Analyze Data Distributions

Lesson 9.6
Complete the graphic organizer by matching the term with the correct description.

...when many data values are grouped together.

...the most frequently occurring value or interval of value.

A peak is...

An outlier is...

A gap is...

A cluster is...

...where there are no data values.

...a data value that is 1.5 times the interquartile range from the first or third quartiles.
**VOCABULARY:**

**Distribution** -
A way to show the arrangement of data values. It can be described by its center, spread (variation), and overall shape.

**Symmetric** -
A description of the shape of a distribution in which the left side of the distribution looks like the right side.
To describe the shape of a distribution, you must identify and describe each of the following:

1. Symmetry
2. Peaks
3. Clusters
4. Gaps
5. Outliers
The graph shows the weights of adult cats. Identify any symmetry, clusters, gaps, peaks, or outliers in the distribution.

Symmetry: non-symmetric

Peaks: at 10 lb

Clusters: from 7-12 lbs

Gaps: between 12 and 14 lbs

Outliers: none

The distribution is non-symmetric. There is a cluster from 7-12 with a peak at 10. There is a gap between 12 and 14. There are no outliers.
Describe the shape of the distribution. Identify any clusters, gaps, peaks, or outliers. (Example 1)

Symmetry: not symmetric

Clusters: from 23-27 and 29-31

Gaps: between 20-23 and 27-29

Peaks: at 27

Outliers: none

The distribution is non-symmetric. There is a cluster from 23-27 and from 29-31 with a peak at 27. There is a gap between 20-23 and between 27-29. There are no outliers.
The winning scores for twenty Super Bowls are shown in the histogram below. Describe the shape of the distribution. Identify any clusters, gaps, peaks, or outliers.

Symmetry: symmetric

Clusters: none

Gaps: none

Peaks: at 31-40

Outliers: none

The distribution is symmetric. The peak of the data is at the interval 31-40. There are no clusters, gaps, or outliers.
Symmetry: not symmetric

Clusters: from 60-63

Gaps: between 58-60 and 63-71

Peaks: at 61

Outliers: at 71

The distribution is non-symmetric. There is a cluster from 60-63 with a peak at 61. There is a gap between 58-60 and between 63-71. There is an outlier at 71.
Describe the Center and Spread of a Distribution

The shape of a distribution tells you which measures are most appropriate for describing the center and spread of a distribution. The mean and mean absolute deviation are affected by outliers, while the median and interquartile range are resistant to outliers.

Is the data distribution symmetric?

- Use the **median** to describe the center and the **interquartile range** to describe the spread.

- Use the **mean** to describe the center and the **mean absolute deviation** to describe the spread.
2. Mr. Watkin’s class charted the high temperatures in various cities. The results are shown in the line plot.

Describe the center and spread of the distribution. Justify your response based on the shape of the distribution.

a) Is it symmetric?  No

b) What do you need to use to describe the distribution and spread?  
Median and Interquartile Range

The distribution is not symmetric. So, the median and interquartile range are the appropriate measures to use. The data are centered around the median of 84°. The first quartile is 80 and the third quartile is 95.5. So, the interquartile range is 95.5 – 80 or 15.5°. The spread of the data around the center is 15.5°.

The line plot shows:

- Median = 84
- Interquartile Range = Q3 - Q1 = 95.5 - 80 = 15.5
The graph shows the hours per week that dance students practice their dances. Describe the center and spread of the distribution. Justify your response based on the shape of the distribution. Round to the nearest tenth if necessary.

a) Is it symmetric? Yes

b) What do you need to use to describe the distribution and spread? Mean and Mean Absolute Deviation
**Example**

**Mean:**
Use the peak as the average = 5

**Mean Absolute Deviation:**

\[
egin{align*}
|2-5| &= 3 & |5-5| &= 0 & |6-5| &= 1 \\
|3-5| &= 2 & |5-5| &= 0 & |6-5| &= 1 \\
|3-5| &= 2 & |5-5| &= 0 & |6-5| &= 1 \\
|4-5| &= 1 & |5-5| &= 0 & |7-5| &= 2 \\
|4-5| &= 1 & |5-5| &= 0 & |7-5| &= 2 \\
|8-5| &= 3
\end{align*}
\]

\[
3 + 2 + 2 + 1 + 1 + 0 + 0 + 0 + 0 + 0 + 1 + 1 + 1 + 2 + 2 + 3 = 19 \div 16 = 1.1875
\]
The distribution is symmetric, so the mean and mean absolute deviation are appropriate measures to use. The data is centered around the mean of 5 hours. The spread of the data around the center is about 1.2 hours.
a) Is it symmetric? Yes

b) What do you need to use to describe the distribution and spread?

Mean and Mean Absolute Deviation

The distribution is symmetric, so the mean and mean absolute deviation are appropriate measures to use. The data are surrounded by the mean of 2.4. The spread of the data around the center is 0.1375.
Describe the center and spread of the distribution. Justify your response based on the shape of the distribution. (Example 2)

The distribution is not symmetric, so the median and interquartile range are appropriate measures to use. The data are centered around the median of 25. The spread of the data around the center is 3.
Homework:

Pg. 721-723
# 1-9 (all)
and
Pg. 728-729
all problems