

Lesson 20: Describing Center, Variability, and Shape of a Data

Distribution from a Graphic Representation

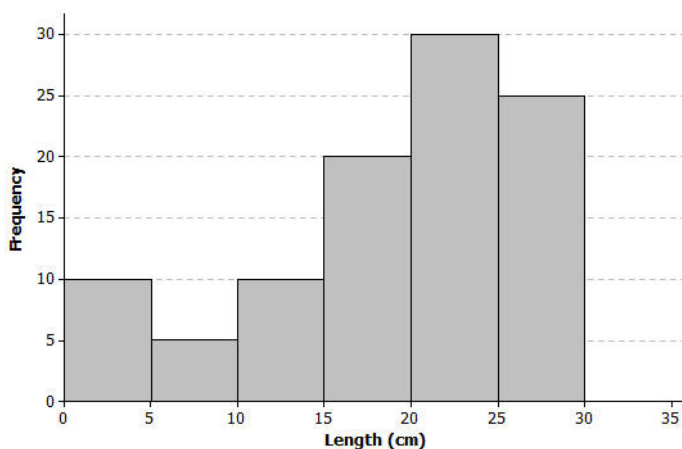
Great Lakes Yellow Perch are fish that live in each of the five Great Lakes and many other lakes in the eastern and upper Great Lakes regions of the United States and Canada. Both countries are actively involved in efforts to maintain a healthy population of perch in these lakes.

Classwork

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.



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?

2. Use the histogram to complete the following table:

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

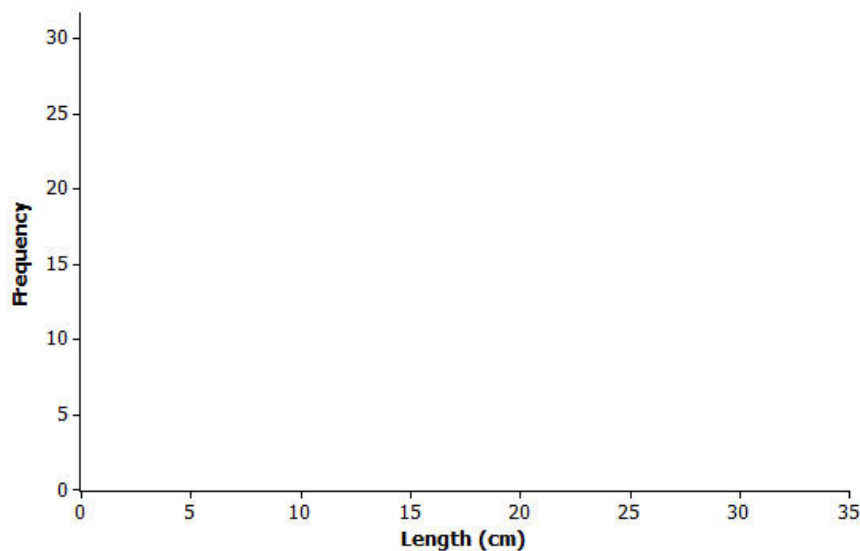
3. The length of each fish was measured and recorded before the fish was released back into the lake. How many yellow perch were measured in this sample?
4. Would you describe the distribution of the lengths of the fish in the sample as a skewed data distribution or as a symmetrical data distribution? Explain your answer.
5. What percentage of fish in the sample were less than 10 centimeters in length?
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.

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

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?

**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 =

Q1 value =

Median =

Q3 value =

Max or maximum value =

13. Based on the 5-number summary, what is an estimate of the value of the interquartile range (IQR) for this data distribution?
14. Sketch a box plot representing the lengths of the yellow perch in this sample.



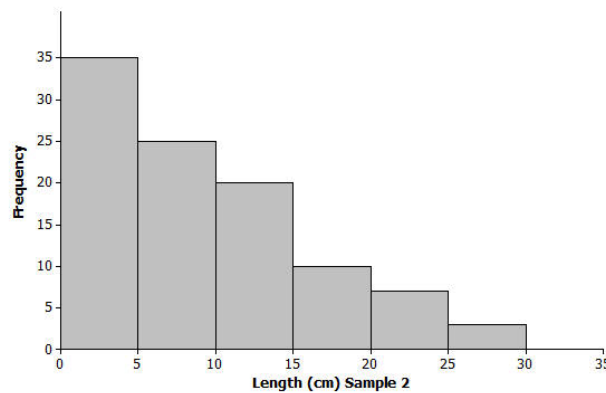
15. Which measure of center, the median or the mean, is closer to where the lengths of yellow perch tend to cluster?
16. What value would you report as a typical length for the yellow perch in this sample?
17. The mean absolute deviation (or MAD) or the interquartile range (IQR) are used to describe the variability of a data distribution. Which measure of variability would you use for this sample of perch? Explain your answer.

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.

Problem Set

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.
2. Does this histogram represent a data distribution that is skewed or that is nearly symmetrical?
3. What measure of center would you use to describe a typical length of a yellow perch in this second sample? Explain your answer.
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:
Min or minimum value =
Q1 value =
Median =
Q3 value =
Max or maximum value =

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.
6. Estimate the mean value of this data distribution.
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.
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.