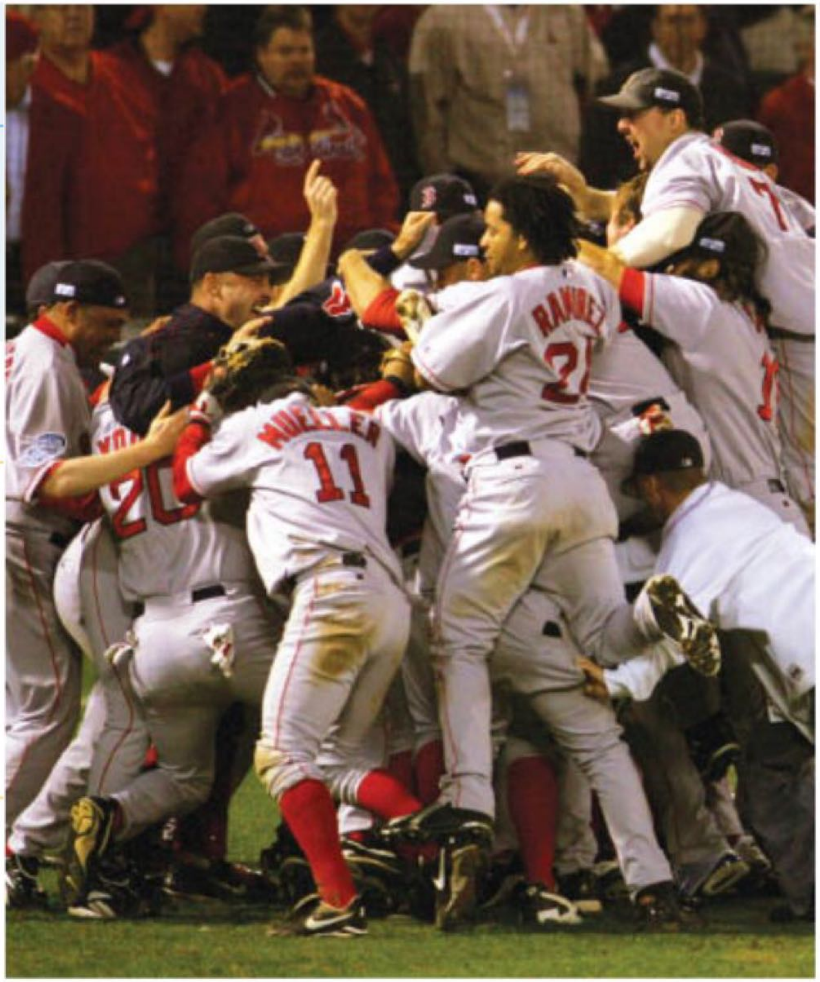
Who uses this?

Statisticians can use measures of central tendency and variation to analyze World Series results. (See Example 2.)



Lesson Objective(s):

- Find measures of central tendency and measures of variation for statistical data.
- Examine the effects of outliers on statistical data.

Recall that the *mean, median,* and *mode* are measures of central tendency—values that describe the center of a data set.

The *mean* is the sum of the values in the set divided by the number of values. It is often represented as $\bar{x}$. The *median* is the middle value or the mean of the two middle values when the set is ordered numerically. The *mode* is the value or values that occur most often. A data set may have one mode, no mode, or several modes.

EXAMPLE 1

Finding Measures of Central Tendency

Find the mean, median, and mode of the data.

Number of days from mailing to delivery: 6, 4, 3, 4, 2, 5, 3, 4, 5, 2, 3, 4

A *weighted average* of a data set gives greater importance, or weight, to some values in the set than to others. To find a weighted average, multiply each value by its weight. Then divide the sum of these products by the sum of the weights.

Suppose a teacher grades students' work in a class by using a weighted average in which homework has a weight of 30%, tests have a weight of 40%, and the final exam has a weight of 30%. Mia has a homework score of 84, a test score of 88, and a final exam score of 91.

$$\text{Mia's weighted average} = \frac{84(0.30) + 88(0.40) + 91(0.30)}{0.30 + 0.40 + 0.30} = \frac{87.7}{1.00} = 87.7$$

For an experiment with numerical outcomes, [ ] the weighted average of the possible outcomes. The weight for each outcome is its probability.

EXAMPLE 2

Finding Expected Value

The probability distribution for the number of games played in each World Series for the years 1923–2004 is given below. Find the expected number of games in a World Series.

World Series Games				
Number of Games n in World Series	4	5	6	7
Probability of n Games	$\frac{5}{27}$	$\frac{5}{27}$	$\frac{6}{27}$	$\frac{11}{27}$

A *box-and-whisker plot* shows the spread of a data set. It displays 5 key points: the minimum and maximum values, the median, and the first and third quartiles.

The quartiles are the medians of the lower and upper halves of the data set. If there are an odd number of data values, do not include the median in either half.

The *interquartile range*, or IQR, is the difference between the 1st and 3rd quartiles, or Q3 - Q1. It represents the middle 50% of the data.

EXAMPLE 3

Making a Box-and-Whisker Plot and Finding the Interquartile Range

Make a box-and-whisker plot of the data. Find the interquartile range.

$\{5, 3, 9, 2, 14, 6, 8, 9, 5, 8, 13, 3, 15, 7, 4, 2, 12, 8\}$

The data sets {19, 20, 21} and {0, 20, 40} have the same mean and median, but the sets are very different. The way that data are spread out from the mean or median is important in the study of statistics.

A *measure of variation* is a value that describes the spread of a data set. The most commonly used measures of variation are the *range*, the interquartile range, the *variance*, and the *standard deviation*.

______ denoted by σ^2 , is the average of the squared differences from the mean. ______ denoted by σ, is the square root of the variance and is one of the most common and useful measures of variation.

Low standard deviations indicate data that are clustered near the measures of central tendency, whereas high standard deviations indicate data that are spread out from the center.

Finding Variance and Standard Deviation	
Step 1.	
Step 2.	
Step 3.	Find the variance, σ^2, by adding the squares of all of the differences from the mean and dividing by the number of data values.
Step 4.	

EXAMPLE 4

Finding the Mean and Standard Deviation

The data represent the number of milligrams of a substance in a patient's blood, found on consecutive doctor visits. Find the mean and the standard deviation of the data.

$\{14, 13, 16, 9, 3, 7, 11, 12, 11, 4\}$

an extreme value that is much less than or much greater than the other data values. Outliers have a strong effect on the mean and standard deviation. If an outlier is the result of measurement error or represents data from the wrong population, it is usually removed. There are different ways to determine whether a value is an outlier. One is to look for data values that are more than 3 standard deviations from the mean.

EXAMPLE 5

Examining Outliers

The number of electoral votes in 2004 for 11 western states are shown. Find the mean and the standard deviation of the data. Identify any outliers, and describe how they affect the mean and the standard deviation.