Grade/Subject	Grade 6/ Mathematics
Unit Title	Unit 6: Statistics and Distribution
	In this unit, students will manipulate data so that it can be displayed using dot plots, box plots and histograms. Students will also analyze this data using measures of center, variability, and patterns.
Pacing	28 days

Background Information For The Teacher

In the 6th grade Statistics and Distribution unit, teachers will have to have good understanding of the development of their students' statistical and data analysis learning progression as it is varied from past practices (as outlined below). Major points include students creating and beginning to analyze bar graphs in grade 3, making line plots in grades 4 and 5, and then analyzing several measures as they relate to statistics and distributions in grade 6. Also, classic graphs such as pictographs, line graphs and circle graphs may not necessarily enter the realm of the students' knowledge concerning pictorial representations of data. Additionally, probability is a Common Core 7th grade standard where it had traditionally been taught in 6th grade with statistics.

In 3rd grade, students will have:

- Draw a scaled picture graph and a scaled bar graph to represent a data set with several categories
- Solved problems using information presented in scaled bar graphs
- Generated measurement data by measuring lengths using rulers marked with halves and fourths of an inch.
- · Shown the data by making a line plot, where the horizontal scale is marked off in appropriate units

In 4th grade, students will have:

- Made a line plot to display a data set of measurements in fractions of a unit (1/2, 1/4, 1/8)
- Solved problems involving addition and subtraction of fractions by using information presented in line plots

In 5th grade, students will have:

- Made a line plot to display a data set of measurements in fractions of a unit (1/2, 1/4, 1/8)
- Used operations on fractions for this grade to solve problems involving information presented in line plots

In 6th grade, students will:

- Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for the data in the answers
- Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread and overall shape
- · Recognize that a measure of center summarizes all its values with a single number
- Recognize that a measure of variation describes how its values vary with a single number
- Display numerical data in plots on a number line including dot pl plots, histograms, and box plots.
- Summarize numerical data sets in relation to their context

Essential Questions (and Corresponding Big Ideas)

Why is it important to summarize and display numerical data?

- Line plots, Histograms, Box plots are all helpful ways to summarize and display numerical data for organizational purposes and analysis.
- Measures of center and variations and distributions can describe a data set's center, spread and overall shape.

What are measures of center and how are they helpful in describing data?

• Median, mode and mean are data analysis tools that summarize all of a data set's values with a single number. They are helpful in describing the overall shape of the data.

What are measures of variation and how are they helpful in describing data?

• Interquartile range and mean absolute deviation are data analysis tool that describes how a data set's values vary with a single number. They are helpful in describing the spread of the data.

Core Content Standards	Explanations and Examples
6.SP.1. Recognize a statistical question as one	6.SP.1. Statistics are numerical data relating to an aggregate of individuals; statistics is
that anticipates variability in the data related	also the name for the science of collecting, analyzing and interpreting such data. A

to the question and accounts for it in the answers. For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages.

The focus for this standard is identifying the difference between statistical and non-statistical questions and formulating/writing simple questions to provide differences in responses. A statistical question must be stated so that responses will allow for differences. In the example, "What color are the shoes I am wearing?" only one response can be given. However, with the example, "What colors of shoes are the students in our class wearing?" a variety of responses can be collected.

What the teacher does:

- Allow students to discover that statistics is the study of numerical information, called data, through classroom discussion of examples. Statisticians collect, organize, and analyze data around a statistical question. A statistical question is used to collect statistical information. Discuss with students that a statistical question must be stated so the responses will allow for differences. In the example, "What color are the shoes I am wearing?" only one response can be given. However, with the example, "What colors of shoes are the students in our class wearing?" a variety of responses can be collected.
- Provide numerous examples for students to sort and categorize questions as statistical or non-statistical.
- Invite students to write and share their own statistical questions that can be used to survey and collect data from other classmates. Sixth graders need multiple experiences writing statistical questions.
- Focus on the following vocabulary terms: statistics, data, and variability.
- Allow students to talk with one another and the teacher to make sense of the definition of statistics, recognizing statistical questions and anticipating variability in the data related to the question.

6.SP.2. Understand that a set of data collected to answer a statistical question has a

statistical question anticipates an answer that varies from one individual to the next and is written to account for the variability in the data. Data are the numbers produced in response to a statistical question. Data are frequently collected from surveys or other sources (i.e. documents).

Questions can result in a narrow or wide range of numerical values. For example, asking classmates "How old are the students in my class in years?" will result in less variability than asking "How old are the students in my class in months?"

Students might want to know about the fitness of the students at their school. Specifically, they want to know about the exercise habits of the students. So rather than asking "Do you exercise?" they should ask about the amount of exercise the students at their school get per week. A statistical question for this study could be: "How many hours per week on average do students at Jefferson Middle School exercise?"

To collect this information, students might design a survey question that anticipates variability by providing a variety of possible anticipated responses that have numerical answers, such as: 3 hours per week, 4 hours per week, and so on. Be sure that students ask questions that have specific numerical answers.

What the students do:

- Understand that data generated from statistical questions vary.
- Identify the difference between a statistical and non-statistical question.
- Recognize that responses to statistical questions have variations that can be used to draw conclusions about the data set
- Formulate and write simple statistical questions that provided differences in responses.

Misconceptions and Common Errors:

Since sixth grade introduces the first formal study of statistics, many students may write questions that do not allow for differences or a variety of responses. To help with this, provide repetitive practice and guidance to help the students write statistical questions having variations that may be used to draw conclusions about the data. Do many sorting activities with these students. Play the example/non-example game where the teacher displays two columns. In the first column, the teacher lists two or three examples, and in the second, the teacher lists two or three non-examples. From these lists, students try to guess what concept the teacher is making examples of. As students think they know the concept or what to test their hypothesis of what is the same about all the examples in the first column, the students add examples and non-examples to the lists, thereby providing more examples for the students who do not understand the concept. This aids students who need more time and more examples

distribution, which can be described by its center, spread, and overall shape.

Standard 2 focuses on the understanding that data collected to answer a statistical question can be analyzed by their distribution. A distribution is the arrangement of the values of a data set and is described as using its center (median or mean) and spread. The single value for each of the measure of center (mean, median, or mode) and measures of spread (range) is used to summarize the data. By finding the measures of center for a set of data, students use the value to describe the data in words. Students use histograms and box plots to describe a set of data using its center (mean, median, and mode), spread (range), and overall shape.

What the **teacher** does:

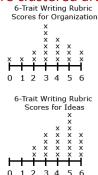
- Pose questions to the class to collect and analyze data, such as, "How long does it take you to do your homework?" Or, "How many text messages do you estimate getting each day?" Have the students create dot plots to examine the distribution of the data set and discuss the center, spread, and overall shape.
- Examine the data by finding the mean, median, mode and range of the data. These concepts are new to the students, so let them discover and report back to the class what each term means and how they will find the mean, media, mode and rage for their dot plots. Divide the class into groups of four. Each group will study one of the concepts and report back to the class what they learned.
- One group of students should report that the mean is a measure of center of that data summarized by a single number and that it represents the arithmetic average of the data. To find the mean of a data set, add all the values

t help them focus on the concept.

6.SP.2.

The two-dot plots show the 6-trait writing scores for a group of students on two different traits, organization and ideas. The center, spread and overall shape can be used to compare the data sets. Students consider the context in which the data were collected and identify clusters, peaks, gaps, and symmetry. Showing the two graphs vertically rather than side by side helps students make comparisons.

For example, students would be able to see from the display of the two graphs that the ideas scores are generally higher than the organization scores. One observation students might make is that the scores for organization are clustered around a score of 3 whereas the scores for ideas are clustered around a score of 5.



What the students do:

- Understand that data collected to answer a statistical question can be analyzed by the distribution.
- Calculate mean, median, mode and range.
- Describe a set of data using its center (mean, median, and mode), speed (range), and overall shape.
- Create a line plot, histogram, and box plot.

Misconceptions and Common Errors:

Some students may describe the spread of the data as low to high such as 6-21. Remind students that spread (range) is stated as a single number such as 15 and describes how the values vary across the data set. Stress that the purpose of the number is not

 Another group should say that the median is a measure of center of that data summarized by a single number and represents the point at which 50% of the data is greater than or equal to that number and 50% is less than or equal to that number. To find the median, place the numbers in

value order and hen find the middle.

together and divide by the number of values in the set.

- The third group of students should state that the **mode** of a set of numerical data represents the center of that data summarized by a single number. It represents the most frequent value of a set of data. To find the mode, put the numbers in order from least to greatest and count how many times each number occurs. The number that occurs the most is the mode. Note that some data sets are bimodal, meaning two numbers occur equally as much or more than the rest of the data points.
- The last group of students should explain that the range of a set of numerical data is a measure of how the data vary, summarized by a single number. Find the highest and lowest numbers in the set of data.
- Note that the suggest ideas will likely take 1 week to 10 days of instructional time. Keep in mind that students are beginning to develop their ability to think statistically with the key focus of describing and summarizing numerical data sets. After the graphs have been created by each student, focus on summarizing the data represented by the graphs. Describe the data using measures of center (mean, median, mode) and the spread and the overall shape of the data.
- Emphasize the following vocabulary: distribution, center, spread, overall shape, histograms, line plots, box plots, outliers, quartiles, upper and lower extreme, and whiskers.
- Provide cyclical, distributed reviews over time to practice collecting data to answer a statistical question with a distribution described by its center, spread, and overall shape.

6.SP.3. Recognize 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 its values vary with a single number.

the value itself but the interpretation it provides for the variation of the data.

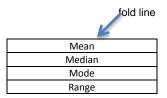
Box plots seem complicated to some students because of the numerous steps involved in creating them. Keep in mind that students are beginning to develop their ability to think statistically, and this new learning must focus around students learning to describe and summarize numerical data sets and not just the procedure of creating the graphs. Technology can be used to create these graphs. Free programs are available on the Internet as are YouTube videos and commercially produced software packages.

The terms needed for describing data, including distribution, center, spread, overall shape, histograms, line plots, box plots, outliers, quartiles, upper and lower extreme, and whiskers, are all new to the students. Creating visual graphic vocabulary organizers such as the Frayer model will help students make sense of the terms they need to describe the data.

This standard helps students understand that a data distribution may not have a definite center. Sixth graders discover that different ways to measure center produce different values. The median measures center as the middle value. The mean measures center as the value that each data point would take on if the total of the data values were redistributed equally. It is a balance point. Students recognize that a measure of variability can also summarize data because two very different sets of data can have the same median and mean but differ by their variability.

What the teacher does:

- Explain that interpreting different measures of center for the same data develops the understanding of how each measure can change how the data get interpreted.
- Plan activities that require students to match graphs to explanations or measures of center.
- Have students create a paper foldable such as a layered book with two sheets of layered paper to illustrate and compare the similarities and differences of mean, median, and mode. Include range, although it is not a measure of center. Stack and layer two sheets of 8 ½ x 11-inch paper. Be sure that the bottom sheet is about 1 inch higher than the top sheets in the stack. Create the layered book by folding the two sheets from the bottom up so that the layers leave flaps about the same distance apart. Write the terms as shown below and instruct students to flip up a tab to illustrate and write comparisons to show the similarities and differences. This may help students begin to understand the purposes among the measures of center (mean, median, and mode) and the distinction between center and spread.



- Focus on the following vocabulary terms: measure of center, mean, median, mode, and range.
- Allow students to talk with each other and the teacher to make sense of and recognize that a measure of center for a numerical data set summarizes all of it's values with a single

6.SP.3 When using measures of center (mean, median, and mode) and range, students are describing a data set in a single number. The range provides a single number that describes how the values vary across the data set. The range can also be expressed by stating the minimum and maximum values.

Example:

- Consider the data shown in the dot plot of the six trait scores for organization for a group of students.
 - o How many students are represented in the data set?
 - o What are the mean, median, and mode of the data set? What do these values mean? How do they compare?
 - o What is the range of the data? What does this value mean?



number, while a measure of variation describe how its values vary with a single number.

 Provide cyclical distributed reviews over time to promote the understanding of measure of center and identify mean, median, and mode given different data sets.

6.SP.4 Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

Students learn how to display data on dot plots, histograms, and box plots (also known as box and whisker plots). A dot plot is appropriate for small to moderate size data sets up to 25 numbers and is useful for highlighting the distribution and spread of the data, including clusters, gaps, and outliers. Histograms display the distributions of continuous data using intervals in a number line. Box plots display the distribution of values in a data set by dividing the set into quartiles. After creating the plots students interpret them, giving meaning of the context with statements such as, "There is little variation in these data because the range on this box plot is 3." Sixth graders learn to select the most appropriate display to represent the given data.

What the teacher does:

- Provide experiences for students to learn about displaying numerical data in plots on a number line, including dot plots, histograms, and box plots. Students must be able to choose an appropriate graph and display the data. They should be able to explain the data generated from the graphs.
- Pose a question or idea for students to create a dot plot.
 Dot plots are plots on a number line where each dot is representative of a piece of data in the set. To make a dot plot, first title the plot. Draw a horizontal line segment on grid paper. Make a scale of numbers below the line. The numbers should include the least value and the greatest value in the set of data. For each piece of data, draw an X or dot above the corresponding number.

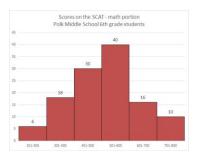
What the students do:

- Understand how to find mean, median, ode, and range.
- Model with examples the difference between measure of center and measures of spread.
- Determine appropriate center and variation for various data sets.

Misconceptions and Common Errors:

Some sixth graders may forget to divide when finding the mean. Model use of the term average as an inclusive term for many different measures of center, not just the mean. The word average connotes many different concepts for students that are not mathematical because the word has many uses in everyday language. A short lesson where students offer mathematical terms that have different meanings in everyday life can be enlightening for these students. A few of those terms are: table (a kitchen table vs. a chart), pie (apple pie vs. π), and ruler (king vs. a measuring tool).

- Ask students questions about the dot plots to ensure understanding, such as, "What does each X on the graph represent?" or "What is the range of the data?" Ask them to interpret the plot to give meaning to the context of the data with a summary.
- Pose a question or idea for students to create and explore histograms. The height of each bar represents the number of data values in that interval. Data are grouped into interval ranges. To make a histogram, it is helpful to make a frequency table first. Choose a range that contains all the data and divide it into equal intervals. After making the frequency table, it is easy to make the histogram because it is like making a bar graph, except each bar represents an interval and there are not spaces between the bars. Ask students questions about the histograms to ensure understanding, such as, "How do the intervals on a histogram present different information than the bars on the bar graph?" Asking them to interpret the histogram they create to give meaning to the context of the data with a summary. An example of a histogram follows:



Pose a question or an idea to help students experience a box plot. It can be graphed either vertically or horizontally and includes a lower quartile, a median, an upper quartile, and maximum. These values give a summary of the shape of a distribution. Students should understand that the size of the box or whiskers represents the middle 50% of the data. To make a box and whisker plot, first collect data. Organize them from least to greatest. Draw a number line that will show the data in equal intervals. Mark the median. Mark the median of the upper half (upper quartile). Mark the median of the lower half (lower quartile). Mark the upper extreme. Mark the lower extreme. Draw a box between the lower quartile and the upper quartile. Split the box by drawing a vertical line

6.SP.4. In order to display numerical data in dot plots, histograms or box plots, students need to make decisions and perform calculations. Students are expected to display data graphically in a format appropriate for that data set as well as reading data from graphs generated by others students or contained in reference materials. Students can use applets to create data displays. Examples of applets include the Box Plot Tool and Histogram Tool on NCTM's Illuminations.

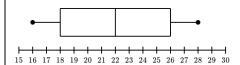
Box Plot Tool - http://illuminations.nctm.org/ActivityDetail.aspx?ID=77
Histogram Tool -- http://illuminations.nctm.org/ActivityDetail.aspx?ID=78
Dot plots are simple plots on a number line where each dot represents a piece of data in the data set. Dot plots are suitable for small to moderate size data sets and are useful for highlighting the distribution of the data including clusters, gaps, and outliers.

In most real data sets, there is a large amount of data and many numbers will be unique. A graph (such as a dot plot) that shows how many ones, how many twos, etc. would not be meaningful; however, a histogram can be used. Students bin the data into convenient ranges and use these intervals to generate a frequency table and histogram. Note that changing the size of the bin changes the appearance of the graph and the conclusions you may draw from it.

Box plots are another useful way to display data and are plotted horizontally or vertically on a number line. Box plots are generated from the five number summary of a data set consisting of the minimum, maximum, median, and two quartile values. Students can readily compare two sets of data if they are displayed with side by side box plots on the same scale. Box plots display the degree of spread of the data and the skewness of the data.

Examples:

through the median. Draw 2 whiskers from the quartile to the extremes. An example of the a box plot f or test graders from a mathematics class:



- After students create a box plot, ask questions to check for understanding about a box plot, such as, "What does a whisker tell you about the data? What questions can you NOT answer from a box and whisker plot?" Ask students to interpret the box and whisker plot to give meaning to the context of the data with a summary.
- Allow students to explore contextual sets of data and have them decide which statistical graph is best suited to the data.
- Model correct use of the following terms: dot plots, histograms, and box plots.
- Provide cyclical, distributed practice over time to review displaying numerical data in dot plots, histograms, and box plots and interpreting these graphs.

• Nineteen students completed a writing sample that was scored using the six traits rubric. The scores for the trait of organization were 0, 1, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 6, 6. Create a data display. What are some observations that can be made from the data display?



Grade 6 students were collecting data for a math class project. They decided they would survey the other two grade 6 classes to determine how many DVDs each student owns. A total of 48 students were surveyed. The data are shown in the table below in no specific order. Create a data display. What are some observations that can be made from the data display?

11	21	5	12	10	31	19	13	23	33
10	11	25	14	34	15	14	29	8	5
22	26	23	12	27	4	25	15	7	
2	19	12	39	17	16	15	28	16	

A histogram using 5 bins (0-9, 10-19, ...30-39) to organize the data is displayed below.



6.SP.5. Summarize numerical data sets in relation to their context, such as by:

• Ms. Wheeler asked each student in her class to write their age in months on a sticky note. The 28 students in the class brought their sticky note to the front of the room and posted them in order on the white board. The data set is listed below in order

- a. Reporting the number of observations.
- Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
- c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), 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.
- d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

This standard emphasizes summarizing data. Students communicate a deep understanding of (1) observations (sample size, sometimes labeled as the n of the data), (2) appropriate measure of center and spread for a particular data set, (3) appropriate section of a graph to represent data collected, and (4) overall patterns in a distribution, including outliers, through statistical investigations.

What the teacher does:

- Begin by exploring plots (graphs) and asking students to find and report the number of observations. Ensure that students understand observations means sample size or n size and how it relates to numerical data sets. For example, a set of data with 10 data points has 10 observations, or we can say n = 10. Ensure that students understand and can explain why the number of observations is important to summarizing numerical data sets. Ensure students know the difference between intervals and observations.
- Provide a variety of samples of plots from newspapers and

from least to greatest. Create a data display. What are some observations that can be made from the data display?

ı	130	130	131	131	132	132	132	133	134	136
ı	137	137	138	139	139	139	140	141	142	142
1	142	143	143	144	145	147	149	150		

Five number summary

Minimum - 130 months

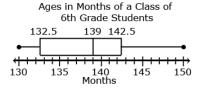
Quartile 1 (Q1) – (132 + 133) \div 2 = 132.5 months

Median (Q2) - 139 months

Quartile 3 (Q3) – $(142 + 143) \div 2 = 142.5$ months

Maximum - 150 months

(Continued on next page)



This box plot shows that

- ¼ of the students in the class are from 130 to 132.5 months old
- ¼ of the students in the class are from 142.5 months to 150 months old
- ½ of the class are from 132.5 to 142.5 months old
- The median class age is 139 months.

What the students do:

- Understand that data are organized in graphs for the purpose of analyzing the data.
- Represent given data on the most appropriate graph (dot plot, histogram, or box plot).
- Interpret data represented on dot plots, box plots and histograms for given situations.

Misconceptions and Common Errors:

Confusing histograms and bar graphs is a common error students have that makes it difficult for them to interpret the intervals. Display histograms and bar graphs side by side with related data and ask students to compare and contrast what you can learn from each graph.

allow students to identify attributes under investigation, including how the attributes were measured and their units of measure.

- Invite students to ask a statistical question of interest, collect data to answer the question, and display the data on an appropriate graph. Ask students to summarize the data in a presentation given to the class that includes: (1) the number of observations; (2) a description of the attributes investigated, including how they were measured and the units used; (3) a measure of center and a measure of variability along with a defense for the various measures chosen; and (4) any overall pattern in the distribution or outliers noted.
- Teach mean absolute deviation (MAD) as a measure of variability. The mean absolute deviation of a set of data is the average distance between each data value and the mean. To find the mean absolute deviation, first find the mean. Next find the distance between each data value in the set and the mean. Then find the absolute value of the differences. Finally, find the mean of those differences. For an activity on mean absolution deviation,.
- Design a class activity to interpret data with a box and whisker plot. Have each student estimate and record how many hours of TV he or she watch in a month and then graph the data as a class on a line plots. Interpret the data and create a box and whisker plot. Use the box and whisker plot to show variation of data.
- Provide cyclical, distributed practice over time to review and summarize numerical data sets in relation to their contexts.
- Allow students to talk with each other and their teacher to make sense of Standard 5.

6.SP.5. Students summarize numerical data by providing background information about the attribute being measured, methods and unit of measurement, the context of data collection activities, the number of observations, and summary statistics. Summary statistics include quantitative measures of center, spread, and variability including extreme values (minimum and maximum), mean, median, mode, range, quartiles, interquartile ranges, and mean absolute deviation.

The measure of center that a student chooses to describe a data set will depend upon the shape of the data distribution and context of data collection. The mode is the value in the data set that occurs most frequently. The mode is the least frequently used as a measure of center because data sets may not have a mode, may have more than one mode, or the mode may not be descriptive of the data set. The mean is a very common measure of center computed by adding all the numbers in the set and dividing by the number of values. The mean can be affected greatly by a few data points that are very low or very high. In this case, the median or middle value of the data set might be more descriptive. In data sets that are symmetrically distributed, the mean and median will be very close to the same. In data sets that are skewed, the mean and median will be different, with the median frequently providing a better overall description of the data set.

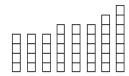
Understanding the Mean

The mean measures center in the sense that it is the value that each data point would take on if the total of the data values were redistributed equally, and also in the sense that it is a balance point. Students develop understanding of what the mean represents by redistributing data sets to be level or fair.

The leveling process can be connected to and used to develop understanding of the computation of the mean. For example, students could generate a data set by measuring the number of jumping jacks they can perform in 5 seconds, the length of their feet to the nearest inch, or the number of letters in their names. It is best if the data generated for this activity are 5 to 10 data points, which are whole numbers between 1 and 10 that are easy to model with counters or stacking cubes.

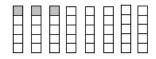
Students generate a data set by drawing eight student names at random from the popsicle stick cup. The number of letters in each of the names is used to create the data set. If the names drawn were Carol, Mike, Maria, Luis, Monique, Sierra, John, and Karen there would be 3 names with 4 letters each, 3 names with 5 letters each, 1 name with 6 letters and 1 name with 7 letters.

This data set could be represented with stacking cubes.



Students can model the mean by "leveling" the stacks or distributing the blocks so the stacks are "fair". Students are seeking to answer the question "If all of the students had the same number of letters in their name, how many letters would each person have?"

One block from the stack of six and two blocks from the stack of 7 can be moved down to the stacks of 4 and then all the stacks have five blocks. If all students had the same number of letters in their name, they would have five letters. The mean number of letters in a name in this data set is 5.



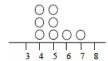
If it was not possible to make the stacks exactly even, students could begin to consider what part of the extra blocks each stack would have.

Understanding Mean Absolute Deviation

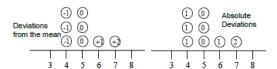
The use of mean absolute deviation in 6th grade is mainly exploratory. The intent is to build a deeper understanding of variability. Students would understand the mean

distance between the pieces of data and the mean of the data set expresses the spread of the data set. Students can see that the larger the mean distance, the greater the variability. Comparisons can be made between different data sets.

In the previous data set, the names drawn were Carol, Mike, Maria, Luis, Monique, Sierra, John, and Karen. There were 3 names with 4 letters each, 3 names with 5 letters each, 1 name with 6 letters and 1 name with 7 letters. This data can be represented on a dot plot. The mean of the data set is 5.



To find the mean absolute deviation, students examine each of the data points and its difference from the mean. This analysis can be represented on the dot plot itself or in a table. Each of the names with 4 letters has one fewer letter than the mean, each of the names with 5 letters has zero difference in letters as compared to the mean, each of the names with 6 letters has one more letter than the mean, and each of the names with 7 letters has two more letters than the mean. The absolute deviations are the absolute value of each difference.



Name	Number of letters in	Deviation from	Absolute Deviation
	a name	the Mean	from the Mean
John	4	-1	1
Luis	4	-1	1
Mike	4	-1	1
Carol	5	0	0
Maria	5	0	0
Karen	5	0	0
Sierra	6	+1	1
Monique	7	+2	2
Total	40	0	6

The mean of the absolute deviations is found by summing the absolute deviations and dividing by the number of data points. In this case, the mean absolute deviation would be $6 \div 8$ or $\frac{3}{4}$ or 0.75. The mean absolute deviation is a small number, indicating that there is little variability in the data set.

Consider a different data set also containing 8 names. If the names were Sue, Joe, Jim, Amy, Sabrina, Monique, Timothy, and Adelita. Summarize the data set and its variability. How does this compare to the first data set? (continued on the next page)

The mean of this data set is still 5.

$$\frac{(3+3+3+3+7+7+7)}{8} = \frac{40}{8} = 5$$

Name	Number of letters in	Deviation from	Absolute Deviation
	a name	the Mean	from the Mean
Sue	3	-2	2
Joe	3	-2	2
Jim	3	-2	2
Amy	3	-2	2
Sabrina	7	+2	2
Timothy	7	+2	2
Adelita	7	+2	2
Monique	7	+2	2
Total	40	0	16

Consider a different data set also containing 8 names. If the names were Sue, Joe, Jim, Amy, Sabrina, Monique, Timothy, and Adelita. Summarize the data set and its variability. How does this compare to the first data set?

The mean of this data set is still 5.

$$\frac{(3+3+3+3+7+7+7)}{8} = \frac{40}{8} = 5$$

Name	Number of letters in	Deviation from	Absolute Deviation
	a name	the Mean	from the Mean
Sue	3	-2	2
Joe	3	-2	2
Jim	3	-2	2
Amy	3	-2	2
Sabrina	7	+2	2
Timothy	7	+2	2
Adelita	7	+2	2
Monique	7	+2	2
Total	40	0	16

The mean deviation of this data set is $16 \div 8$ or 2. Although the mean is the same, there is much more variability in this data set.

Understanding Medians and Quartiles

Students can also summarize and describe the center and variability in data sets using the median and a five number summary consisting of the minimum, quartiles, and maximum as seen in the box plot example in 6.SP.4. The median is the middle number of the data set with half the number below the median and half the numbers above the median. The quartiles partition the data set into four parts by dividing each of the halves of the data set into half again. Quartile 1 (Q1 or the lower quartile) is the middle value of the lower half of the data set and quartile 3 (Q3 or the upper quartile) is the middle value of the upper half of the data set. The median can also be referred to as quartile 2 (Q2). The range of the data is the difference between the minimum and maximum values. The interquartile range of the data is the difference between the lower and upper quartiles (Q3 – Q1). The interquartile range is a measure of the dispersion or spread of the data set: a small value indicates values that are clustered near the median whereas a larger value indicates values that are more distributed.

Consider the first data set again. Recall that the names drawn were Carol, Mike, Maria, Luis, Monique, Sierra, John, and Karen. The data set can be represented in a numerical

list. To find the median and quartile, the values are placed in order from least to greatest.

The middle value in the ordered data set is the median. If there are an even number of values, the median is the mean of the middle two values. In this case, the median would be 5 because 5 is the average of the 4^{th} and 5^{th} values which are both 5. Students find quartile 1 (Q1) by examining the lower half of the data. Again there are 4 values which is an even number of values. Q1 would be the average of the 2^{nd} and 3^{rd} value in the data set or 4. Students find quartile 3 (Q3) by examining the upper half of the data. Q3 would be the average of the 6^{th} and 7^{th} value in the data set or 5.5. The mean of the data set was 5 and the median is also 5, showing that the values are probably clustered close to the mean. The interquartile range is 1.5 (5.5 – 4). The interquartile range is small, showing little variability in the data.



What the students do:

- Perform a statistical investigation, including the collection, organization, and analysis of the data. Analysis should include the appropriate statistics from mean, median, interquartile range, measures of center, measures of variability, data, mean absolute deviation, quartiles, lower quartile (first quartile or Q1), and upper quartile (third quartile or Q3)
- Calculate mean absolute deviation (MAD) for data set and explain that MAD is the mean of the absolute values of the differences of each point in the data set from the man of the data set.
- Communicate a deep understanding of (1) observations, (2) appropriate measure of center and spread for a particular data set, (3) appropriate section of a graph to represent data collected, and (4) overall patterns in a distribution, including outliers, through statistical investigation.

Misconceptions and Common Errors:

Mean absolute deviation (MAD) is problematic for many students both procedurally and conceptually. Although the computation is clear, students forget steps, confuse the steps and so on. To help with the procedure, it is useful to create a three-column table with the following headings: data, distance from the mean, and absolute value of distance from the mean. For example, if the mean of the following set of data is 60, a table to find he MAD would look like this:

Data Points	Distance from Mean	Absolute Value of Distances
65	+5	5

	- - - - - - - - - -					
	58	-2	2			
	60	0	0			
Standards for Mathematical Practice	From the table, students need to only calculate the mean of the values in column 3. The concept is also very difficult at this age level. Be aware of this and allow adequate time to develop the concept of mean absolute deviation. This will make a difference with student understanding and application. Many of the challenges students often have will be avoided if students deeply understand the concept of MAD.					
Standards for Mathematical Fractice	Explanations and Exa	inipies				
Develop understanding of statistical variability. 6.SP.1, 6.SP.2, 6.SP.3 At the sixth-grade level, students learn how to write statistical questions used to survey and collect data. They study the key idea that data distribution may not have a definite center. Students discover that different ways to measure center produce different values and that interpreting measures of center for the same data develops the understanding of how each measure can change how the data get interpreted.						
MP3. Construct viable arguments and critique the reasoning of others.	Students write and share their cexplain their learning to others	· · · · · · · · · · · · · · · · · · ·		t data from other classmates. They		
MP4. Model with mathematics.	Students use measures of center sets.	rs and variability and data displ	ays to draw inferences about a	nd make comparisons between data		
MP6. Attend to precision.		•	hematical language when descr	ribing and explaining the connection		
Summarize and describe distributions. 6.SP.4, 6.SP.5	between different representation	ons of data sets.				
Students display data on line plots, histograms, and box plots with the focus of summarizing and describing distributions. Students summarize numerical data by providing background information about the attribute being measured, methods and unit of measurements,						
and the context of data collection activities, including random samples, the number of observations, and summary statistics.						
MP4. Model with mathematics.						
MP6. Attend to precision.						

Students display data graphically in a format appropriate for a particular set of data.
Students communicate precisely with others and use clear mathematical language when summarizing, describing, and reading data from graphs generated by other students.

K-L	J-D
KNOW Facts, formulas, information, vocabulary	DO Skills of the discipline, social skills, production skills, processes (usually verbs/verb phrases)
 Measure of center Median Mean Measure of variation Range Interquartile range Mean absolute deviation Numerical data sets Observations Attributes Overall pattern Deviations from Choice of measures of center and variability Shape of the data distribution Center Spread Variability Statistical question Numerical data displays 	Determine and utilize appropriate measures of center and variability for a data set to analyze it
Number line	

Dot plot
 Histogram
 Box plot

UNDERSTAND

Big ideas, generalizations, principles, concepts, ideas that transfer across situations

- RECOGNIZE (measure of center and measure of variation)
- SUMMARIZE (numerical data sets)
- REPORT (observations)
- DESCRIBE (attribute)
- GIVE/FIND (measure of center and measure of variation)
- DESCRIBE (overall pattern)
- RELATE (choice of measure to shape of the data)
- RECOGNIZE (a statistical question)
- UNDERSTAND (data distribution is described by its center, spread, and overall shape)
- DISPLAY (numerical data)

Common Student Misconceptions for this Unit

- Students may associate mode with "most" as in the greatest number instead of the number that appears most often.
- Students may know that median is the number in the middle but may not put the numbers in order.
- Students may misrepresent intervals when they create or read graphs.
- Students may ignore or forget to divide their calculated sum of values by the total numbers of values in a data to set to find mean.
- Students may not understand "distance" when finding mean absolute deviation refers to an individual range of two values in a data set.
- Students may not understand mean absolute deviation as a measure of how far, on average, each value is from the mean of a set of data.

Unit Assessment/Performance Task	DOK

Unit 6 Test

Unit 6 Performance Task Priority "What is the Center"

Unit 6 Performance Task Optional "Time Spent on Homework"

Unit 6 Performance Task Optional "Baseball Players"

Vocabulary

Box Plot

Center

Deviation

Distribution

Dot Plot

Extreme Values

Histogram

Inter-quartile Range

Lower-quartile

Maximum

Mean

Mean Absolute Deviation

Measure of Center

Measure of Variation

Median

Minimum

Observation

Sample

Sample Space

Shape

Spread

Upper-quartile

Variability

Key Learning Activities/Possible Lesson Focuses (order may vary)

The following activities are broken into "lessons," even though each may take more or less than one class period depending on school schedule.

These are ideas for lessons. Teachers can refer to lesson outline in Math In Focus Chapters 13 and 14. The Math In Focus Performance Tasks for Chapters 13 and 14 will be used as classroom activities.

Pre-assessment (Recall prior knowledge) and Pre-requisite skills review (if needed)

Lesson 1: Line Plots

Dot plots (line plots) are simple plots on a number line where each dot represents a piece of data in the data set. Dot plots are suitable for small to moderate size data sets and are useful for highlighting the distribution of the data including clusters, gaps, and outliers. A line plot uses x's to represent a piece of data in the data set.

Examples:

• Nineteen students completed a writing sample that was scored using the six traits rubric. The scores for the trait of organization were 0, 1, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 6, 6. Create a data display. What are some observations that can be made from the data display?



Several observations can be made from the line plot. Some of them include, there is a range (spread of numbers) of 6. The median and mode of the scores of the writing sample are 3. The line plot resembles a bar graph. There are 19 x's in the display that represent each of the

students. Four students have a score below 3.

Lesson 2: Frequency Distribution

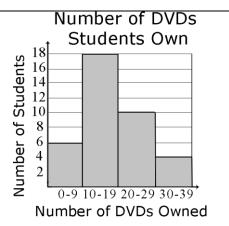
In most real data sets, there is a large amount of data and many numbers will be unique. A graph (such as a dot plot) that shows how many ones, how many twos, etc. would not be meaningful; however, a histogram can be used. Students organize the data into convenient ranges and use these intervals to generate a frequency table and histogram. Note that changing the size of the range changes the appearance of the graph and the conclusions you may draw from it. There is no space between bars and between numbers along the horizontal axis. There is no overlap of numbers along the horizontal axis, as well.

Examples:

Grade 6 students were collecting data for a math class project. They decided they would survey the another grade 6 class to determine how many DVDs each student owns. A total of 38 students were surveyed. The data are shown in the table below in no specific order. We want to create a data display. We would begin by creating a tally chart with a trial of a range of values (0-4;5-9;10-14; etc). After completing the tally chart and finding the frequency of the values, you may choose to adjust the range.

11	21	5	12	10	31	19	13	23	33
10	11	25	14	34	15	14	29	8	5
22	26	23	12	27	4	25	15	7	
2	19	12	39	17	16	15	28	16	

This histogram using 5 ranges (0-9, 10-19, ...30-39) to organize the data is displayed below.



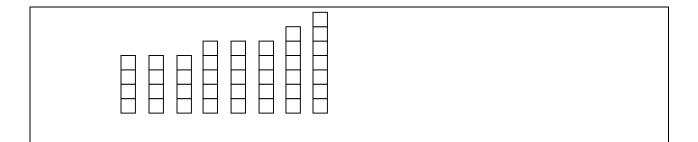
What are some observations that can be made from the data display? Students can discuss in small groups come up with several observations to report to the class on.

Lesson 3: Mean

The mean measures center in the sense that it is the value that each data point would take on if the total of the data values were redistributed equally. Students develop understanding of what the mean represents by redistributing data sets to be level or fair. The leveling process can be connected to and used to develop understanding of the computation of the mean.

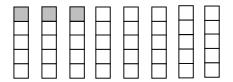
Example- Finding mean by distributing blocks (linking cubes work well):

Students could generate a data set by measuring the number of jumping jacks they can perform in 5 seconds, the length of their feet to the nearest inch, or the number of letters in their names. It is best if the data generated for this activity are 5 to 10 data points which are whole numbers between 1 and 10 that are easy to model with counters or stacking cubes. Students generate a data set by drawing eight student names at random from the popsicle stick cup. The number of letters in each of the names is used to create the data set. If the names drawn were Carol, Mike, Maria, Luis, Monique, Sierra, John, and Karen there would be 3 names with 4 letters each, 3 names with 5 letters each, 1 name with 6 letters and 1 name with 7 letters. This data set could be represented with stacking cubes.



Students can model the mean by "leveling" the stacks or distributing the blocks so the stacks are "fair". Students are seeking to answer the question "If all of the students had the same number of letters in their name, how many letters would each person have?"

One block from the stack of six and two blocks from the stack of 7 can be moved down to the stacks of 4 and then all the stacks have five blocks. If all students had the same number of letters in their name, they would have five letters. The mean number of letters in a name in this data set is 5. (Since there were 8 names you must have 8 stacks).



If it was not possible to make the stacks exactly even, students could begin to consider what part of the extra blocks each stack would have.

Example- Finding mean by calculation:

Kamil earned \$35, \$38, \$26, \$12, and \$14 baby-sitting last month. What is the mean of the amounts she earned?

(Solution: \$25. A review of calculating mean is important here, especially if using a calculator. Make sure that students get a sum of the numbers before they divide. It is important for students to understand that the sum of the numbers is then divided by the amount of numbers added to equal the mean or average.)

Math Test Scores						
90	86	96	89			
100	65	73	85			
71	70	75	80			

Using the Math Test Scores in the table above, find the **mean** or average, of the 12 test scores.

Mean _____ (Solution: 81.7)

If an additional score of 100 was put in with the other math scores, would the mean of the math scores change? Explain your answer. (Yes it would change and increase. Students should understand that the reason it would increase is because 100 is higher than the mean. The actual increase is to 83.1)

Example- using both methods:

A data set consists of the number 1, 6, 6, 2 and 8. What is the mean of the data set? There is a total of 23 blocks.

Represent the numbers with blocks. Make even stacks of blocks. There are 3 stacks of 5 and 2 stacks of 4, so the mean will come between 4 and 5. 3 out of 5 stacks have 5 blocks which can

be written as $\frac{3}{5} = \frac{6}{10}$ which is 6 tenths. Therefore the mean is 4.6

To calculate the mean: Add the numbers, 1, 6, 6, 2 and 8 and get a sum of 23. Divide 23 by 5 and the result is a mean of 4.6

Lesson 4: Interpreting Measures of Center

Consider the data shown in the dot plot of the six trait scores for organization for a group of students.

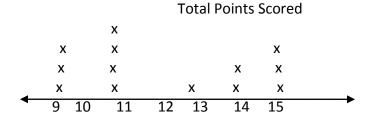
- How many students are represented in the data set? (Solution: 19-count the x's)
- What are the mean, median, and mode of the data set? What do these values mean? How do they compare?

(Solutions: Mean-3.5; median-3; mode-3. These values are the measures of center

- (central tendency) of the data. Various answers.)
- What is the range of the data? What does this value mean? (Solution: 6-highest score minus lowest. There is a spread of 6 scores. Range is a measure of variation, not of center, but it is often used in conjunction with the measures of center.)



• Find the median, mode and range of the total points scored that are represented in the line plot. (Solution: median is 11; mode is 11; range is equal to 4 (14 minus 10); Subtract the lowest value with data, 10, from the highest value with data, 14.)



Objective: I can solve a problem by collecting, organizing, displaying, and interpreting data.

Explanations:

Statistics are numerical data relating to an aggregate of individuals; statistics is also the name for the science of collecting, analyzing and interpreting such data. A statistical question

anticipates an answer that varies from one individual to the next and is written to account for the variability in the data. Data are the numbers produced in response to a statistical question. Data are frequently collected from surveys or other sources (i.e. documents).

Questions can result in a narrow or wide range of numerical values. For example, asking classmates "How old are the students in my class in years?" will result in less variability than asking "How old are the students in my class in months?"

Students might want to know about the fitness of the students at their school. Specifically, they want to know about the exercise habits of the students. So rather than asking "Do you exercise?" they should ask about the amount of exercise the students at their school get per week. A statistical question for this study could be, "How many hours per week on average do students at Jefferson Middle School exercise?"

To collect this information, students might design a survey question that anticipates variability by providing a variety of possible anticipated responses that have numerical answers, such as: 3 hours per week, 4 hours per week, and so on. Be sure that students ask questions that have specific numerical answers.

Examples of statistical questions:

- How much time in minutes do you spend listening to music?
- How many video games do you have?
- What are the heights of the students in the class in inches?

Examples of questions that are not statistical questions:

- What are the names of the teachers in your school?
- What is the title of the book you are reading?

Objective: I can find the measures of variation of a set of data. Explanations:

• Measures of variation are used to describe the distribution of the data. Range, the

difference between the greatest and least data values, is the one most commonly familiar to students. Quartiles are values that divide the set of data into 4 equal parts.

- To find the quartiles you first find the median of the data. Then you find the median of the lower half of the data set (this is the lower quartile). Next you find the median of the upper half of the data set (this is the upper quartile). When you find the difference between the upper quartile and lower quartile, this is called the interquartile range (IQR).
- An outlier is a data value that is either much greater or much less than the rest of the data. (Technically, if a data value is more than 1.5 times the value of the interquartile range beyond the quartiles, it is an outlier).

Example

• The Miami Heat scored 76, 90, 88, 116, 77, 82, and 90 points in their seven game series. Find the measures of variation of the scores. Is there an outlier in the data set?

Solution: First find the median by listing the scores in order from least to greatest. **76, 77, 82, 88, 90, 90, 116** The middle number, or median is 88

The range is 116-76, or 40

The interquartile range (IQR) is 90-77, or 13.

The outlier is 116. It is much greater than the other data values. (13 x 1.5 = 19.5; 90 + 19.5 = 109.5, so any value greater than this value would be considered an outlier. Or, 77 – 19.5 = 57.5, so any value less than this value is also considered an outlier.)

Objective: I can display and interpret data in box-and-whisker plots.

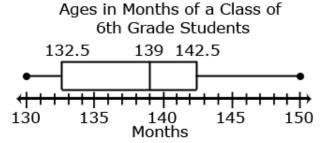
Explanation:

 Box plots are another useful way to display data and are plotted horizontally or vertically on a number line. Box plots are generated from the five number summary of a data set consisting of the minimum value, maximum value, median, and two quartile values. Students can readily compare two sets of data if they are displayed with side by side box plots on the same scale. Box plots display the degree of spread of the data and if the data is skewed.

Examples:

• Ms. Wheeler asked each student in her class to write their age in months on a sticky note. The 28 students in the class brought their sticky note to the front of the room and posted them in order on the white board. The data set is listed below in order from least to greatest. Create a data display. What are some observations that can be made from the data display?

130	130	131	131	132	132	132	133	134	136
137	137	138	139	139	139	140	141	142	142
142	143	143	144	145	147	149	150		



Observations include: this box plot shows that ¼ of the students in the class are from 130 to 132.5 months old ½ of the class are from 132.5 to 142.5 months old. ¼ of the students in the class are from 142.5 months to 150 months old. The median class age is 139 months.

The five number summary determined from the box plot above:

Minimum, also known as lower extreme: 130 months

Quartile 1 (Q1): $(132 + 133) \div 2 = 132.5$ months

Median (Q2): 139 months

Quartile 3 (Q3): $(142 + 143) \div 2 = 142.5$ months

Maximum, also known as upper extreme: 150 months

Supplemental Materials and Resources

<u>Teaching the Common Core Math Standards with Hands-On Activities</u> (Jossey-Bass): "Statistical Questions vs. Non-statistical questions", "And the answer is..", "Measure of center versus measure of variation", "Creating Data Displays", and "Summarizing Data"

Illuminations: http://illuminations.nctm.org/Search.aspx?view=search&st=d&gr=6-8

Common Core Mathematics (Newmark Learning)

Common Core Standards for Math (Steck-Vaughn)

Brain Pop

LearnZillion

Literature Connections

- -Have students explore newspapers and magazines to find real-world examples of data.
- -The Inch Boy by Junko Morimoto
- -Swamp Angel by Anne Isaacs
- -<u>Tiki, Tiki, Tembo</u> by Arlene Mosel

Interdisciplinary Connections

- -Students will need to be able to read different graphs that they come across in different subject areas.
- -Students will need to create graphs and interpret data in Science.

Graph paper Rulers Number lines Unifix cubes

Suggested Formative Assessment Practices/Processes

Teacher created exit slips, teacher created quizzes

Exit Slins: Give students a set of data and ask them to determine measures of center (example: Differentiation and Accommodations

the most sense for that particular data set.

- Provide additional examples and opportunities for repetition
- Provide tutoring opportunities
- Provide retesting opportunities after remediation (up to teacher and district discretion)
- Teach for mastery not test
- Teaching concepts in different modalities
- Adjust homework assignments