

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

Classwork

In Lesson 12, the median was used to describe a typical value for a data set. But the values in a data set vary around the median. What is a good way to indicate how the data vary when we use a median as an indication of a typical value? These questions are explored in the following exercises.

Exercises 1–4: More French Fries

- 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.
 - How do you think the data were collected, and what problems might have come up in collecting the data?
 - What scenario(s) would give counts that might not be representative of typical bags?
- The medians of the top half and the medians of the bottom half of the data for each of the three restaurants are as follows: Restaurant A—87.5 and 77; Restaurant B—83 and 76; Restaurant C—84 and 78. The difference between the medians of the two halves is called the *interquartile range*, or IQR.
 - What is the IQR for each of the three restaurants?
 - Which of the restaurants had the smallest IQR, and what does that tell you?

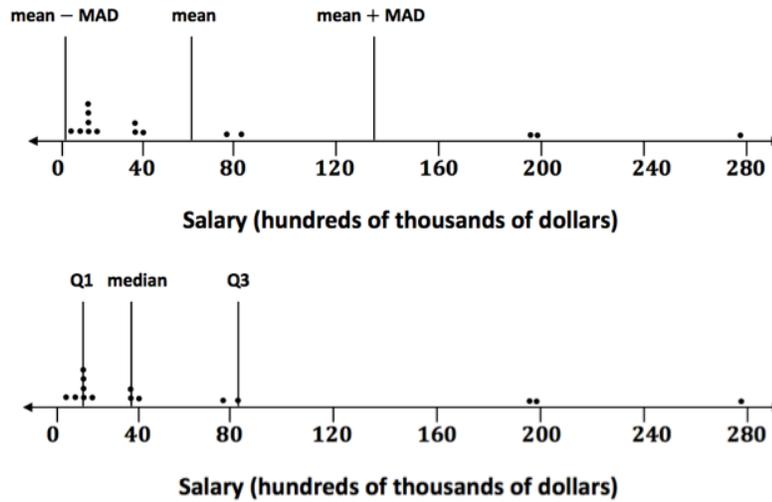
- c. The median of the bottom half of the data is called the *lower quartile* (denoted by Q_1), and the median of the top half of the data is called the *upper quartile* (denoted by Q_3). About what fraction of the data would be between the lower and upper quartiles? Explain your thinking.
3. Why do you think that the median of the top half of the data is called the *upper quartile* and the median of the bottom half of the data is called the *lower quartile*?
- 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 large bag? Explain your thinking.

Exercise 5: When Should You Use the IQR?

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.)



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

Exercise 6: On Your Own with IQRs

6. Create three different examples where you might collect data and where that data might have an IQR of 20. Define a median in the context of each example. 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 context.

a.

b.

c.

Lesson Summary

To find the IQR, you order the data, find the median of the data, and then find the median of the bottom half of the data (the lower quartile) and the median of the top 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. 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.

Problem Set

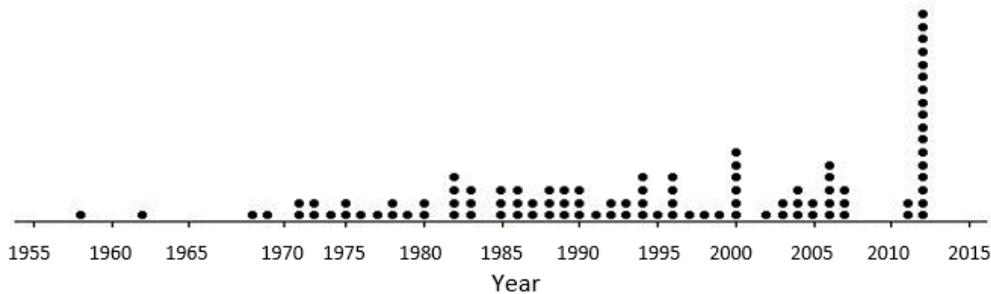
1. The average monthly high temperatures (in degrees Fahrenheit) for St. Louis and San Francisco are given in the table below.

| | 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 |

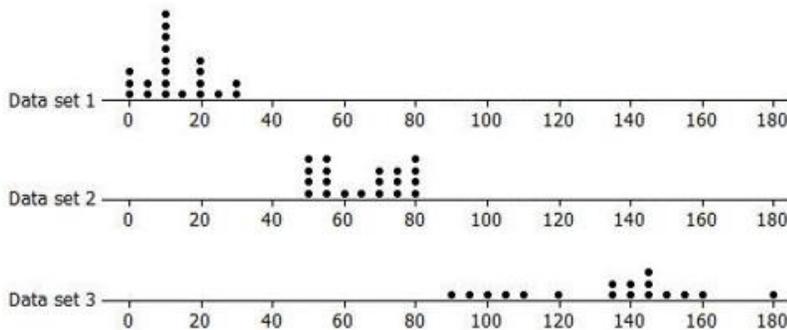
Data Source: <http://www.weather.com>

- 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°F or above? Why or why not?
- Make a prediction about how the values 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 what you predicted?

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 these pennies in 2014?
 - b. Here is some information about the sample of 100 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. In each of parts (a)–(c), create a data set with at least 6 values such that it has the following properties:
- a. A small IQR and a big range (maximum – minimum)
 - b. An IQR equal to the range
 - c. 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 number of 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, 77
- a. Suppose one bag of fries had been overlooked and that bag had only 50 fries. If that value is added to the data set, would the IQR change? Explain your reasoning.
 - b. Will adding another data value always change the IQR? Give an example to support your answer.