Part 1: Finding the Median

Classwork

How do we summarize a data distribution? What provides us with a good description of the data? The following exercises help us to understand how a numerical summary provides an answer to these questions.

Example 1: The Median—A Typical Number

Suppose a chain restaurant (Restaurant A) advertises that a typical number of french fries in a large bag is . The dot plot shows the number of french fries in a sample of twenty large bags from Restaurant A.

Sometimes it is useful to know what point separates a data distribution into two equal parts, where one part represents the upper half of the data values and the other part represents the lower half of the data values. This point is called the *median*. When the data are arranged in order from smallest to largest, the same number of values will be above the median point as below the median.

Exercises 1–3

1. You just bought a large bag of fries from the restaurant. Do you think you have exactly french fries? Why or why not?
2. How many bags were in the sample?
3. Which of the following statement(s) would seem to be true for the given data? Explain your reasoning.
   1. Half of the bags had more than fries in them.
   2. Half of the bags had fewer than fries in them.
   3. More than half of the bags had more than fries in them.
   4. More than half of the bags had fewer than fries in them.
   5. If you got a random bag of fries, you could get as many as fries.

Exercises 4–5: A Skewed Distribution

1. The owner of the chain decided to check the number of french fries at another restaurant in the chain. Here are the data for Restaurant B: ,,,,,,, ,, , , , ,,,, , ,
   1. How many bags of fries were counted?
   2. Sallee claims the median is because she sees that is the middle number in the data set listed above. She thinks half of the bags had fewer than fries because there are data values that come before in the list, and there are data values that come after in the list. Do you think she would change her mind if the data were plotted in a dot plot? Why or why not?
   3. Jake said the median was . What would you say to Jake?
   4. Betse argued that the median was halfway between and , or . Do you think she is right? Why or why not?
   5. Chris thought the median was . Do you agree? Why or why not?
2. Calculate the mean, and compare it to the median. What do you observe about the two values? If the mean and median are both measures of center, why do you think one of them is smaller than the other?

Exercises 6–8: Finding Medians from Frequency Tables

1. A third restaurant (Restaurant C) tallied the number of fries for a sample of bags of french fries and found the   
   results below.

|  |  |
| --- | --- |
| **Number of Fries** | **Frequency** |
|  | || |
|  | | |
|  | || |
|  | ||| |
|  | ~~||||~~ |
|  | |||| |
|  | | |
|  | | |
|  |  |
|  | ||| |
|  | ||| |
|  | | |

* 1. How many bags of fries did they count?
  2. What is the median number of fries for the sample of bags from this restaurant? Describe how you found your answer.

1. Which of the three restaurants seems most likely to really have fries in a typical bag? Explain your thinking.

Summary

The **median** is the middle value (or the mean of the two middle values) in a data set that has been ordered from smallest to largest. The median separates the data into two parts with the same number of data values below the median as above the median in the ordered list. To find a median, you first have to order the data. For an even number of data values, you find the average of the two middle numbers. For an odd number of data values, you use the middle value.

Part 2: Finding the Upper Quartile, Lower Quartile and Innerquartile Range

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 8–12: More French Fries

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 . Then, you looked at data on the number of fries in a bag from three of the restaurants.
   1. How do you think the data were collected, and what problems might have come up in collecting the data?
   2. What scenario(s) would give counts that might not be representative of typical bags?
2. Find the median of the top half and the median of the bottom half of the data for each of the three restaurants.
   1. Restaurant A
      1. Top half median:
      2. Bottom half median:
   2. Restaurant B
      1. Top half median:
      2. Bottom half median:
   3. Restaurant C
      1. Top half median:
      2. Bottom half median:
3. The difference between the medians of the two halves is called the *interquartile range,* or IQR.
   1. What is the IQR for each of the three restaurants?
   2. Which of the restaurants had the smallest IQR, and what does that tell you?
   3. The median of the bottom half of the data is called the *lower quartile* (denoted by Q1), and the median of the top half of the data is called the *upper quartile* (denoted by Q3). About what fraction of the data would be between the lower and upper quartiles? Explain your thinking.
4. 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*?
5. Mark the quartiles for each restaurant on the graphs below.



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

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 of the data in each section. Two of the sections are between the quartiles, so the interval between the quartiles would contain about of the data.

Part 3: Building a Boxplot

Classwork

A box plot is a graph that is used to summarize a data distribution. What does the box plot tell us about the data distribution? How does the box plot indicate the variability of the data distribution? These questions are explored in this lesson.

Example 2: Making a Box Plot

A box plot is a graph made using the following five numbers: the smallest value in the data set, the lower quartile, the median, the upper quartile, and the largest value in the data set.

To make a box plot:

* Find the median of all of the data.
* Find Q1, the median of the bottom half of the data, and Q3, the median of the top half of the data.
* Draw a number line, and then draw a box that goes from Q1 to Q3.
* Draw a vertical line in the box at the value of the median.
* Draw a line segment connecting the minimum value to the box and a line segment that connects the maximum value to the box.

You will end up with a graph that looks something like this:



Now, use the given number line to make a box plot of the data below.

, ,,,,,,,

The five-number summary is as follows:

Min

Q1

Median

Q3

Max

Exercises 13 - 16: French Fries Boxplots

1. Build a boxplot for Restaurant A

The five-number summary is as follows:

Min

Q1

Median

Q3

Max

1. Build a boxplot for Restaurant B

The five-number summary is as follows:

Min

Q1

Median

Q3

Max

1. Build a boxplot for Restaurant C

The five-number summary is as follows:

Min

Q1

Median

Q3

Max

1. Write 4 questions you could answer using the box plots you have created above.

\***Adapted from Eureka Math and EngageNY licensed by Great Minds.**