The **normal distribution**, also known as the **Gaussian distribution** or the **bell curve**, is one of the most important tools in statistics and probability theory. The normal distribution arises from running repeated trials of a random event, such as tossing a coin or rolling dice. For instance, if a fair coin is tossed N times, we expected to land on heads roughly N/2 times, or half the time. However, we do not expect it to land on heads exactly N/2 out of N times, especially when N is large, but rather somewhere near N/2 times. It is possible to calculate the probability of the coin landing on heads exactly n times for any value of n from 0 to N. The following four graphs show these probability distributions for N = 10, 20, 50, and 100.
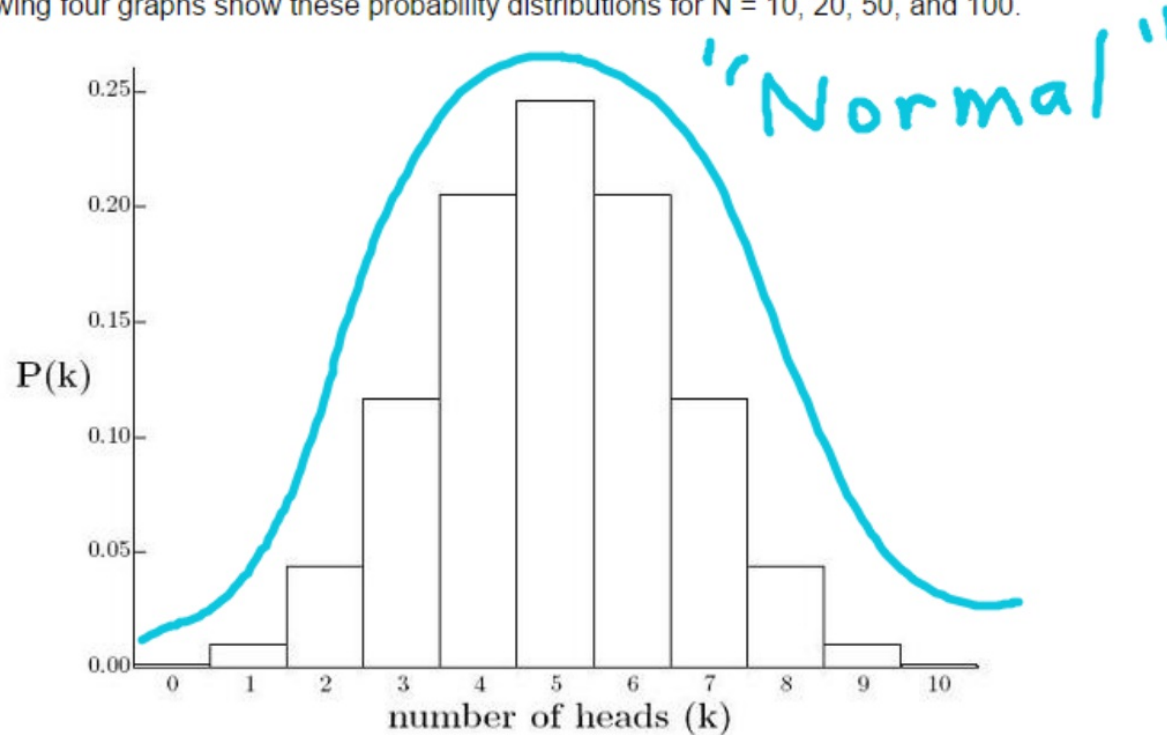

Figure 5.5.5.1: Probability Distribution of Heads for 10 Coin Tosses

Check ✓

1. The histogram shows the wait times in minutes for entering a concert. Describe the shape of the distribution. (Example 1)

Sample answer: The shape of the distribution is not symmetric. There is a cluster from 0–49 minutes. The distribution has a gap from 50–89. The peak of the data is on the left side of the data in the interval 20–29. There is an outlier in the interval 90–99.

Concert Entrance Wait Times (min)

Histogram: Frequency (y-axis, 0 to 20) vs Wait Times (min) (x-axis, intervals 0–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, 90–99)

2. The box plot shows the weights in pounds of several dogs. Describe the shape of the distribution. (Example 2)

Sample answer: The shape of the distribution is not symmetric since the lengths of each box and whisker are not the same. There is an outlier at 20.

Weights of Dogs (lb)

Box plot with number line: 20 40 60 80 100 120 140 160 180 200
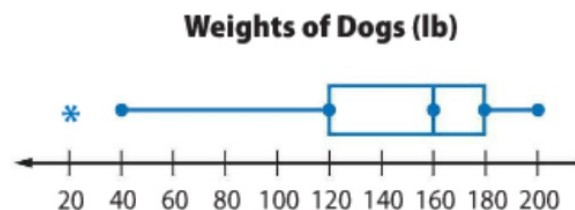
3. The line plot shows the number of hours several students spent on the Internet during the week. (Example 3)

Number of Hours Spent on the Internet

a. Choose the appropriate measures to describe the center and spread of the distribution. Justify your response based on the shape of the distribution. _**mean and mean absolute**_

deviation; Sample answer: The distribution is symmetric

and there are no outliers.

b. Write a few sentences describing the center and spread of the distribution using the appropriate measures. Round to the nearest tenth if necessary.

Sample answer: The data are centered around the

mean of 4 hours. The spread of the data around the

center is about 1.2 hours, which is the mean absolute

deviation.

4. 🄮 **Building on the Essential Question** Why does the choice of measure of center and spread vary based on the type of data display? **Sample answer: Different displays**

show different information. Median can be found in line

plots, histograms, and box plots, but mean can only be

found when looking at individual data. Not all displays

show gaps, clusters, and peaks.

Rate Yourself!

How well do you understand how to describe the shape of a distribution? Circle the image that applies.

Clear   Somewhat Clear   Not So Clear

**For more help, go online to access a Personal Tutor.**

Tutor

1 The histogram shows the average animal speeds in miles per hour of several animals. Describe the shape of the distribution. (Example 1)

Sample answer: The shape of the distribution is not symmetric. There is a cluster from 1–79. The distribution has a gap from 80–199. The peak of the distribution is on the left side of the data in the interval 20–39. There is an outlier in the interval 200–219.

Average Animal Speeds

Number of Animals vs. Speed (miles per hour)
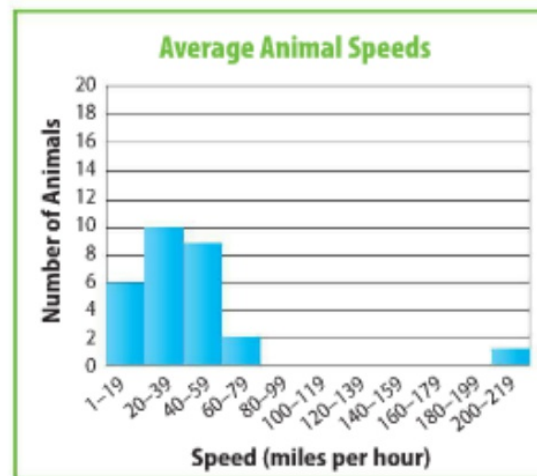
2. The box plot shows the science test scores for Mrs. Everly's students. Describe the shape of the distribution. (Example 2)

Sample answer: The shape of the distribution is symmetric. The left side of the data looks like the right side. There are no outliers.

Science Test Scores (%)

60 65 70 75 80 85 90 95 100
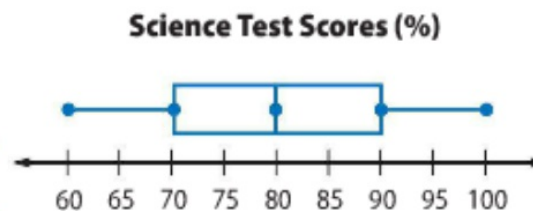
3 The line plot shows the number of text messages sent by different students in one day. (Example 3)

Number of Text Messages Sent Today

17 18 19 20 21 22 23 24 25

a. Choose the appropriate measures to describe the center and spread of the distribution. Justify your response based on the shape of the distribution.

median and interquartile range; Sample answer:

The distribution is not symmetric.

b. Write a few sentences describing the center and spread of the distribution using the appropriate measures.

**Sample answer: The data are centered around 23.5 text messages.**

The spread of the data around the center is about 3 text messages.

4. **CCSS** **Identify Structure** Fill in the graphic organizer to show when to use each measure regarding the shape of the distribution.

Measure	Symmetric or Not Symmetric
mean	**symmetric**
median	**not symmetric**
interquartile range	**not symmetric**
mean absolute deviation	**symmetric**

5. A distribution that is not symmetric is called *skewed*. A distribution that is *skewed left* shows data that is more spread out on the left side than on the right side. A distribution that is *skewed right* shows data that is more spread out on the right side than on the left side. The box plot shows the heights in feet of several trees.

Height of Trees (ft)

10 15 20 25 30 35 40 45 50
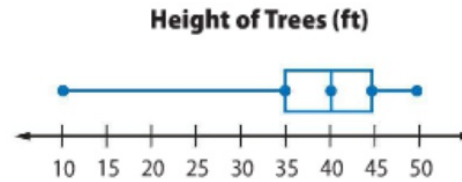
a. Explain how you know the distribution is not symmetric.

Sample answer: The lengths of the whiskers are not equal.

b. Is the distribution skewed left or skewed right? Explain.

skewed left; Sample answer: The data are more spread out on the left side

due to the long left whisker.

c. Use appropriate measures to describe the center and spread of the distribution. Justify your choice of measure based on the shape of the distribution. Sample answer: Use the median and interquartile range to describe the

center and spread since the distribution is not symmetric. The data are centered around

40 feet. The spread of the data around the center is 10 feet.

H.O.T. Problems Higher Order Thinking

6. **CCSS** **Model with Mathematics** Draw a line plot for which the median is the most appropriate measure to describe the center of the distribution.

See students' work; Distributions should not be symmetric.

7. **CCSS** **Persevere with Problems** Explain why you cannot describe the specific location of the center and spread of the box plot shown using the most appropriate measures.

Calories in Servings of Fruits
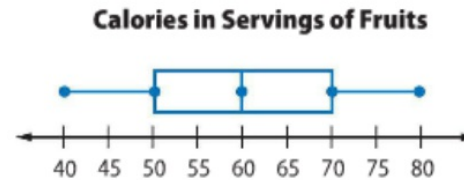
40 45 50 55 60 65 70 75 80

Sample answer: The distribution is symmetric. The appropriate measures to describe the center and spread are the mean and mean absolute deviation. A box plot shows the location of the median and interquartile range but it does not show the location of the mean or the mean absolute deviation.

8. **CCSS** **Justify Conclusions** Tyra created the dot plot shown to represent the ages of the staff of the community pool. She concludes that since there is a peak at 19, the median is 19. She also concludes the two data values that are 25 to be outliers, so there are no gaps. Evaluate her conclusions.

14 15 16 17 18 19 20 21 22 23 24 25

Ages of Staff Members

Sample answer: The median of the data set is 19.5. Peaks in a data set do not necessarily correspond to the median. The data values of 25 are outliers, but there is a gap between 21 and 25.