

## Lesson 13: Describing Variability using the Interquartile Range (IQR)

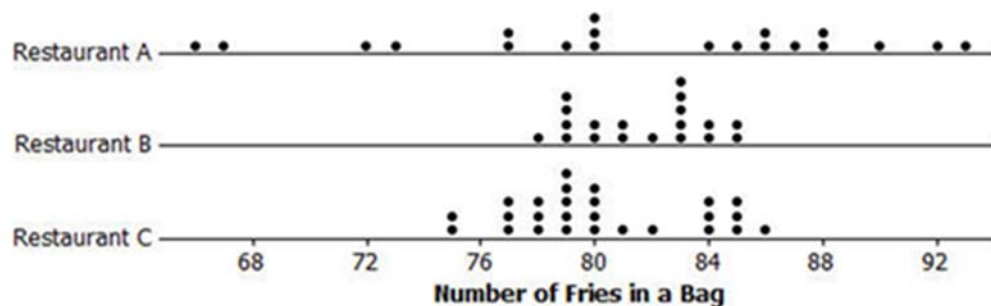
The median was used to describe the typical value of our data in Lesson 12. Clearly, not all of the data is described by the value. How do we find a description of how the data vary? What is a good way to indicate how the data vary when we use a median as our typical value? These questions are developed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

### Classwork

#### 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?
  - b. What scenario(s) would give counts that might not be representative of typical 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?

- b. Which of the restaurants had the smallest IQR, and what does that tell you?
- c. About what fraction of the counts would be between the quartiles? Explain your thinking.
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?
- 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.

### 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, 17

- III. Find the median: There are 17 data points so the 9<sup>th</sup> one from the smallest or from the largest will be the median.

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

median

- IV. Find the lower quartile and upper quartile: The lower quartile (Q1) will be half way between (the mean) of the 4<sup>th</sup> and 5<sup>th</sup> data points (4 and 6), or 5 and the upper quartile (Q3) will be half way between the 13<sup>th</sup> and the 14<sup>th</sup> data points (15 and 15), or 15.

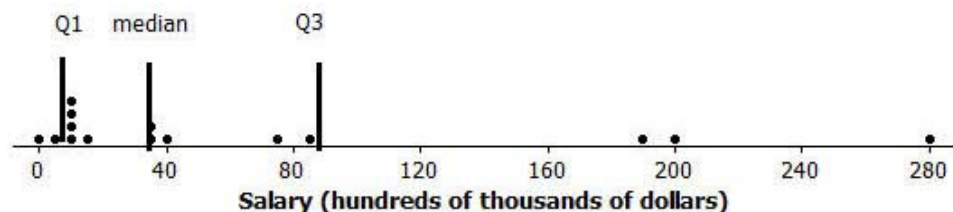
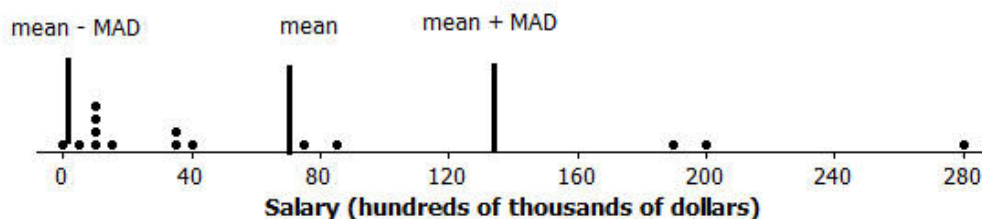
1, 1, 3, 4, 6, 6, 7, 8, 10, 11, 11, 12, 15, 15, 17, 17, 17

Q1 is 5      Q3 is 15

- V. Find the difference between Q3 and Q1: The **IQR** =  $15 - 5 = 10$ .

### Exercises 5–6

5. When should you use the IQR? The data for the 2012 salaries for the Lakers basketball team are given in the two plots below (see problem 5 in the Problem Set from Lesson 12).



- a. The data are given in hundreds of thousands of dollars. What would a salary of 40 hundred thousand dollars be?
- b. The vertical lines on the top plot show the mean and the mean  $\pm$  the MAD. The bottom plot shows the median and the IQR. Which interval is a better picture of the typical salaries? Explain your thinking.

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.

a.

b.

c.

### 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.

### Problem Set

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

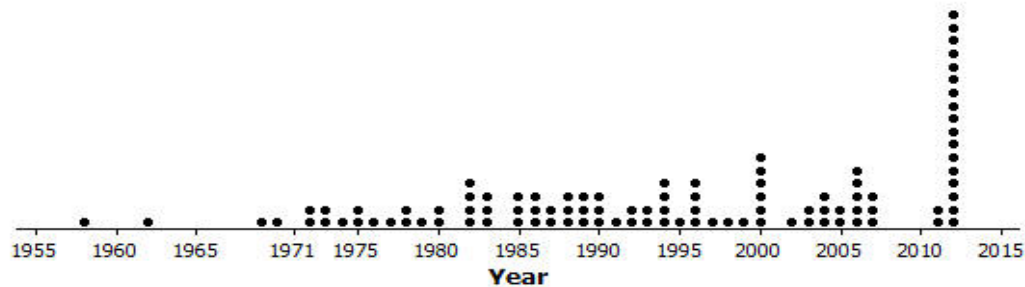
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
St. Louis	40	45	55	67	77	85	89	88	81	69	56	43
San Francisco	57	60	62	63	64	67	67	68	70	69	63	57

Data Source: [www.weather.com/weather/wxclimatology/monthly/graph/USCA0987](http://www.weather.com/weather/wxclimatology/monthly/graph/USCA0987)

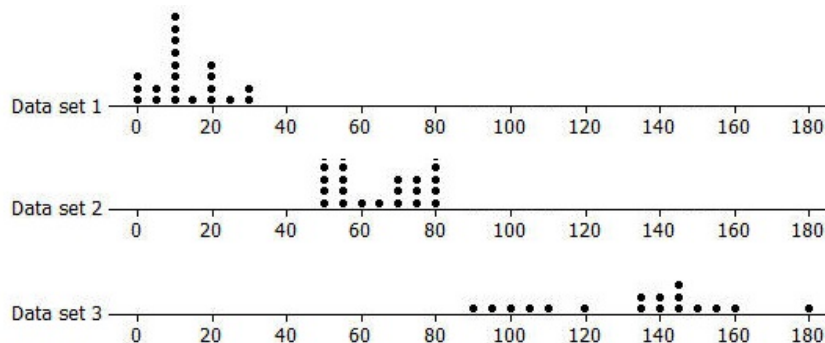
[www.weather.com/weather/wxclimatology/monthly/graph/USMO0787](http://www.weather.com/weather/wxclimatology/monthly/graph/USMO0787)

- How do you think the data might have been collected?
- 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?
- Make a prediction about how the sizes of the IQR for the temperatures for each city compare. Explain your thinking.
- Find the IQR for the average monthly high temperature for each city. How do the results compare to your conjecture?

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



- What does the stack of 17 dots at 2012 representing 17 pennies tell you about the “age” of the pennies in 2014?
  - 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.
3. Create a data set with at least 6 elements such that it has the following:
- A small IQR and a big range (maximum-minimum).
  - An IQR equal to the range.
  - The lower quartile is the same as the median.
4. Rank the following three data sets by the value of the 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.
- 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.
  - Will adding another data value always change the IQR? Give an example to support your answer.