\*Material below is adapted and modified from EngageNY, derived from Eureka Math and licensed by GreatMinds.

How Many French Fries?

Part I: Finding the Median

Classwork

Students read the following paragraph silently.

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 (5 minutes): The Median—A Typical Number

The activity begins with a set of data represented as a dot plot. The concept of median is then developed by having students consider a sequence of questions. Once the concept has been developed, the median is formally defined. Begin by introducing the data presented in the example.

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.

* What could the restaurant mean when they say that the typical number of french fries in a large bag is ?
* *Answers will vary, but students may suggest that the mean is about .*
* Locate on the dot plot. What do you notice about the number of data values that are above and the number of data values that are below ?
  + *There are the same number of data values on either side of —ten data values are greater than , and data values are less than .*
* The restaurant used a summary measure called the *median* to describe the typical number of french fries. The median represents the middle value in a data set when the data values are arranged in order from smallest to largest. The same number of values will be above the median as are below the median.

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 (4 minutes)

Students work independently on the exercises and confirm answers with a neighbor.

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?

The number of fries in a bag seems to vary greatly from bag to bag. No bag had exactly fries, so mine probably will not. The bags that were in the sample had from to french fries.

1. How many bags were in the sample?

bags were part of the sample.

1. 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.

Statements (a) and (b) are true because there are bags above fries and bags below fries. Also, statement (e) is true because that happened once, so it could probably happen again.

Discuss the answers to the two questions below with the whole class.

* How do you find the median if there are an odd number of data points?
  + *Put the data values in order from smallest to largest, or construct a dot plot of the data.*
  + *Find the middle number in the ordered list or on the dot plot. One way to do this is to divide the number of observations by and then round up to get an integer. This identifies the position of the median. For example, if there are observations, dividing by gives us , which rounds up to . Then, starting with the smallest observation, count up to find the th number in the ordered list. This number is the median.*
* How do you find the median if there are an even number of data points?
  + *Put the data values in order from smallest to largest, or construct a dot plot of the data.*
  + *Find the middle two numbers in the ordered list or on the dot plot. One way to do this is to divide the number of observations by . This identifies the position of the first of the two middle numbers. For example, if there are observations, dividing by gives us observations. Then, starting with the smallest observation, count up to find the th number in the ordered list. This number and the next number in the list are the two middle values.*
  + *Find the mean of the two middle values. This number is the median.*

Exercises 4–5 (10 minutes): A Skewed Distribution

In this set of exercises, students have to put the data in order from smallest to largest before they find the median. There are values, so the median is the th value with data values above and data values below. Another way to determine the median after ordering the data is to cross out the maximum and minimum values, then the next largest and smallest values, and so on until students are left with just one number in the middle if there are an odd number of data values, or two numbers if there are an even number of data values. If this process results in a single number, that number is the median. If this process results in two numbers in the middle, students would then find the mean of the two values to get the value of the median.

MP.3

The questions are designed to help students confront some common misconceptions and errors: not ordering the data before counting to the middle, confusing median and mode (most frequent value), and confusing median and midrange (halfway between the maximum and the minimum). They also compute the mean and compare the median to the mean, noting that the median might be more reflective of a typical value because several bags with low numbers of french fries pulled the mean down.

Consider the following questions as students are completing the exercises:

* Why is it necessary to order the data before you find the median?
  + *In order to find the value that has half of the data values smaller and half of the data values larger, the data values must first be put in order.*
* Is the median connected to the range (maximum minimum) of the data? Why or why not?
  + *The median is not connected to the range because it is possible for the range to change while the median stays the same.*
* What is the difference in the effect of a few very extreme values on the mean and on the median?
  + *Extreme values will have a big effect on the mean but will not affect the median.*

Students should work in small groups to complete exercises 4-5. Teacher should circulate to support students in building understanding. Teacher might want to pose questions listed above to get students thinking as they work.

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?

bags of fries were counted.

* 1. Sallee claims the median is because she sees that is the middle number in the data set listed on the previous page. 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?

*Scaffolding:*

If students are struggling, allow them to determine the median before answering the questions.

Yes. You cannot find the median unless the data are organized from least to greatest. Plotting the number of fries in each bag on a dot plot would order the data correctly. You would probably get a different halfway point because the data above are not ordered from least to greatest.

* 1. Jake said the median was . What would you say to Jake?

is the most common number of fries in the bags ( bags had fries), but it is not in the middle of the data.

* 1. Betse argued that the median was halfway between and , or . Do you think she is right? Why or why not?

She is wrong because the median is not calculated from the distance between the largest and smallest values in the data set. This is not the same as finding a point that separates the ordered data into two parts with the same number of values in each part.

* 1. Chris thought the median was . Do you agree? Why or why not?

Chris is correct because if you order the numbers, the middle number will be the th number in the ordered list, with at most bags that have more than fries and at most bags that have fewer than fries.

1. 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?

The mean is , and the median is . The bag with only fries decreased the value of the mean.

Exercises 6–8 (10 minutes): Finding Medians from Frequency Tables

**MP.4**

In these exercises, students find the median using data summarized in a frequency table. The median falls halfway between the th and th data value when the data are ordered from smallest to largest. They also find the medians of the top and bottom halves of the data set, the th value from the top and from the bottom, as a precursor to finding quartiles and the interquartile range in a later lesson. Consider having students write out the individual counts in a long ordered list. For example, the first counts would be as follows:

Median of the lower half

In these exercises, students need to deal with repeated data values when finding the median. In this case, have students find the median by counting from the top and bottom of the list, noting that values for bags with the same count can fall on both sides of the median. It might help to think about the individual bags: One of the bags with fries is in the first half, one of the bags with fries is in the second half, and one of the bags divides the two halves and marks the median of the data set. At this point, the important idea is that students get a sense of how to find a median: Order the values, and find a middle value in the ordered list of data values.

Students should work in small groups to complete exercises 6 – 8. Teacher might want to model finding the median with an arbitrary set of numbers to remind students what to do, especially if this is the start of a second class period.

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.

*Scaffolding:*

If students are struggling, encourage them to list out the values in a numerical list or to create a dot plot from the frequency table.

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

* 1. How many bags of fries did they count?

They counted bags of fries.

* 1. What is the median number of fries for the sample of bags from this restaurant? Describe how you found your answer.

; I took half of , which is , and then counted tallies from to reach . I also counted tallies from to reach . The point halfway between and is the median.

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

Answers will vary. The data sets for Restaurants A and B both have a median of . Look for answers that consider how much the data values vary around . Restaurant B seems to have the most bags closest to a count of . The data set for Restaurant C has a median of , but the data values are not very spread out, and most are close to , so some students might make a case for Restaurant C.

Closing (3 minutes) – Have students turn and talk, then share answers with the class.

* Does the median have to be a value in the data set?
  + *The median does not have to be a value in the data set.*
* Is calculating the median the same as calculating the middle of the range? Explain.
  + *The median is not the same as calculating the middle of the range. Finding the point that is halfway between the largest and the smallest data value is not the same as finding a value that will have half of the data values above and half of the data values below. For example, think about the data set consisting of . The median is , but halfway between the largest and smallest data values is .*

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

Classwork

Exercises 8–12 (16 minutes): More French Fries

These exercises return 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 where they found medians of the lower half and upper half of a data set and use these medians to calculate the interquartile range (IQR).

MP.1

These exercises build an understanding of quartiles. Quartiles divide a data set into quarters, with of the data falling below the lower quartile, falling between the lower quartile and the median (which is sometimes called the *middle quartile*), falling between the median and the upper quartile, and falling above the upper quartile. The upper quartile is found by finding the median of the top half of the data set, and the lower quartile is found by finding the median of the bottom half of the data set. This is why students began looking at medians of the top and bottom halves in Lesson 12. The upper and lower quartiles are then used to compute the interquartile range (IQR), which is a measure of variability in a data set. Students should be able to approximate the number of elements in each section in terms of , or , of the data or by giving an estimate of the actual number of data values as well as knowing that , or , of the data values are between the lower and upper quartiles. They should also recognize that the IQR is a measure of spread around the median (it is the length of the interval that captures the middle of the data).

Data: Number of french fries by Restaurant (*Post on the board or give to each group of students)*

Restaurant A: ,,,,,, ,,,,,,, ,, ,,,,

Restaurant B: ,,,,,,, ,, , , , ,,,, , ,

Restaurant C:

Students work in small groups to complete 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?

Answers will vary. They probably went to the restaurants and ordered a bunch of large 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.

* 1. 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 bought the bags 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.

1. 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: 87.5
      2. Bottom half median: 77
   2. Restaurant B
      1. Top half median: 83
      2. Bottom half median: 76
   3. Restaurant C
      1. Top half median: 84
      2. Bottom half median: 78
2. 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?

The IQR for Restaurant A is ; Restaurant B is ; Restaurant C is .

* 1. Which of the restaurants had the smallest IQR, and what does that tell you?

Restaurant C 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 a range of fries and near the median, so the median is a pretty good description of what is typical.

* 1. 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.

About , or , of the counts would be between the quartiles because about of the counts are between the median and the lower quartile, and of the counts are between the median and the upper quartile.

1. 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*?

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 about one fourth, or a quarter, of the data values in each section.

* 1. 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.

The IQR does help decide which restaurant is most likely to have fries in a typical large bag because the IQR explains the variability of the data. Because Restaurant C has the smallest IQR, the middle half of the counts of the number of fries in a bag is really close to the median. In addition, Restaurant C also has the smallest range.

Part 3: Building the Boxplot

Example 2 (7 minutes): Making a Box Plot

This example describes the procedure for finding a box plot. Consider having students read through the steps themselves, and then ask them to restate the directions.

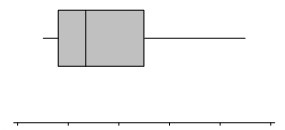
Example 2: Making a Box Plot

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:

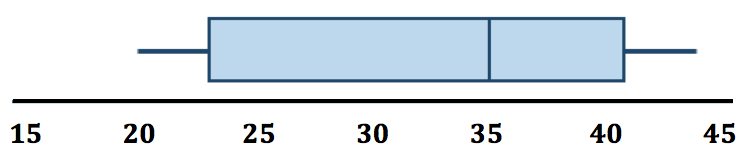


Now, use the given number line to make a box plot of the data below.

, ,,,,,,,

The five-number summary is as follows:

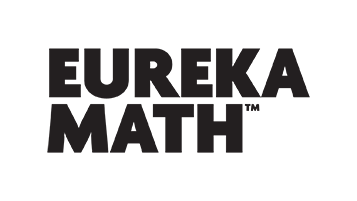
Min   
Q1   
Median   
Q3   
Max



Ask students the following when the box plot is complete.

* Why is it important to have a standard way to make a box plot?
  + *It is important to have a standard way to make a box plot because it guarantees that everyone will have the same plot when given the same set of data.*
* What does a box plot tell you about the *story* in the data? What does it not tell you?
  + *A box plot divides the data in four parts, which each contain about of the data values. The “box” part of the box plot shows where the middle of the data values are*
  + *The box plot does not show us specific data values.*
* What proportion (percent) of the data are in each of the sections of the box plot? How do you know?
  + *Each section of the box plot represents of the data because the median and quartiles divide the data into four sections that have the same number of data values.*

Exercises 13 – 16. Students will build boxplots to represent each of the Restaurants A, B and C. Students will then develop questions that could be answered using the boxplots they have created. Students should work in their original groups to complete these exercises.

[http://mirrors.creativecommons.org/presskit/buttons/80x15/png/by-nc-sa.png](http://creativecommons.org/licenses/by-nc-sa/3.0/deed.en_US)

This work is licensed under a   
[Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.](http://creativecommons.org/licenses/by-nc-sa/3.0/deed.en_US)

[Thi](http://creativecommons.org/licenses/by-nc-sa/3.0/deed.en_US)s work is derived from Eureka Math ™ and licensed by G[reat Minds. ©2015 Gr](http://creativecommons.org/licenses/by-nc-sa/3.0/deed.en_US)eat Minds. eureka-math.org This file derived from G6-M6-TE-1.3.0-10.2015

**S.97**