



Lesson Overview

In this TI-Nspire lesson, students associate distributions with appropriate measures of center and spread. They investigate the difference between histograms and bar graphs. Finally they explore the differences in the information given by dot plots, histograms, and boxplots.



The tools that are useful in the analysis and visualization of data depend on the distribution and type of data.

Learning Goals

1. Associate a graphical representation of a set of data with measures of center and spread;
2. compare different measures of center and spread for a given distribution;
3. identify the advantages and disadvantages of different graphical representations of the same data;
4. recognize the difference between bar graphs and histograms.

Prerequisite Knowledge

Analyzing Distributions is the ninth lesson in a series of lessons that investigates the statistical process. In this lesson, students analyze data represented on dot plots, bar graphs, and histograms. This lesson builds on the concepts of the previous lessons. Prior to working on this lesson students should have completed *Introduction to Data* and *Introduction to Histograms*. Students should understand:

- how to read a bar graph;
- how to interpret data represented on box plots, dot plots, and histograms;
- how to find measures of center and spread.

Vocabulary

- **symmetric:** when one side is the exact image or reflection of the other
- **skewed:** data that clusters towards one end of a graphical display
- **mound shaped:** data that clusters towards the middle of a graphical display
- **bimodal:** data distribution having two equal, most common values
- **outlier:** a value that lies outside most of the other values in a set of data
- **median:** the value that separates the upper half of the distribution of a set of data values from the lower half
- **mean:** the sum of all the data values in a set of data divided by the number of data values
- **interquartile range:** the difference between the upper quartile and the lower quartile
- **mean absolute deviation:** the mean of the absolute values of all deviations from the mean of a set of data

Lesson Pacing

This lesson should take 50–90 minutes to complete with students, though you may choose to extend, as needed.



Lesson Materials

- Compatible TI Technologies:



TI-Nspire CX Handhelds,



TI-Nspire Apps for iPad®,



TI-Nspire Software

- Analyzing Distributions_Student.pdf
- Analyzing Distributions_Student.doc
- Analyzing Distributions.tns
- Analyzing Distributions_Teacher Notes
- To download the TI-Nspire activity (TNS file) and Student Activity sheet, go to <http://education.ti.com/go/buildingconcepts>.

Class Instruction Key

The following question types are included throughout the lesson to assist you in guiding students in their exploration of the concept:



Class Discussion: Use these questions to help students communicate their understanding of the lesson. Encourage students to refer to the TNS activity as they explain their reasoning. Have students listen to your instructions. Look for student answers to reflect an understanding of the concept. Listen for opportunities to address understanding or misconceptions in student answers.



Student Activity: Have students break into small groups and work together to find answers to the student activity questions. Observe students as they work and guide them in addressing the learning goals of each lesson. Have students record their answers on their student activity sheet. Once students have finished, have groups discuss and/or present their findings. The student activity sheet can also be completed as a larger group activity, depending on the technology available in the classroom.



Deeper Dive: These questions are provided for additional student practice and to facilitate a deeper understanding and exploration of the content. Encourage students to explain what they are doing and to share their reasoning.



Mathematical Background

As in earlier work, students should view statistical reasoning as a four-step investigative process. Steps one and two relate to posing a question and collecting data to answer the question; this lesson focuses on step three, analyzing data with appropriate methods. Students use the terms (*skewed, symmetric, mound shaped*) from previous lessons such as, *Introduction to Data* and *Introduction to Histograms* and investigate how the shape of a distribution might affect the choice of summary measures of center and spread. They investigate whether the mean or the median is a more useful measure of center, taking into account distributions with very long tails. Similar thinking guides their choice of measures of spread as either the interquartile range (IQR) or the mean absolute deviation (MAD). Symmetry and lack of symmetry can be useful as a way of choosing summary measures for center and spread. Students should consider that a distribution of data that is mound shaped and symmetric, suggests the mean and mean \pm MAD are reasonable measures to use for summarizing the data. But when a distribution is skewed, the median and the interval associated with the IQR will typically give a better picture of the data because the values in the tail will have large deviations from the mean, and thus may increase the size of the mean beyond what is usual for most of the data. Encourage students to think about the MAD in terms of deviations (distance) from the mean, while the IQR is based on order alone not on the weight of the actual values in the distribution.

The lesson also involves bar graphs. Students should recognize that in a bar graph, bars represent quantitative measures associated with categorical data—such as favorite colors (red, blue, green), participation in sports (football, basketball, tennis), gender (boys, girls). Bar graphs differ from histograms in that the bars in histograms represent frequencies associated with quantitative values—such as the number of hours watching television, miles per gallon, life expectancy—where the height of the bar is the frequency of data values in that bar. Bar graphs indicate the value for an individual entry in a category (e.g., number of cars of a certain color) and can be moved and arranged in any order (least to smallest, alphabetical), while histograms cannot because histograms must be positioned relative to a number line. Bar graphs can have any amount of space between the bars; histograms can only have empty spaces between bars if no data values occur in the interval represented by these spaces on the number line.

Students tend to graph every data set using a bar graph, perhaps because of a reluctance to abandon an individual piece of information rather than looking for overall trends and patterns in a distribution where the names of individuals are not present. (A common misconception is thinking they have to see which height is Suzie's or which is Jim's; you might want to revisit the last section of Lesson 5, *Mean as Balance Point*, where the transition from picture graphs to dot plots is visualized.) Bar graphs are not easily used for finding measures of center and spread or for recognizing the shape of a distribution.



Data Sources:

- *Source: Natural History Magazine, March 1974, copyright 1974; The American Museum of Natural History; and James G. Doherty, general curator, The Wildlife Conservation Society;*
<http://www.thetravelalmanac.com/lists/animals-speed.htm>;
- <http://www.demogr.mpg.de/longevityrecords/0203.htm>
- http://web.stanford.edu/group/stanfordbirds/text/essays/How_Long.html
- <http://www.vetstreet.com/our-pet-experts/meet-eight-of-the-fastest-dogs-on-the-planet;>
<http://canidaepetfood.blogspot.com/2012/08/which-dog-breeds-are-fastest.html>
- <http://slimdoggy.com/dog-life-spans/>
- <http://users.pullman.com/lostriver/breeddata.htm>
- <http://visual.ly/shark-speed-worlds-fastest-sharks><http://www.thetravelalmanac.com/lists/fish-speed.htm>
- http://www.elasmo-research.org/education/topics/r_haulin'_bass.htm
- <http://whaleopedia.org/animalfund/harbor-porpoise/>
- <http://www.enchantedlearning.com/subjects/whales/species/>
- http://www.panda.org/what_we_do/endangered_species/cetaceans/about/fin_whale/
- http://www.speedofanimals.com/animals/sea_otter
- <http://www.petcarerx.com/article/the-average-lifespan-of-a-cat-breed-by-breed-chart/1698>
- <http://www.petcarerx.com/article/the-average-lifespan-of-a-cat-breed-by-breed-chart/1698>
- <http://www.dfo-mpo.gc.ca/Library/333800.pdf>
- <http://www.infoplease.com/ipa/A0004737.html>
- <http://a-z-animals.com/animals/donkey/>
- <http://a-z-animals.com/animals/galapagos-giant-tortoise/>
- http://en.wikipedia.org/wiki/Fastest_animals



Part 1, Page 1.3

Focus: Students associate distributions of data with appropriate measures of center (median or mean) and spread (interquartile range or mean absolute deviation).

On page 1.3, the two graphs represent the shoulder heights in centimeters of a collection of domestic dogs.

Outliers displays or does not display outliers.

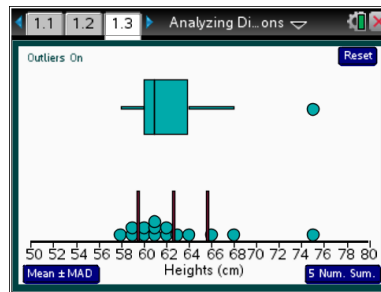
5 Num Sum. displays the values for the five number summary.

Mean +/- MAD displays the values for the mean and the mean +/- MAD.

Segments highlights the points on the number line associated with that segment.

Points can be moved using the arrow keys or by dragging to a new location.

Reset returns to the original screen.



TI-Nspire Technology Tips

menu accesses page options.

tab cycles through points in the dot plot or through the summary measures and buttons.

Up/Down arrow control what tab selects.

enter selects a highlighted point.

esc releases all selected points and measures.

ctrl del resets the page.



Class Discussion

The following questions focus on the relationship between the shape of a distribution of data and the summary measures for the data.

On page 1.3, the two graphs represent the shoulder heights in centimeters of a collection of domestic dogs.

- **Write down three things you observe about the distribution of heights.**

Answers will vary. Possible responses: the range of heights is about 18 centimeters, from about 57 centimeters to 75 centimeters; the dog that is 75 centimeters tall (a Great Dane) is an outlier, much taller than the other dogs; most of the dogs are between 59 and 62 centimeters tall; the distribution seems to be skewed right; the distribution shows the heights of 15 dogs.



Class Discussion (continued)

Have students...

- **How do you think the mean and median will compare? Explain your thinking.**
- **Select the vertical segment in the middle of the dot plot to check your conjecture in part.**
- **How much does the height of the tallest dog deviate from the mean?**

Select 5 Num Sum and the other two vertical segments in the dot plot.

- **How does the interval associated with the IQR compare to the length of the interval from the mean + MAD to the mean – MAD (mean+/-MAD)?**
- **Select each region of the box plot. How many of the dogs have heights in each region?**

Look for/Listen for...

Answers will vary. The outlier will probably make the mean larger than the median because 75 centimeters is part of calculating the mean, but 75 could be any number above the median and not affect the value of the median.

Answer: The height of the tallest dog deviates from the mean by $75 - 62.6 = 12.4$ centimeters.

Answer: The IQR = $64 - 60 = 4$ centimeters, and the mean+/-MAD = $65.66 - 59.62 = 6.04$ centimeters. The mean+/-MAD is larger than the IQR, which is probably because the 75 centimeter tall dog is so much taller than the others.

Answer: The lower and upper tails each have the heights for three dogs; there are nine dog heights in the box including those on the edge.



Student Activity Questions—Activity 1

1. **Suppose the heights of the three tallest dogs had been entered wrong and they were all 10 centimeters too high.**
 - a. **Make a conjecture about how the IQR and the mean+/-MAD will change. Then move the three dots to the suggested heights to check your reasoning. (Select the white space to deselect a dot.)**

Answers will vary. Some may suggest that both get smaller; others that only the interval for the mean+/-MAD will get smaller. After the dots are moved, both the mean+/-MAD and the IQR get smaller because the heights are now all closer together.

- b. **Look at the distribution of heights after the heights were adjusted. Which of the following words would you use to describe the new distribution of heights: *skewed right*, *skewed left*, *symmetric*, *mound shaped*? Explain why.**

Answer: The new distribution of heights of the dogs is mound shaped and relatively symmetric because the heights are clustered between about 58 and 63 centimeters with about three dogs with heights from 56 to 59 centimeters and three dogs with heights from 62 to 65 centimeters, about three centimeters on either side of the cluster.



Student Activity Questions—Activity 1 (continued)

- c. **How do the mean and median of the adjusted distribution of the dog heights compare?**

Answer: They are close together; the mean is 60.6 centimeters, and the median is 61 centimeters.

- d. **Reset. Which of the following words would you use to describe the original distribution of the dog heights: *skewed right*, *skewed left*, *symmetric*, *mound shaped*? Explain your reasoning.**

Answer: The distribution of heights of the dogs is skewed right because the heights of the dogs are between about 58 and 64 centimeters except for three dogs, who have heights that go all the way up to 75 centimeters, making it skewed in the direction of the tail.

- e. **How do the mean and median of the distribution of dog heights compare?**

Answer: The mean height for the dogs is 62.6 centimeters, 1.6 centimeters larger than the median of 61 centimeters.

2. **Which of the following do you think are true statements about a distribution of data? Use your answers for the previous question to support your thinking in each case.**

- a. **If a distribution is skewed, the mean and the median will be close together.**
- b. **If a distribution is skewed, the median probably represents a better measure of the center than the mean.**
- c. **In a symmetric mound shaped distribution, the mean and the median will be close together.**
- d. **If a distribution is skewed, the measures that best summarize the data are the median and the IQR.**
- e. **If a distribution is mound shaped, the measures that best summarize the data are the mean and mean \pm MAD.**

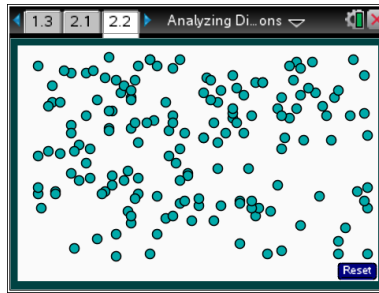
Answer: Statements b, c, and d are true, given the difference between the skewed distribution of heights and the mound shaped and symmetric distribution of height in question 1. Statement a is not true because a skewed distribution will have either some values larger than all of the others or smaller than all of the others. These values will make the means larger or smaller because their absolute deviations will be large. Statement e is not true because the distribution could be mound shaped at one end, which is kind of the case in question 1. When the distribution is both mound shaped and symmetric, the mound will be in the center of the distribution making the mean and median close together.



Part 2, Page 2.2

Focus: Students identify which measures of center and spread might be most appropriate for different distribution shapes.

On page 2.2, data used are the maximum recorded speeds and the longest recorded life spans of animals. Selecting up to six points will display the type of animal associated with the point.



Type chooses the animal type.

Attribute chooses between maximum speed or life span.

Graph type chooses among dot plot, box plot, histogram with bins of 1, 5, 10, or 20 units, and a bar graph.

Summary measurements shows median and IQR or mean \pm MAD on the screen.

TI-Nspire Technology Tips

menu accesses page options.

tab cycles through the points in the dot plot.

enter selects up to six points.

ctrl del resets the page.



Class Discussion

The following questions ask students to apply what they learned in Part 1 to different contextual situations. The tools they have to use include choices for different graphical representations of the data and the ability to display and move a segment representing the IQR and median and a segment representing the mean and the mean \pm MAD.

The data used for page 2.2 are the maximum recorded speeds of animals and the longest recorded life spans from Introduction to Data.

Select menu> Type> All Animals and menu> Attribute> Max. Speed.

- **How would you describe the distribution of maximum-recorded speeds of all animals? Select menu> Graph Type> Dot plot and then examine Graph Type> Histogram with bin width of 5 mph to confirm your thinking. (Note that selecting a dot will show the animal associated with the dot.)**
- **Return to Dot plot. Select three dots representing three animals you think will have speeds typical for most of the other land animals in the data.**

Answer: The dot plot and histogram suggest the speeds are skewed to the right with a few types of animals having faster maximum speeds than is typical for most of them with one animal, the falcon, clearly being an outlier.

Answers will vary. Students will probably select animals from or near the tallest column. They might mention zebra, great white shark, red tailed hawk, camel—all with maximum recorded speeds around 47 mph.



Class Discussion (continued)

- ***How do you think the median/IQR and the mean and mean \pm MAD will compare?***

Answers will vary. Some may think the mean and mean \pm MAD will be affected by the speed of the falcon, so the mean speed will be larger than the median speed.

Select menu> Summary Measurements> both to check your answer to the question above.

- ***Without a key, how could you tell which of the two segments is around the mean and which is around the median? Explain your reasoning.***
- ***Were you surprised when you saw the summary measures? Why or why not?***

Answer: The brown segment is around the mean because the segment is divided into two equal parts by the center value, which is what would happen when adding and subtracting the MAD from the mean.

Answers will vary. The IQR was 20.4 mph, and the interval determined by the mean \pm MAD was a lot larger, at 35.4 mph (twice the MAD), which is probably because of the speed of the falcon. The mean is larger at 37.5 mph than the median at 32 mph.

Reset. Use the menu to create a boxplot of the maximum-recorded speed for dogs.

- ***Describe the distribution of maximum-recorded speeds for dogs. Explain your reasoning. Then select another plot to see if a different representation supports your thinking.***
- ***Estimate the mean and median maximum recorded speeds. Give a reason for your estimates.***
- ***Estimate the IQR and the mean \pm MAD. Give a reason for your estimates.***
- ***Select menu> Summary Measuresm> both. How good were your estimates?***

Answer: The distribution of maximum-recorded speeds for dogs is relatively symmetric, although the segment to the right is a bit longer than the segment to the left. It could be mound shaped as well but cannot tell from a box plot. Using either a dot plot or a histogram shows that the distribution is fairly mound shaped as well as symmetric.

Answer: The mean and median maximum-recorded speeds are both about 30 mph because 30 is the center of the mound shape, it is the most common speed, and about half of the speeds are below or the same as 30 mph and half are above or the same.

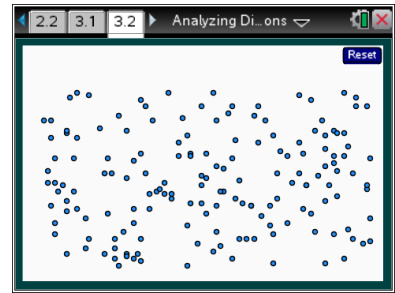
Answers may vary. Students might reason that both measures will be from 25 mph to 36 mph or 11 mph because that is where the mound in the center begins and ends.

Answers may vary. The mean and median speed for the dogs are both about 30 mph, but the length of the segment mean \pm MAD, 9.8 mph, is a bit larger than the IQR, 8.



Part 2, Page 2.2

Focus: Students interpret bar graphs and identify the difference between a bar graph and a histogram.



Student Activity Questions—Activity 2

These questions focus on categorical data represented in bar graphs. Students consider what they can learn from bar graphs of categorical data as well as what they cannot learn. (Note that no data are in the file for maximum-recorded speed of cats.)

1. Reset. Choose *all animals, life span, bar graph*.

- a. What category has the life span of the most types of animals? How do you know? (Note that dogs are not included in the domestic animal category and that hovering over a bar shows the frequency of the types of animals represented in the bar.)**

Answer: There were 35 types of birds, more types than any other kind of animals. You can tell because the bar is highest, which means the frequency is the largest.

b. Which of the statements is true? Explain why.

- i. The number of types of wild animals is more than twice as many as the number of types of domestic animals (excluding dogs).**

Answer: There are 31 types of wild animals and 14 types of domestic animals so there are more than twice as many types of wild animals.

- ii. The total number of types of fish and sea mammals is more than the total number of wild and domestic land animals (excluding dogs and cats).**

Answer: There are 41 types of fish and sea mammals and 45 types of wild and domestic land animals, so the statement is not true.

- iii. The difference between the number of types of birds and types of cats is more than the difference between the number of types of dogs and number of types of wild land animals.**

Answer: The difference between the number of types of birds and cats is 18 and the difference between the number of types of dogs and number of types of wild animals is 5, so the statement is true.



Student Activity Questions—Activity 2 (continued)

2. **What is the difference between a bar graph and a histogram? Use examples from the TNS activity to support your reasoning.**

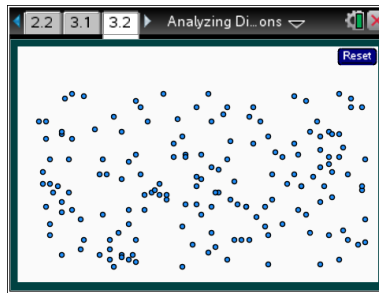
Answer (examples may vary): A bar graph shows the number of types for a category or kind of animal, while a histogram does not show the type but does show how many types have a specific maximum speed or life span located on the number line. For example, if the attribute is life span for all animals and the graph is a bar graph, there are 17 types of cats; while, from the all animals histogram with bin width 5, there are 21 types of animals with maximum life spans between 20 and 25 years. The bar graph allows you to compare the number of each type (there are more dog types than fish types in the data), while the histogram allows you to say something overall about the life spans of the animals – the distribution is skewed right with the most common life span (46 animal types) from 10 to 15 years and four animal types had a life span from 75 to 80 years.

Part 3, Page 3.2

Focus: Students consider what they can learn from different representations of the same data using a dot plot, a box plot and a histogram.

Page 3.2 functions in the same way as page 1.3, but the data are displayed simultaneously in three different graphs: box plot, histogram and dot plot.

Bin Width allows a change in bin width on the histogram. Students can select a point in the dot plot to display the name of the associated animal.



TI-Nspire Technology Tips

- menu** accesses page options.
- tab** cycles through the points.
- enter** selects points.
- esc** releases all selected points.
- ctrl del** resets the page.



Student Activity Questions—Activity 3

This last set of questions asks students to consider three representations of the same data: a dot plot, a box plot and a histogram. They compare what they can learn about a distribution of data from each representation and think about the advantages and disadvantages of each in helping understand the story in the data.

1. **Choose menu > Type > Birds, and menu > Attribute > Life Span.**
 - a. **How wide are the bins in the histogram? Explain how you know.**

Answer: The bins are 5 units—they go from 5 years up to and including 9 years, 10 years up to and including 14 years and so on. You can tell by looking at the scale and checking with the dot plot or by hovering over the bar. You know that 5 cannot be in both bins.



Student Activity Questions—Activity 3 (continued)

- b. **What can you learn from each of kind of plot that you cannot learn from the others? Use examples from part a. to support your thinking.**

Answers will vary. Students should note that the dot plot shows all of the life spans, which none of the other plots do (note the Amazon parrot has the longest recorded life span at about 104 years). The histogram gives a frequency for the number of birds with life span within a fixed interval (five birds that have a life span from 5 to 9 years) and allows you to compare the frequencies in these intervals by the height of the bins. The box plot allows you to get a good estimate of the LQ (about 15 years), UQ (about 33 years) and median (about 22 years); you could also do this on the dot plot, but it would involve a lot of counting.

- c. **What are the disadvantages of each plot?**

Answer: If there are a lot of animals, the dots get all piled on top of each other; the histogram groups values into the same bin so you lose the exact value, and sometimes you lose any gaps or clusters or the clusters show up in strange places because of the way the values go into the bins; the box plot loses all of the exact values and does not show gaps or clusters.

2. **Reset. Choose *menu*> *Type*> *all*, *menu*>*Attribute*> *Max Speed*. Set the bin width of the histogram to 10. Which plot seems to be the best for displaying the data? Explain your reasoning.**

Answer: The histogram or the boxplot seem to be the best; the histogram shows the shape of the distribution, and you can see the number of data values in each bin. It shows that the speed of the outlier, the peregrine falcon, is really far away from the other speeds. The boxplot shows the peregrine falcon as the maximum and how far away the other speeds are. The boxplot does show the IQR and median; you would have to estimate these from the histogram. The dot plot has a lot of dots, and they get all scrambled on top of each other.



Deeper Dive — page 1.3

Remember that a statistical question is one where the responses to the question will vary. Formulate a statistical question related to either the maximum-recorded speed or the maximum recorded life span of the collection of animals that was not considered in the exercises. Use the data and the tools in the file to answer your question.

Answers will vary.



 Deeper Dive — page 2.2

Identify the statements as true or false. Explain your reasoning in each case.

- ***If a distribution is skewed and has an outlier, the mean+/MAD will be greater than the IQR.***

Answer: True because the MAD will be affected by the values in the tail, especially the outlier, while the IQR will not be.

- ***If a distribution is symmetric and mound shaped, the mean should be a good measure of the center.***

Answer: True because the mean will be in the center of the mound and there would not be an outlier on one side that might influence the mean.

- ***The bins in a histogram can be exchanged as long as you keep the right height for the bar.***

Answer: No, because the bins in a histogram are fixed on the number line, and the frequency represents the values on the number line within that bin width.

- ***The bars in a bar graph can be moved as long as you keep the right height for the bar.***

Answer: True. The order does not make a difference because the bars are not fixed or attached to anything except the label for the category.

 Deeper Dive — page 3.2

Which plot would typically be a good choice for each:

- ***to display a measure of center and spread***

Answer: box plot

- ***if you have a large number of data values***

Answer: histogram

- ***if you want to see the gaps and clusters***

Answer: dot plot or maybe histogram

- ***the shape of a distribution***

Answer: dot plot or histogram

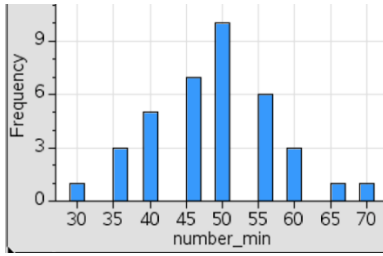


Sample Assessment Items

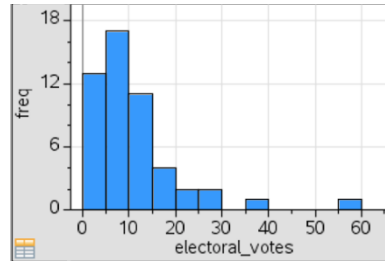
After completing the lesson, students should be able to answer the following types of questions. If students understand the concepts involved in the lesson, they should be able to answer the following questions without using the TNS activity.

1. Which summary measures should be used to describe the distributions: mean and mean \pm MAD or median and IQR?

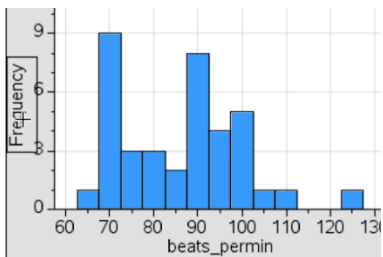
a. Time studied for test



b. Number of electoral votes by states



c. Heart beats per minute for middle school students



Answer: a. mean and mean \pm MAD; b. median and IQR; c. could be either



2. The table shows the number of customers at Malcolm's Bike Shop for 5 days, as well as the mean (average) and the median number of customers for these 5 days.

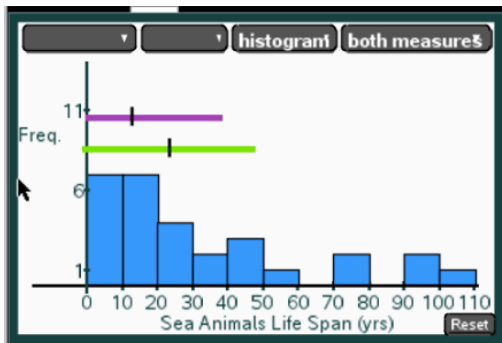
Number of customers at Malcolm's Bike Shop	
Day 1	100
Day 2	87
Day 3	90
Day 4	10
Day 5	91
Mean (average)	75.6
Median	90

Which statistic, the mean or the median, will be more typical of the number of customers at Malcolm's Bike Shop for these 5 days? Explain your reasoning.

NAEP 2007 grade 8

Answer: The median because the number of customers on Day 4 (10) is an outlier. Maybe the store was only open half the day, there was a big storm, or something.

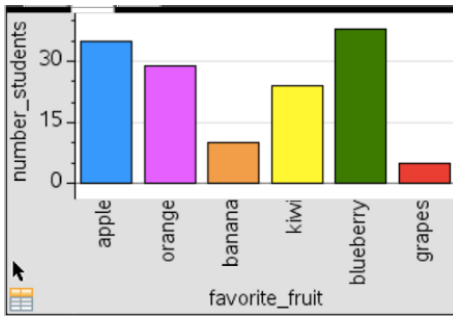
3. Which of the segments represents the median and IQR?



Answer: The shorter segment that goes from 0 to just less than 40 because the distribution is skewed, the median is not in the center of the segment



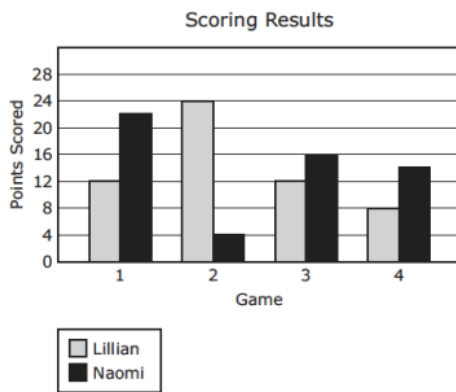
4. The bar graph below displays the favorite fruit of students in a sixth grade class.



Which fruit do most of the students prefer?

Answer: Blueberries

5. The number of points scored by Lillian and Naomi during four basketball games is shown in the graph below. Which statement is best supported by the information in the graph?



- a. In Game 1 the number of points scored by Lillian was more than half the number of points scored by Naomi.
- b. The total number of points scored by Lillian and Naomi in Game 4 was more than the number of points scored by Lillian in Game 2.
- c. In Game 4 the number of points scored by Naomi was two times the number of points scored by Lillian.
- d. The total number of points scored by Lillian and Naomi in Game 3 was seven times the number of points scored by Lillian in Game 2.

Adapted from Texas Grade 7 STAAR, 2013

Answer: a. In Game 1 the number of points scored by Lillian was more than half the number of points scored by Naomi.



Student Activity Solutions

In these activities you will compare different measures of center and spread for a given distribution, interpret bar graphs, and compare different representations of the same data. After completing the activities, discuss and/or present your findings to the rest of the class.



Activity 1 [Page 1.3]

1. Suppose the heights of the three tallest dogs had been entered wrong and they were all 10 centimeters too high.
 - a. Make a conjecture about how the IQR and the mean \pm MAD will change. Then move the three dots to the suggested heights to check your reasoning. (Select the white space to deselect a dot.)

Answers will vary. Some may suggest that both get smaller; others that only the interval for the mean \pm MAD will get smaller. After the dots are moved, both the mean \pm MAD and the IQR get smaller because the heights are now all closer together.

- b. Look at the distribution of heights after the heights were adjusted. Which of the following words would you use to describe the new distribution of heights: skewed right, skewed left, symmetric, mound shaped? Explain why.

Answer: The new distribution of heights of the dogs is mound shaped and relatively symmetric because the heights are clustered between about 58 and 63 centimeters with about three dogs with heights from 56 to 59 centimeters and three dogs with heights from 62 to 65 centimeters, about three centimeters on either side of the cluster.

- c. How do the mean and median of the adjusted distribution of the dog heights compare?

Answer: They are close together; the mean is 60.6 centimeters, and the median is 61 centimeters.

- d. Reset. Which of the following words would you use to describe the original distribution of the dog heights: skewed right, skewed left, symmetric, mound shaped? Explain your reasoning.

Answer: The distribution of heights of the dogs is skewed right because the heights of the dogs are between about 58 and 64 centimeters except for three dogs, who have heights that go all the way up to 75 centimeters, making it skewed in the direction of the tail.

- e. How do the mean and median of the distribution of dog heights compare?

Answer: The mean height for the dogs is 62.6 centimeters, 1.6 centimeters larger than the median of 61 centimeters.



2. Which of the following do you think are true statements about a distribution of data? Use your answers for the previous question to support your thinking in each case.
 - a. If a distribution is skewed, the mean and the median will be close together.
 - b. If a distribution is skewed, the median probably represents a better measure of the center than the mean.
 - c. In a symmetric mound shaped distribution, the mean and the median will be close together.
 - d. If a distribution is skewed, the measures that best summarize the data are the median and the IQR.
 - e. If a distribution is mound shaped, the measures that best summarize the data are the mean and mean \pm MAD.

Answer: Statements b, c, and d are true, given the difference between the skewed distribution of heights and the mound shaped and symmetric distribution of height in question 1. Statement a. is not true because a skewed distribution will have either some values larger than all of the others or smaller than all of the others. These values will make the means larger or smaller because their absolute deviations will be large. Statement e is not true because the distribution could be mound shaped at one end, which is kind of the case in question 1. When the distribution is both mound shaped and symmetric, the mound will be in the center of the distribution making the mean and median close together.



Activity 2 [Page 2.2]

1. Reset. Choose **all animals, life span, bar graph**.
 - a. What category has the life span of the most types of animals? How do you know? (Note that dogs are not included in the domestic animal category and that hovering over a bar shows the frequency of the types of animals represented in the bar.)

Answer: There were 35 types of birds, more types than any other kind of animals. You can tell because the bar is highest, which means the frequency is the largest.
 - b. Which of the statements is true? Explain why.
 - i. The number of types of wild animals is more than twice as many as the number of types of domestic animals (excluding dogs).

Answer: There are 31 types of wild animals and 14 types of domestic animals so there are more than twice as many types of wild animals.
 - ii. The total number of types of fish and sea mammals is more than the total number of wild and domestic land animals (excluding dogs and cats).

Answer: There are 41 types of fish and sea mammals and 45 types of wild and domestic land animals, so the statement is not true.



- iii. The difference between the number of types of birds and types of cats is more than the difference between the number of types of dogs and number of types of wild land animals.

Answer: The difference between the number of types of birds and cats is 18 and the difference between the number of types of dogs and number of types of wild animals is 5, so the statement is true.

2. What is the difference between a bar graph and a histogram? Use examples from the TNS activity to support your reasoning.

Answer (examples may vary). A bar graph shows the number of types for a category or kind of animal, while a histogram does not show the type but does show how many types have a specific maximum speed or life span located on the number line. For example, if the attribute is life span for all animals and the graph is a bar graph, there are 17 types of cats; while, from the all animals histogram with bin width 5, there are 21 types of animals with maximum life spans between 20 and 25 years. The bar graph allows you to compare the number of each type (there are more dog types than fish types in the data), while the histogram allows you to say something overall about the life spans of the animals – the distribution is skewed right with the most common life span (46 animal types) from 10 to 15 years and four animal types had a life span from 75 to 80 years.



Activity 3 [Page 3.2]

1. Choose **menu> Type> Birds** and **menu> Attribute> Life Span**.

- a. How wide are the bins in the histogram? Explain how you know.

Answer: The bins are 5 units—they go from 5 years up to and including 9 years, 10 years up to and including 14 years and so on. You can tell by looking at the scale and checking with the dot plot or by hovering over the bar. You know that 5 cannot be in both bins.

- b. What can you learn from each of kind of plot that you cannot learn from the others? Use examples from part a to support your thinking.

Answers will vary. Students should note that the dot plot shows all of the life spans, which none of the other plots do (note the Amazon parrot has the longest recorded life span at about 104 years). The histogram gives a frequency for the number of birds with life span within a fixed interval (five birds that have a life span from 5 to 9 years) and allows you to compare the frequencies in these intervals by the height of the bins. The box plot allows you to get a good estimate of the LQ (about 15 years), UQ (about 33 years) and median (about 22 years); you could also do this on the dot plot, but it would involve a lot of counting.

- c. What are the disadvantages of each plot?

Answer: If there are a lot of animals, the dots get all piled on top of each other; the histogram groups values into the same bin so you lose the exact value, and sometimes you lose any gaps or clusters or the clusters show up in strange places because of the way the values go into the bins; the box plot loses all of the exact values and does not show gaps or clusters.



2. Reset. Choose **menu> Type> all**, **menu> Attribute> Max Speed**. Set the bin width of the histogram to 10. Which plot seems to be the best for displaying the data? Explain your reasoning.

Answer: The histogram or the boxplot seem to be the best; the histogram shows the shape of the distribution, and you can see the number of data values in each bin. It shows that the speed of the outlier, the peregrine falcon, is really far away from the other speeds. The boxplot shows the peregrine falcon as the maximum and how far away the other speeds are. The boxplot does show the IQR and median; you would have to estimate these from the histogram. The dot plot has a lot of dots, and they get all scrambled on top of each other.