WIIT 7751

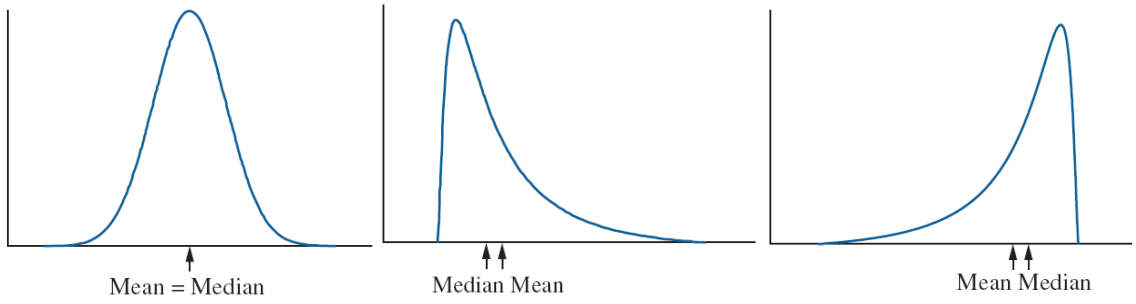
Lesson E Guided Notes

Describing the Shape of a Data Set

The mean and median measure the center of a data set in different ways. When a data set is **symmetric**, the mean and median are equal.

When a data set is skewed to the right, there are large values in the right tail. Because the median is resistant while the mean is not, the mean is generally more affected by these large values. Therefore for a data set that is **skewed to the right**, the mean is often greater than the median.

Similarly, when a data set is **skewed to the left**, the mean is often less than the median.



Outliers

An **outlier** is a value that is considerably larger or considerably smaller than most of the values in a data set.

Some outliers result from errors; for example a misplaced decimal point may cause a number to be much larger or smaller than the other values in a data set.

Some outliers are correct values, and simply reflect the fact that the population contains some extreme values.

Example 1 The temperature in a downtown location in a certain city is measured for eight consecutive days during the summer. The readings, in degrees Fahrenheit, are

81.2, 85.6, 89.3, 91.0, 83.2, 8.45, 79.5, and 87.8.

Which reading is an outlier? Is it certain that the outlier is an error, or is it possible that it is correct? Should the outlier be deleted?

Interquartile Range

One method for detecting outliers involves a measure called the **Interquartile Range**.

The interquartile range is found by subtracting the first quartile from the third quartile.

IQR = $Q_3 - Q_1$

IQR Method for Detecting Outliers

The most frequent method used to detect outliers in a data set is the **IQR Method**. The procedure for the IQR Method is:

Step 1: Find the first quartile Q_1, and the third quartile Q_3.

Step 2: Compute the interquartile range: IQR = $Q_3 - Q_1$.

Step 3: Compute the outlier boundaries. These boundaries are the cutoff points for determining outliers:

$$\text{Lower Outlier Boundary} = Q_1 - 1.5(\text{IQR})$$

$$\text{Upper Outlier Boundary} = Q_3 + 1.5(\text{IQR})$$

Step 4: Any data value that is less than the lower outlier boundary or greater than the upper outlier boundary is considered to be an outlier.

Example 2 The following table presents the number of students absent in a middle school in northwestern Montana for each school day January 2008. Identify any outliers.

Jan. 2	65	Jan. 10	44	Jan. 18	45	Jan. 28	100
Jan. 3	67	Jan. 11	41	Jan. 21	77	Jan. 29	59
Jan. 4	71	Jan. 14	59	Jan. 22	44	Jan. 30	53
Jan. 7	57	Jan. 15	49	Jan. 23	42	Jan. 31	51
Jan. 8	51	Jan. 16	42	Jan. 24	45		
Jan. 9	49	Jan. 17	56	Jan. 25	46		