Lesson 20: Describing Center, Variability, and Shape of a Data Distribution from a Graphic Representation

Student Outcomes

- Given a frequency histogram, students are able to describe the data collected, including the number of responses, an estimate of the mean or median, and an estimate of the interquartile range (IQR) or the mean absolute deviation (MAD).

Lesson Notes

In each lesson of this module, students were either given graphic representations of a data distribution or they were expected to construct a graphic representation. The graphic representations used in this module are dot plots, box plots, and histograms. Each of these representations provides a summary of the data. If actual data are provided and the data set is not too large, a dot plot is generally a good way to display the data distribution. If the data set is large, a histogram is generally used to display the data distribution. Histograms are challenging for students at this grade level. A histogram provides a display of the data distribution, but the shape of a histogram can sometimes depend on the intervals used to construct the histogram. It is important to investigate the shape of the data distribution because the shape influences the choice of numerical summaries—the mean and MAD are used for data distributions that are approximately symmetric, and the median and IQR are used for data distributions that are skewed.

In this lesson, students consider a histogram of the length of yellow perch in the Great Lakes region. The data presented in this lesson are based on various scientific research studies of yellow perch during the 1990s. Before students analyze the data, share with them that the histogram is part of the yellow perch story that they will uncover in this lesson. The yellow perch is a valuable resource for the fishing industry, as well as a food source for several other species of fish and wildlife. Although the sample presented in this lesson has been simplified for students, it provides a graphic representation of data that was particularly disturbing to scientists researching the yellow perch. The histogram is a graphic representation of the length of the yellow perch. The length of the yellow perch is used to estimate the age of the fish. As a result, the histogram indicates that most of the fish were older or adult fish. It would be preferable for the younger fish to represent a larger proportion of the population of yellow perch. If the intervals representing the younger fish are less than the older intervals, then in time there will be fewer and fewer fish, possibly even indicating that the yellow perch will not survive.

Classwork

> Great Lakes yellow perch are fish that live in each of the five Great Lakes and many other lakes in the eastern and upper Great Lakes regions of the United States and Canada. Both countries are actively involved in efforts to maintain a healthy population of perch in these lakes.

Example 1 (10 minutes): The Great Lakes Yellow Perch

Example 1: The Great Lakes Yellow Perch

Scientists collected data from many samples of yellow perch because they were concerned about the survival of the yellow perch. What data do you think researchers might want to collect about the perch?

Scientists captured yellow perch from a lake in this region. They recorded data on each fish, and then returned each fish to the lake. Consider the following histogram of data on the length (in centimeters) for a sample of yellow perch.
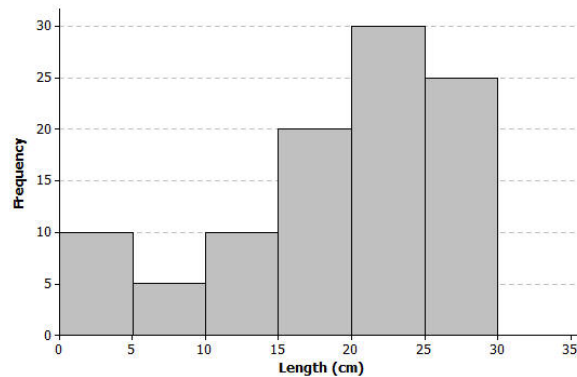
Discuss the following questions with students as you present the histogram.

MP.1

- What data do you think researchers might want to collect about the perch?
- How many fish captured had a length of 20 to 25 centimeters?
- Do you know how many fish had a length of 22 centimeters? Why or why not? (Remind students to understand that a histogram does not provide the frequency of a specific value, only the frequency for an interval of values.)
- Why were the scientists concerned about what they saw in the histogram of the lengths of yellow perch?

Exercises 1–11 (15 minutes)

Students should work individually or in small groups as they answer the questions in these exercises. Discuss as a group answers to these questions.

Exercises 1–11

Scientists were concerned about the survival of the yellow perch as they studied the histogram.

1. **What statistical question could be answered based on this data distribution? How do you think the scientists collected these data?**

   Answers will vary: A possible statistical question would be, "What is a typical length of the Great Lakes yellow perch?" Remind students that a statistical question is a question that can be answered by data that you anticipate will vary.

2. Use the histogram to complete the following table:

Length of fish in centimeters (cm)	Number of fish
$0 - < 5$ cm	10
$5 - < 10$ cm	5
$10 - < 15$ cm	10
$15 - < 20$ cm	20
$20 - < 25$ cm	30
$25 - < 30$ cm	25

3. The length of each fish was measured and recorded before the fish was released back into the lake. How many yellow perch were measured in this sample?

   100 fish were measured in this sample.

4. Would you describe the distribution of the lengths of the fish in the sample as a skewed data distribution or as a symmetrical data distribution? Explain your answer.

   The data distribution is a skewed distribution, with the tail to the left.

5. What percentage of fish in the sample were less than 10 centimeters in length?

   15 fish had a length of less than 10 centimeters, thus 15% are less than 10 centimeters.

6. If the smallest fish in this sample were 2 centimeters in length, what is your estimate of an interval of lengths that would contain the lengths of the shortest 25% of the fish? Explain how you determined your answer.

   25% of the fish are represented in the first three intervals. If the smallest value in the first interval is known, then an estimate of the interval of the smallest 25% of the fish is 2 centimeters to 15 centimeters. Students would determine this by considering the histogram bars at the low end and looking for an interval that would represent 25 fish.

7. If the length of the largest yellow perch was 29 centimeters, what is your estimate of an interval of lengths that would contain the lengths of the longest 25% of the fish?

   In a similar way, there are 25 fish in the interval 25 to 30 centimeters in length. If the longest fish were measured at 29 centimeters, then an estimate of the upper 25% would be 25 to 29 centimeters.

8. Estimate the median length of the yellow perch in the sample. Explain how you determined your estimate.

   To estimate the median length, students would identify a length in which approximately 50% of the fish would be above and approximately 50% of the fish would be below the estimate. If a student would start from the smallest lengths, an estimate of the median would be located within the 20 to 25 centimeters interval. The same interval would be identified if students started with the largest lengths. As the actual values of the lengths of the fish are not known, any estimate within that interval would be a good estimate. For example, an estimate of 23 centimeters would be a good estimate.

9. Based on the shape of this data distribution, do you think the mean length of a yellow perch would be greater than, less than, or the same as your estimate of the median? Explain your answer.

   Because the data distribution is skewed, the smaller lengths pull an estimate of the mean to the left of the median. Therefore, an estimate of the mean would be less than the estimate of the median.

10. Recall that the mean length is the balance point of the distribution of lengths. Estimate the mean length for this sample of yellow perch.

If students think of the mean as a balance point, they will estimate a length that is less than the median. Answers will vary, but an estimate of a length in the 15 to 20 centimeters interval would show an understanding of this idea. For example, 17 or 18 centimeters would be a good estimate of the mean.

11. The length of a yellow perch is used to estimate the age of the fish. Yellow perch typically grow throughout their lives. Adult yellow perch have lengths between 10 and 30 centimeters. How many of the yellow perch in this sample would be considered adult yellow perch? What percentage of the fish in the sample are adult fish?

85 fish are counted in the intervals that represent 10 to 30 centimeters. Therefore, 85% of the fish in this sample were estimated to be adult fish.

Example 2 (5 minutes): What Would a Better Distribution Look Like?

Example 2: What Would a Better Distribution Look Like?

Yellow perch are part of the food supply of larger fish and other wild life in the Great Lakes region. Why do you think that the scientists worried when they saw the histogram of fish lengths given above?

Sketch a histogram representing a sample of 100 yellow perch lengths that you think would indicate the perch are not in danger of dying out?
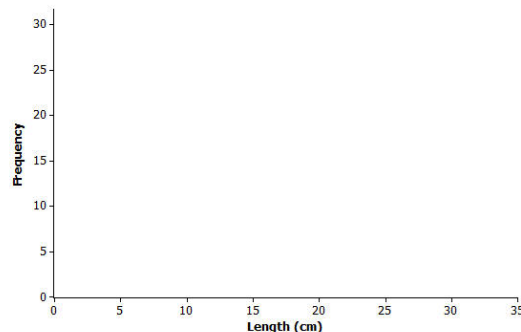
Discuss the summary that was provided at the beginning of the teacher notes. A better distribution of fish would have more young fish. Because age is related to length, a better distribution would have more fish in the smaller length intervals. As the lengths (or ages) increased, you would expect the number of fish to decline. Allow students to sketch their own histogram shapes in response to the discussion question in this exercise. Point out that a histogram with the greater frequencies of fish in the smaller length intervals would be better.

MP.2

Exercises 12–17 (10 minutes): Estimating the Variability in Yellow Perch Lengths

Exercises 12–17: Estimating the Variability in Yellow Perch Lengths

You estimated the median length of yellow perch from the first sample in Exercise 8. It is also useful to describe variability in the length of yellow perch. Why might this be important? Consider the following questions:

12. In several previous lessons, you described a data distribution using the 5-number summary. Use the histogram and your answers to the questions in Exercise 2 to provide estimates of the values for the 5-number summary for this sample:

    Min or minimum value = 2 centimeters

    Q1 value = 15 centimeters

   *Median = 23 centimeters*

    Q3 value = 25 centimeters

    Max or maximum value = 29 centimeters

13. Based on the 5-number summary, what is an estimate of the value of the interquartile range (IQR) for this data distribution?

    Based on the above estimates, an estimate of the interquartile range (IQR) would be as follows: 25 centimeters – 15 centimeters or 10 centimeters.

14. Sketch a box plot representing the lengths of the yellow perch in this sample.
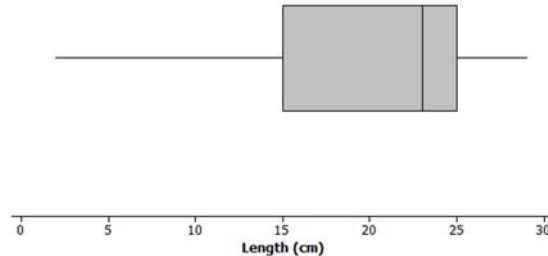
15. Which measure of center, the median or the mean, is closer to where the lengths of yellow perch tend to cluster?

    For a skewed distribution, the median is closer to where the lengths of yellow perch tend to cluster.

16. What value would you report as a typical length for the yellow perch in this sample?

    Encourage students to use the median value they estimated in the previous questions as the typical value of the yellow perch.

17. The mean absolute deviation (or MAD) or the interquartile range (IQR) are used to describe the variability of a data distribution. Which measure of variability would you use for this sample of perch? Explain your answer.

    When the median is selected as the measure of center for a typical value, then the interquartile range would be selected as the measure of variability. In this case, 10 centimeters, or the IQR determined in Exercise 2, would be the measure of the variability.

Closing (5 minutes)

Describe additional questions.

MP.4

- What is the problem with the yellow perch length distribution shown in the opening histogram?
- What is a typical yellow perch length?
  - *(Basically, the answer to this question is the answer to students' statistical question.)*
- What would you use as a measure of the variability of yellow perch lengths?
  - *You would use the measurement of centimeters.*

Lesson Summary

Data distributions are usually described in terms of shape, center, and spread. Graphical displays, such as histograms, dot plots, and box plots, are used to assess the shape. Depending on the shape of a data distribution, different measures of center and variability are used to describe the distribution. For a distribution that is skewed, the median is used to describe a typical value, whereas the mean is used for distributions that are approximately symmetric. The IQR is used to describe variability for a skewed data distribution, while the MAD is used to describe variability for distributions that are approximately symmetric.

Exit Ticket (5 minutes)

Name _____   Date_____

Lesson 20: Describing Center, Variability, and Shape of a Data Distribution from a Graphic Representation

Exit Ticket

1. Great Lake yellow perch continue to grow until they die.   What does the histogram in Example 1 indicate about the ages of the perch in the sample?

2. What feature of the histogram in Example 1 indicates that the values of the mean and the median of the data distribution will not be equal?

3. Adult yellow perch have lengths between 10 and 30 centimeters. Would a perch with a length equal to the median length be classified as an adult or pre-adult fish? Explain your answer.

Exit Ticket Sample Solutions

1. Great Lake yellow perch continue to grow until they die. What does the histogram in Example 1 indicate about the ages of the perch in the sample?

    The histogram indicates that most of the perch are in the intervals corresponding to the longest lengths. Because length is related to age, the histogram indicates that there are more fish estimated as older fish.

2. What feature of the histogram in Example 1 indicates that the values of the mean and the median of the data distribution will not be equal?

    The histogram indicates that the shape of the data distribution is skewed. For skewed distributions, the mean and the median are not equal.

3. Adult yellow perch have lengths between 10 and 30 centimeters. Would a perch with a length equal to the median length be classified as an adult or pre-adult fish? Explain your answer.

    A perch equal to the median length would be classified as an adult fish. The median is estimated to be between 20 and 25 centimeters in length. Adult fish are 10 centimeters or more in length.

Problem Set Sample Solutions

Consider again your students' progress on their projects. If needed, develop expectations for completing the four-steps of the project as part of the problem set. Indicate to students that you expect a brief written summary of their progress or a sample of the graphs or numerical summaries from their data. Assign four or five of the problems in the problem set if completion of the project needs to be addressed.

Another sample of Great Lake yellow perch from a different lake was collected. A histogram of the lengths for the fish in this sample is shown below:
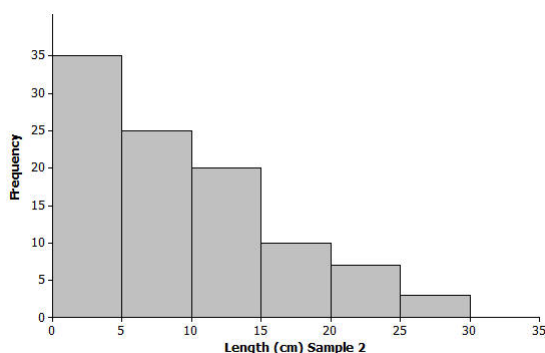
1. If the length of a yellow perch is an indicator of its age, how does this second sample differ from the sample you investigated in the exercises? Explain your answer.

    The second sample has more fish with lengths in the intervals corresponding to shorter lengths. Therefore, more of the fish are younger.

2. Does this histogram represent a data distribution that is skewed or that is nearly symmetrical?

    This distribution is also skewed. However, the tail of this distribution is to the right, or toward the longer lengths.

Lesson 20: | Describing Center, Variability, and Shape of a Data Distribution from a Graphic Representation
Date: | 4/3/14

engage^ny

226

3.   What measure of center would you use to describe a typical length of a yellow perch in this second sample? Explain your answer.

Students should recommend the median of the data distribution as a description of a typical value of the length of the yellow perch because this distribution is also skewed.

4.   Assume the smallest perch caught was 2 centimeters in length, and the largest perch caught was 29 centimeters in length. Estimate the values in the 5-number summary for this sample:

Q1, Q3, and median values are not as clear-cut in this distribution as in Exercise 4, so allow a wider range of acceptable answers.

Min or minimum value = 2 centimeters

Q1 value = 4 centimeters (value greater than 2, but within the interval of 0 to 5 centimeters)

Median value = 7 centimeters (a value within the interval of 5 to 10 centimeters)

Q3 value = 12 centimeters (a value within the interval of 10 to 15 centimeters)

Max or maximum value = 29 centimeters

5.   Based on the shape of this data distribution, do you think the mean length of a yellow perch from this second sample would be greater than, less than, or the same as your estimate of the median? Explain your answer.

An estimate of the mean would be greater than the median length because the values in the tail or to the right of the median pull the mean in that direction. Consider estimating the mean as the balance point of this distribution. (If students have problems with estimating the balance point, consider providing them a representation similar to the representation used to introduce a balance point in earlier lessons. Use a ruler with coins (or weights) taped to locations that would represent a skewed distribution. This representation helps them sense the point of balance.)

6.   Estimate the mean value of this data distribution.

An estimate of the mean would be a value slightly larger than the median value. For example, a mean of 10 or 11 centimeters would be a reasonable estimate of a balance point.

7.   What is your estimate of a typical length of a yellow perch in this sample? Did you use the mean length from problem 5 for this estimate? Explain why or why not.

As the median was selected as the estimate of a measure of center, a value of 7 centimeters (or whatever students used to estimate the median) would be an estimate of a typical value for a yellow perch from this sample.

8.   Would you use the MAD or the IQR to describe variability in the length of Great Lakes yellow perch in this sample? Estimate the value of the measure of variability that you selected.

Students should use the IQR to describe the variability because the data distribution is skewed and the median was used as a measure of a typical value. An estimate of the IQR based on the above estimates would be as follows: 12 − 4 centimeters = 8 centimeters.