



Lesson 13: Describing Variability Using the Interquartile Range (IQR)

Student Outcomes

- Given a set of data, students describe how the data might have been collected.
- Students describe the unit of measurement for observations in a data set.
- Students calculate the median of the data.
- Students describe the variability in the data by calculating the interquartile range.

Lesson Notes

Students should develop understanding of statistical variability, in particular recognizing that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how the values in the data set vary.

Students are also expected to summarize numerical data sets in relation to their context. This is done by reporting the number of observations and describing the nature of the attribute under investigation. Students indicate how data were measured and the units of measurement. They provide quantitative measures of center (median) and variability (interquartile range), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

In this lesson students are engaged in making sense of problems and solving them. As the contexts change, they are asked to pay attention to the units for each. They are also engaged in working in groups, sharing their reasoning, and critiquing the reasoning of others as they create contexts that satisfy given constraints.

Classwork

The median was used to describe the typical value of our data in Lesson 12. Clearly, not all of the data is described by the value. How do we find a description of how the data vary? What is a good way to indicate how the data vary when we use a median as our typical value? These questions are developed in the following exercises.

Exercises 1–4 (15 minutes): More French Fries

MP.1

This exercise returns to the data from Lesson 12, raising questions about how the data might have been collected and whether any bias (the formal word is not used) might be inherent in the process. Students examine work from Lesson 12 finding the medians of the lower half and upper half of the data and use them to calculate the interquartile range for the three data sets.

Be sure that students understand that a quartile is the cut off point for the median of either the top half or bottom half of the data and not the length of the segment from the minimum (maximum) to the quartile. They should be able to

approximate the number of elements in each section in terms of $\frac{1}{4}$ or 25% of the data or by giving an estimate of the actual number data values as well as knowing that $\frac{1}{2}$ or 50% of the data values are between the two quartiles. They should also recognize that the IQR is a measure of spread around the median (the length of the interval that captures the middle 50% of the data).

Consider the following questions as you discuss the example questions with students:

- Approximately how many data values would be between the quartiles? Explain your reasoning.
 - *Answers will vary depending on the data.*
- What other measures of center and spread have you studied and how do you think they would compare to the median and IQR? (If you want to pursue this question, you could have students compute the mean and MAD for one of the data sets.)
 - *This question provides students an opportunity to discuss what they might recall about the mean as a center and the MAD as a measure of variability.*

Data from Lesson 12: Number of french fries

Restaurant A: 80, 72, 77, 80, 90, 85, 93, 79, 84, 73, 87, 67, 80, 86, 92, 88, 86, 88, 66, 77

Restaurant B: 83, 83, 83, 84, 79, 78, 80, 81, 83, 80, 79, 81, 84, 82, 85, 85, 79, 79, 83

Restaurant C: 75, 75, 77, 85, 85, 80, 80, 80, 80, 81, 82, 84, 84, 84, 85, 77, 77, 86, 78, 78, 79, 79, 79, 79

Exercises 1–4

1. In Lesson 12, you thought about the claim made by a chain restaurant that the typical number of french fries in a large bag was 82. Then you looked at data on the number of fries in a bag from three of the restaurants.
 - a. How do you think the data was collected and what problems might have come up in collecting the data?

Answers will vary: They probably went to the restaurants and ordered a bunch of bags of french fries. Sometimes the fries are broken, so they might have to figure out what to do with those - either count them as a whole, discard them, or put them together to make whole fries.
 - b. What scenario(s) would give counts that might not be representative of typical bags?

Answers will vary: Different workers might put different amounts in a bag, so if you took the sample at lunch, you might have different numbers than if you did it in the evening. The restaurants might weigh the bags to see that the weight was constant despite the size of the fries, so you could have the same weight of fries even though you had different counts for the bags.
2. In Exercise 7 of Lesson 12 you found the median of the top half and the median of the bottom half of the counts for each of the three restaurants. These were the numbers you found: Restaurant A – 87.5 and 77; Restaurant B – 82 and 79; Restaurant C – 84 and 78. The difference between the medians of the two halves is called the interquartile range or IQR.
 - a. What is the IQR for each of the three restaurants?

Restaurant A is 10.5; Restaurant B is 3; Restaurant C is 6.
 - b. Which of the restaurants had the smallest IQR, and what does that tell you?

Restaurant B had the smallest IQR. This indicates that the spread around the median number of fries is smaller than for either of the other two restaurants. About half of the data are within 3 fries of the median, so the median is a pretty good estimate of what is typical.

c. About what fraction of the counts would be between the quartiles? Explain your thinking.

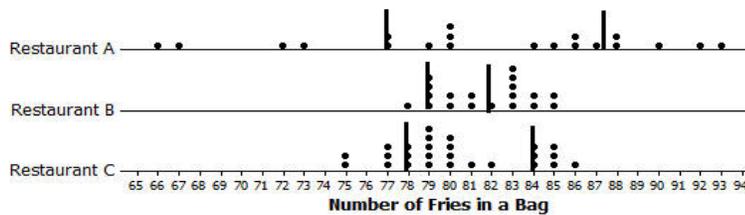
About $\frac{1}{2}$ or 50% of the counts would be between the quartiles because about $\frac{1}{4}$ of the counts are between the median and the lower quartile, and $\frac{1}{4}$ of the counts are between the median and upper quartile.

3. The medians of the lower and upper half of a data set are called quartiles. The median of the top half of the data is called the upper quartile; the median of the bottom half of the data is called the lower quartile. Do these names make sense? Why or why not?

Answers will vary: Students might say that quartile is related to quarter, and the lower quartile, the median, and the upper quartile divide the data into four sections with one fourth or a quarter of the data values in each section.

4.

a. Mark the quartiles for each restaurant on the graphs below.



b. Does the IQR help you decide which of the three restaurants seems most likely to really have 82 fries in a typical bag? Explain your thinking.

Restaurant B has the smallest IQR, which means that the middle half of the counts of the number of fries in a bag is really close to the median. Restaurant B also has the smallest range.

Example 1 (5 minutes): Finding the IQR

The example is intended as a reference for students and not to be reproduced during class unless students need more clarity. If that is the case, have students explain the diagram to each other rather than recreating it on the board. Make sure that students understand that the lower quartile is the number marking the median of the lower half.

Example 1: Finding the IQR

Read through the following steps. If something does not make sense to you, make a note and raise it during class discussion. Consider the data: 1, 1, 3, 4, 6, 6, 7, 8, 10, 11, 11, 12, 15, 15, 17, 17, 17

Creating an IQR:

I. Order the data: The data is already ordered.

1, 1, 3, 4, 6, 6, 7, 8, 10, 11, 11, 12, 15, 15, 17, 17, 17

II. Find the minimum and maximum: The minimum data point is 1, and the maximum is 17.

① 1, 3, 4, 6, 6, 7, 8, 10, 11, 11, 12, 15, 15, 17, 17, ① 17

Exercise 6 (15–20 minutes): On Your Own with IQRs

Students should work together in pairs or groups of three on this exercise. After they develop three examples, encourage them to select one of their examples and explain it more fully by creating a simple poster. If a poster is used for this exercise, indicate to students that they should do the following:

- Draw or write out a specific context for each example that explains the data and how it would be collected.
- Explain on the poster how to find the median, the upper quartile, and the lower quartile.
- Explain what the IQR would mean for the context of the selected example.

Either discuss student examples or design a presentation of the posters.

Encourage students to use their imaginations and identify contexts for which the IQR might provide useful information. If time permits, have students explain their poster to other students. Display posters as possible examples of problems for future lessons or discussions.

6. Create three different contexts for which a set of data collected related to those contexts could have an IQR of 20. Define a median for each context. Be specific about how the data might have been collected and the units involved. Be ready to describe what the median and IQR mean in each case.

Examples that could be included in this exercise are as follows: number of books read by students during a school year (some students read a lot of books, while other students may not read as many), number of movies viewed at a theater during the last year by students in a class, number of text messages students receive during a specific day (for example, on Monday), number of commercials on TV during a specific time period that are about buying a car, number of different states students have visited, number of healthy trees on certain blocks of a city, and number of students in each classroom of a school during a specific time period. Remind students that the goal is to have them think of data that if collected might have an IQR of approximately 20. These ideas also allow students to start thinking of the process of actually collecting data that is needed later.

Closing (2–3 minutes)**Lesson Summary**

One of our goals in statistics is to summarize a whole set of data in a short concise way. We do this by thinking about some measure of what is typical and how the data are spread relative to what is typical.

In earlier lessons, you learned about the MAD as a way to measure the spread of data about the mean. In this lesson, you learned about the IQR as a way to measure the spread of data around the median.

To find the IQR, you order the data, find the median of the data, and then find the median of the lower half of the data (the lower quartile) and the median of the upper half of the data (the upper quartile). The IQR is the difference between the upper quartile and the lower quartile, which is the length of the interval that includes the middle half of the data, because the median and the two quartiles divide the data into four sections, with about $\frac{1}{4}$ of the data in each section. Two of the sections are between the quartiles, so the interval between the quartiles would contain about 50% of the data.

Small IQRs indicate that the middle half of the data are close to the median; a larger IQR would indicate that the middle half of the data is spread over a wider interval relative to the median.

Exit Ticket (5 minutes)



Name _____

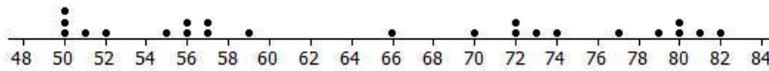
Date _____

Lesson 13: Describing Variability Using the Interquartile Range (IQR)

Exit Ticket

1. On the graph below, insert the following words in approximately the correct position.

Maximum Minimum IQR Median Lower Quartile (Q1) Upper Quartile (Q3)



2. Estimate the IQR based on the data set above.

Exit Ticket Sample Solutions

1. On the graph below, insert the following words in approximately the correct position.

Maximum Minimum IQR Median Lower Quartile (Q1) Upper Quartile (Q3)

2. Estimate the IQR based on the data set above.

The IQR is approximately 22.

Problem Set Sample Solutions

1. The average monthly high temperatures (in °F) for St. Louis and San Francisco are given in the table below.

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
St. Louis	40	45	55	67	77	85	89	88	81	69	56	43
San Francisco	57	60	62	63	64	67	67	68	70	69	63	57

Data Source: www.weather.com/weather/wxclimatology/monthly/graph/USCA0987
www.weather.com/weather/wxclimatology/monthly/graph/USMO0787

a. How do you think the data might have been collected?

Someone at a park or the airport or someplace probably records the temperature every hour of every day and then takes all of the highest ones and finds the mean.

b. Do you think it would be possible for $\frac{1}{4}$ of the temperatures in the month of July for St. Louis to be 95° or above? Why or why not?

Yes, it is possible. The mean temperature in St. Louis for July is 89°. There are 31 days in July, so $\frac{1}{4}$ of the days would be about 8 days. If the temperature was 95° for 5 days, 100° for 3 days, and 87° for all of the rest of the days, it would work.

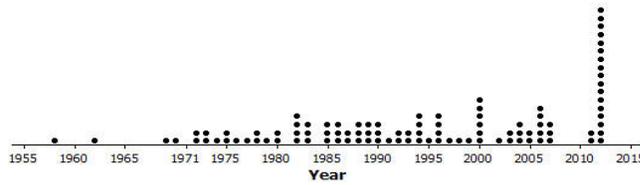
c. Make a prediction about how the sizes of the IQR for the temperatures for each city compare. Explain your thinking.

San Francisco probably has the smaller IQR because those temperatures don't seem to vary as much as the St. Louis temperatures.

d. Find the IQR for the average monthly high temperature for each city. How do the results compare to your conjecture?

For San Francisco the IQR is 6.5°; St. Louis is 33°.

2. The plot below shows the years in which each of 100 pennies were made.



a. What does the stack of 17 dots at 2012 representing 17 pennies tell you about the “age” of the pennies in 2014?

17 pennies were made in 2012, and they would be 2 years old in 2014.

b. Here is some information about the sample of pennies. The mean year they were made is 1994; the first year any of the pennies were made was 1958; the newest pennies were made in 2012; Q1 is 1984, the median is 1994, and Q3 is 2006; the MAD is 11.5 years. Use the information to indicate the years in which the middle half of the pennies was made.

In this case, the IQR is 22 years and the mean ± the MAD gives an interval of 23, so the middle half of the pennies were made over an interval of 22 years.

3. Create a data set with at least 6 elements such that it has the following:

a. A small IQR and a big range (maximum-minimum).

Answers will vary: {0, 100, 50, 50, 50, 50} where the range is 100 and the IQR is 0.

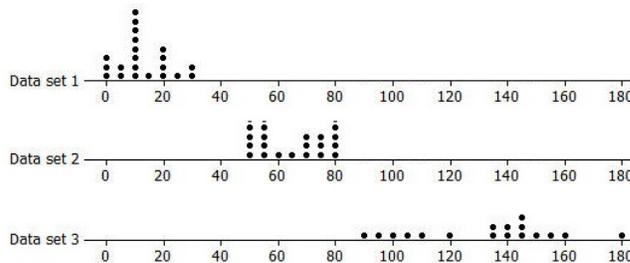
b. An IQR equal to the range.

Answers will vary: {10, 10, 10, 15, 20, 20, 20}

c. The lower quartile is the same as the median.

Answers will vary: {1, 1, 1, 1, 1, 5, 6, 7}

4. Rank the following three data sets by the value of the IQR.



Data set 1 has the smallest IQR at about 14, data set 2 the next smallest at about 22, and data set 3 the largest at about 41. (Be sure students do not confuse data set with the larger quartiles as having the larger IQR.)

5. Here are the counts of the fries in each of the bags from Restaurant A:

80, 72, 77, 80, 90, 85, 93, 79, 84, 73, 87, 67, 80, 86, 92, 88, 86, 88, 66, and 77.

a. Suppose one bag of fries had been overlooked in the sample and that bag had only 50 fries. Would the IQR change? Explain your reasoning.

The IQR would be larger, 12.5, because the median number of fries would be at 80 now instead of 82, which would make the lower quartile at 75 instead of 77.

b. Will adding another data value always change the IQR? Give an example to support your answer.

No, it depends on how many values you have in the data set. For example, if the set of data is {2, 2, 2, 6, 9, 9, 9}, the IQR is 9 – 2 = 7. If you add another 6, the IQR would stay at 7.