

Statistics are numerical values used to summarize and compare sets of data. Two important types of statistics are *measures of central tendency* and *measures of dispersion*.

A **measure of central tendency** is a number used to represent the center or middle of a set of data values. The *mean*, *median*, and *mode* are three commonly used measures of central tendency.

Measures of Central Tendency

- The **mean**, or *average*, of n numbers is the sum of the numbers divided by n . The mean is denoted by \bar{x} , which is read as "x-bar." For the data set

$$x_1, x_2, \dots, x_n, \text{ the mean is } \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

- The **median** of n numbers is the middle number when the numbers are written in order. (If n is even, the median is the mean of the two middle numbers.)
- The **mode** of n numbers is the number or numbers that occur most frequently. There may be one mode, no mode, or more than one mode.

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

1. 1, 6, 3, 9, 6, 8, 4, 4, 4

Mean = _____ Median = _____ Mode = _____

2. 1, 5, 6, 2, 6, 1, 7, 6, 2

Mean = _____ Median = _____ Mode = _____

3. **Quiz Scores** The data set below gives the quiz scores for a student on quizzes consisting of 10 questions each.

7, 9, 7, 10, 8, 7, 9

Mean = _____ Median = _____ Mode = _____

4. **Travel Distance** The data set below gives the distance (in miles) that several people travel to and from work each day.

12, 15, 11, 8, 11, 13, 10, 16

Mean = _____ Median = _____ Mode = _____

5. Oil Change The data set below gives the waiting time (in minutes) of several people having the oil changed in their car at an auto mechanics shop.

22, 18, 25, 21, 28, 26, 20, 28, 20

Mean = _____ Median = _____ Mode = _____

6. Hockey The data set below gives the number of goals for the 10 players who scored the most goals during the 2003 -2004 National Hockey League regular season.

41, 41, 41, 38, 38, 36, 35, 35, 34, 33

Mean = _____ Median = _____ Mode = _____

Measures of Dispersion A *measure of dispersion* is a statistic that tells you how dispersed, or spread out, data values are. One simple measure of dispersion is the **range**, which is the difference between the greatest and least data values.

Standard Deviation Another measure of dispersion is *standard deviation*, which describes the typical difference (or *deviation*) between a data value and the mean.

Standard Deviation of a Data Set

The **standard deviation** σ (read as “sigma”) of x_1, x_2, \dots, x_n is:

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

Find the range and standard deviation of the data set.

7. 12, 8, 17, 15, 12, 14

Range = _____ σ = _____

8. 17, 14, 24, 21, 30, 20

Range = _____ σ = _____

9. 22, 24, 31, 34, 23, 27, 21

Range = _____ σ = _____

10. 31,46,39,43,32,35,40

Range = _____

σ = _____

Outliers Measures of central tendency and dispersion can give misleading impressions of a data set if the set contains one or more *outliers*. An **outlier** is a value that is much greater than or much less than most of the other values in a data set.

11. Consider the data set: 11, 15, 10, 37, 17, 14, 9, 15

a. Identify the outlier in the data set.

b. Find the range and standard deviation of the data set when the outlier is included and when it is not.

Outlier included: Range= _____ σ = _____

Outlier not included: Range= _____ σ = _____

Identify the outlier in the data set. Then find the mean, median, mode, range, and standard deviation of the data set when the outlier is included and when it is not.

12. 4, 6, 10, 2, 90, 3, 10, 5, 1

Outlier: _____

Outlier included: Mean = _____ Median = _____ Mode _____ Range= _____ σ = _____

Outlier not included: Mean = _____ Median = _____ Mode _____ Range= _____ σ = _____

13. 52,61,55,1,59,68,69,55

Outlier: _____

Outlier included: Mean = _____ Median = _____ Mode _____ Range= _____ σ = _____

Outlier not included: Mean = _____ Median = _____ Mode _____ Range= _____ σ = _____

Outlier Definition: An outlier can be defined as a value in a data set that lies more than three standard deviations from the mean. So, x is an outlier if $\frac{|x - \bar{x}|}{\sigma} > 3$. In parts (a)–(c), use this definition to identify the outlier(s) in the data set. *Justify* your answers mathematically.

14. a. 70, 55, 54, 74, 60, 58, 55, 56, 6, 62, 68, 94, 55, 82, 69, 74

b. 18, 20, 22, 25, 16, 40, 24, 19, 38, 3, 21, 27, 88, 24, 23, 26

c. 50, 93, 81, 84, 88, 85, 90, 99, 92, 199, 96, 89, 87, 94, 37