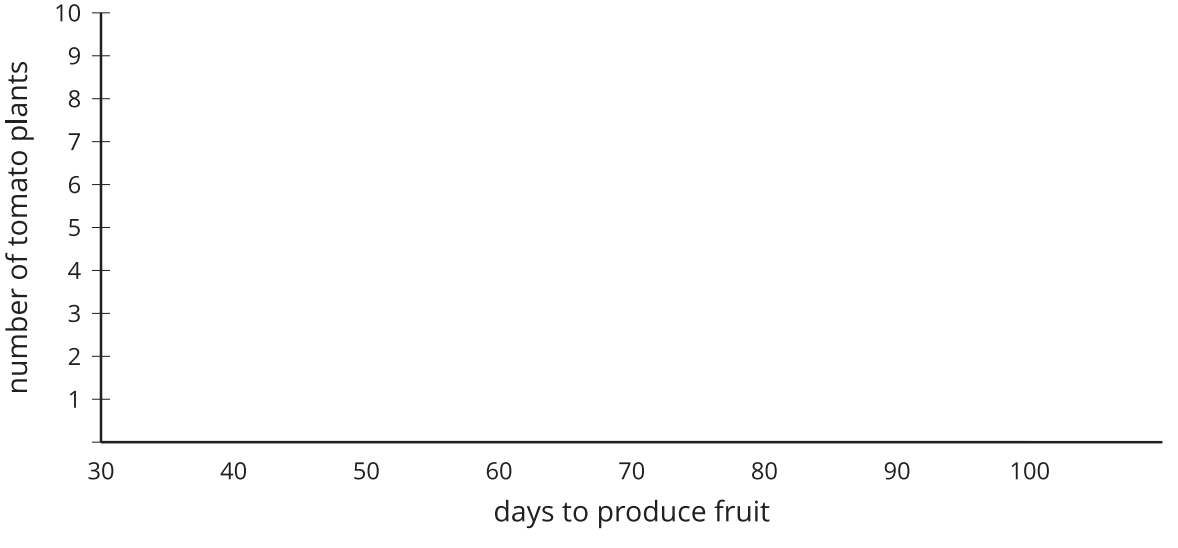
A histogram can be used to represent the distribution of numerical data.

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.

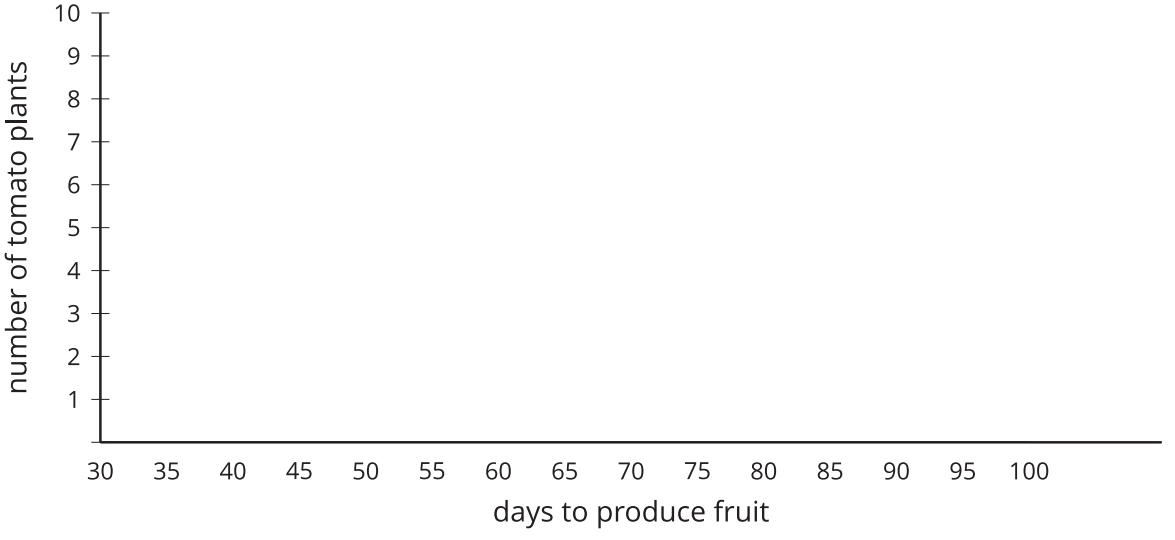
47	52	53	55	57
60	61	62	63	65
65	65	65	68	70
72	72	75	75	75
76	77	78	80	81
82	85	88	89	90

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 lengths to figure out which gives the best sense of the shape of the distribution.

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

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

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

2.3: Tomato Plants: Box Plot

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

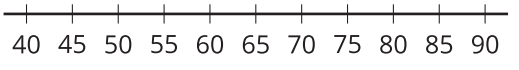
minimum	Q1	median	Q3	maximum

1. Using the same data as 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.

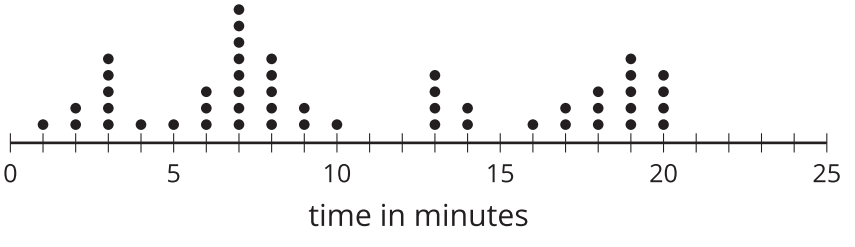


time in days

Lesson 2 Summary

The table shows a list of the number of minutes people could intensely focus on a task before needing a break. 50 people of different ages are represented.

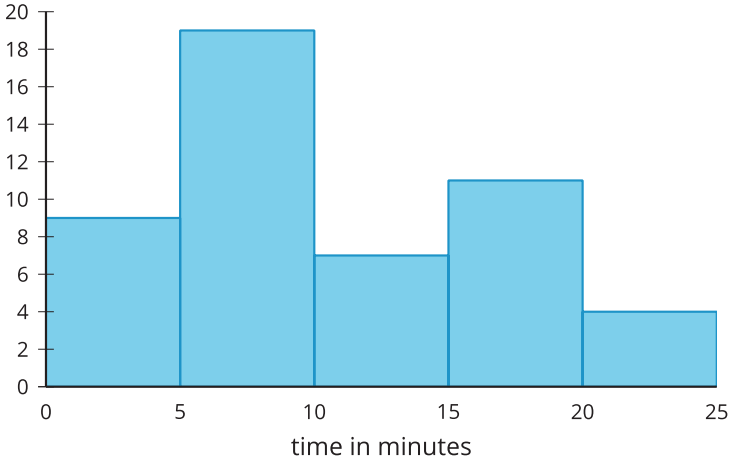
19	7	1	16	20	2	7	19
9	13	3	9	18	13	20	8
3	14	13	2	8	5	17	7
18	17	8	8	7	6	2	20
7	7	10	7	6	19	3	18
8	19	7	13	20	14	6	3
19	4						



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.

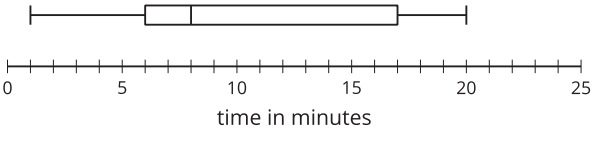
There were quite a few people that lost focus at around 3, 7, 13, and 19 minutes and nobody lost focus at 11, 12, or 15 minutes. Dot plots are useful when the data set is not too large and shows all of the individual values in the data set. In this example, a dot plot can easily show all the data. If 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 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.

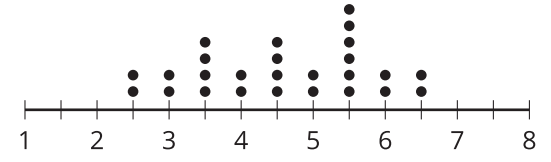


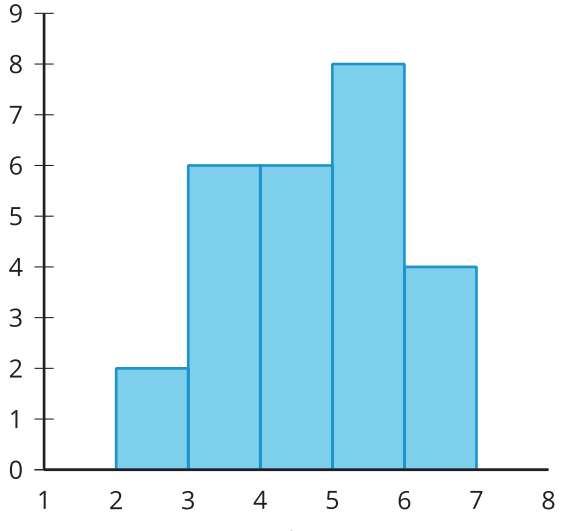
Box plots are created using the **five-number summary**. For a set of data, the five-number summary consists of these five statistics: the minimum value, the first quartile, the median, the third quartile, and the maximum value. These values split the data into four sections each representing approximately one-fourth of the data. The median of this data is indicated at 8 minutes and about 25% of the data falls in the short second quarter of the data between 6 and 8 minutes. Similarly, approximately one-fourth of the data is 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 3 different ways to look at the shape and distribution while highlighting different aspects of the data.

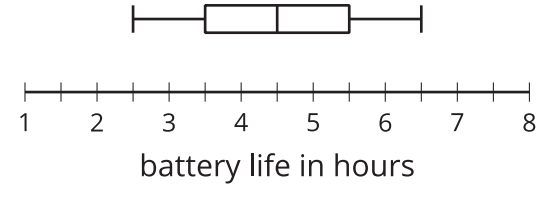
Lesson 2: Data Representations

Cool Down: Reasoning About Representations

The dot plot, histogram, and box plot represent the distribution of the same data in 3 different ways.







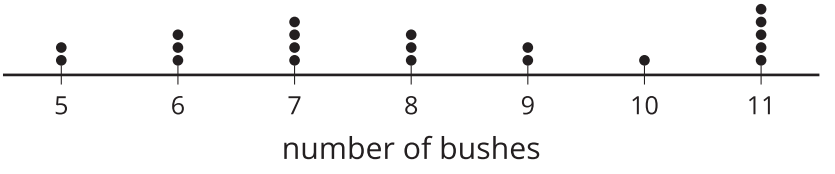
1. What information can be seen most easily in the dot plot?

2. What information can be seen most easily in the histogram?

3. What information can be seen most easily in the box plot?

Unit 1 Lesson 2 Cumulative 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. 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}

   (From Unit 1, Lesson 1.)

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

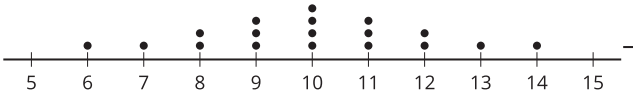
   (From Unit 1, Lesson 1.)

Lesson 12: Standard Deviation

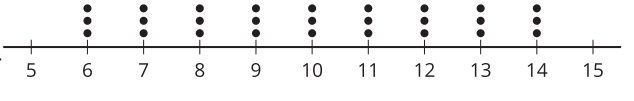
12.1: Notice and Wonder: Measuring Variability

What do you notice? What do you wonder?

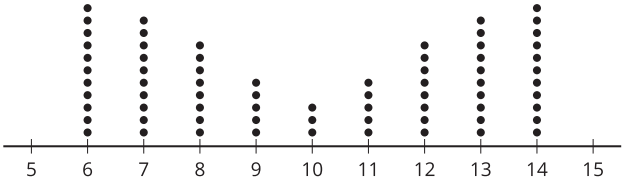
mean: 10, MAD: 1.56, standard deviation: 2



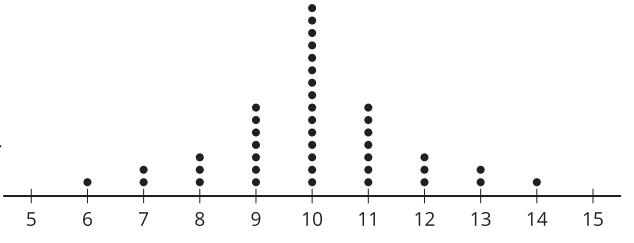
mean: 10, MAD: 2.22, standard deviation: 2.58



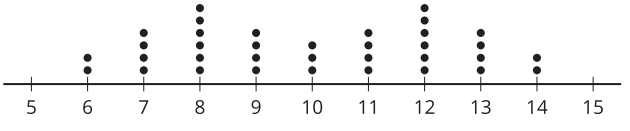
mean: 10, MAD: 2.68, standard deviation: 2.92



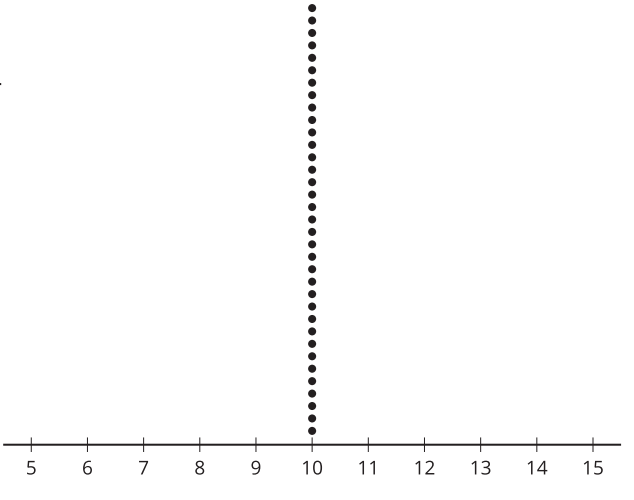
mean: 10, MAD: 1.12, standard deviation: 1.61



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



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



12.2: Investigating Standard Deviation

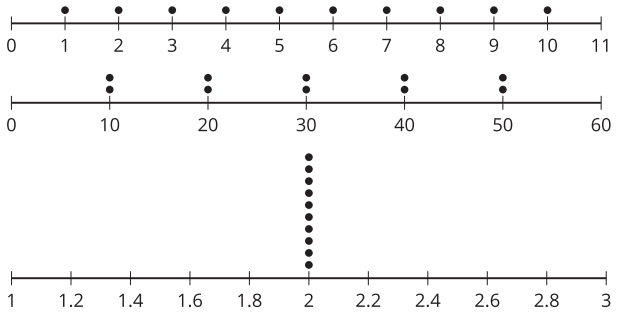
Use technology to find the mean and the standard deviation for the data in the dot plots.

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

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

Partner 1

Dot plots:

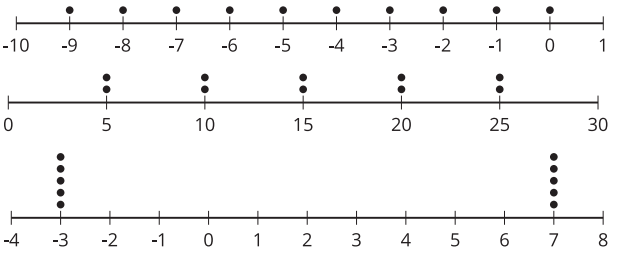


Conditions:

- 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.

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.

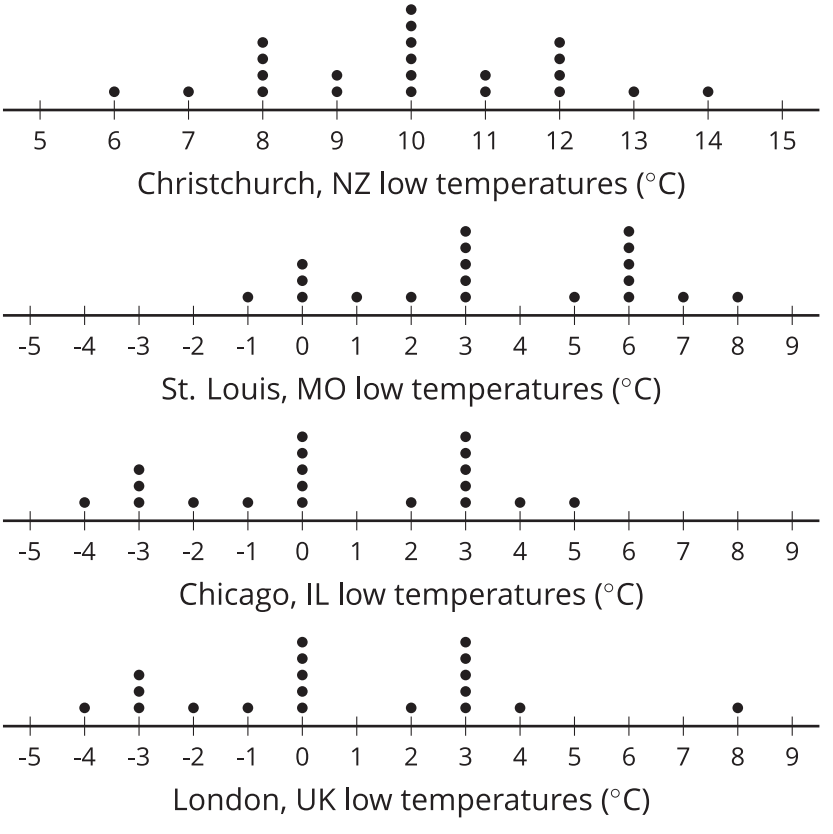
A deeper understanding of the importance of standard deviation as a measure of variability will come with a deeper study of statistics. For now, know that standard deviation is mathematically important and will be used as the appropriate measure of variability when mean is an appropriate measure of center.

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.

Lesson 12: Standard Deviation

Cool Down: True or False: Reasoning with Standard Deviation

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 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.

Unit 1 Lesson 12 Cumulative 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?

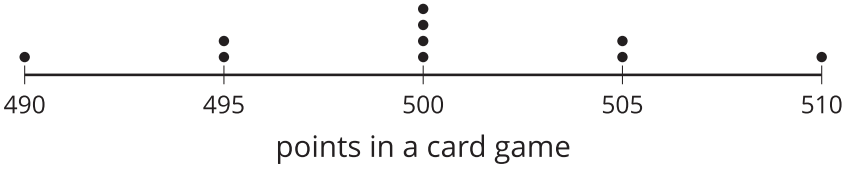
2. Here is a data set:

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. 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?

(From Unit 1, Lesson 11.)

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

   A. bell-shaped

   B. bimodal

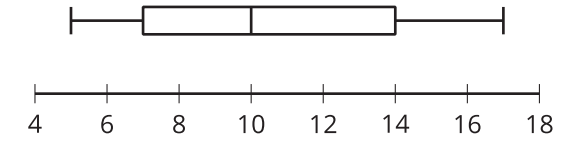
   C. skewed

   D. symmetric

   E. uniform

(From Unit 1, Lesson 10.)

6. What is the IQR?



A. 5 branches

B. 7 branches

C. 10 branches

D. 12 branches

(From Unit 1, Lesson 11.)

7. 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.

(From Unit 1, Lesson 9.)