Mean, Median, and Mode

In many real-life situations, it is helpful to describe data by a single number that is most representative of the entire collection of numbers. Such a number is called a measure of central tendency. The most commonly used measures are as follows.

1. The **mean**, or **average**, of $n$ numbers is the sum of the numbers divided by $n$.

2. The **median** of $n$ numbers is the middle number when the numbers are written in order. If $n$ is even, the median is the average of the two middle numbers.

3. The **mode** of $n$ numbers is the number that occurs most frequently. If two numbers tie for most frequent occurrence, the collection has two modes and is called **bimodal**.

Example 1  ►  Comparing Measures of Central Tendency

On an interview for a job, the interviewer tells you that the average annual income of the company’s 25 employees is $60,849. The actual annual incomes of the 25 employees are shown below. What are the mean, median, and mode of the incomes? Was the person telling you the truth?

$17,305, \ $478,320, \ $45,678, \ $18,980, \ $17,408, \ $25,676, \ $28,906, \ $12,500, \ $24,540, \ $33,450, \ $12,500, \ $33,855, \ $37,450, \ $20,432, \ $28,956, \ $34,983, \ $36,540, \ $36,853, \ $37,450, \ $45,678, \ $48,980, \ $94,024, \ $35,671

Solution

The mean of the incomes is

$$
\text{Mean} = \frac{17,305 + 478,320 + 45,678 + 18,980 + \cdots + 35,671}{25} = \frac{1,521,225}{25} = 60,849.
$$

To find the median, order the incomes as follows.

$12,500, \ $12,500, \ $16,430, \ $17,305, \ $17,408, \ $18,980, \ $20,432, \ $24,540, \ $25,676, \ $28,906, \ $28,956, \ $32,654, \ $33,450, \ $33,855, \ $34,983, \ $35,671, \ $36,540, \ $36,853, \ $37,450, \ $45,678, \ $48,980, \ $94,024, \ $98,213, \ $98,213, \ $250,921, \ $478,320

From this list, you can see that the median (the middle number) is $33,450. From the same list, you can see that $12,500 is the only income that occurs more than once. So, the mode is $12,500. Technically, the person was telling the truth because the average is (generally) defined to be the mean. However, of the three measures of central tendency *Mean*: $60,849  *Median*: $33,450  *Mode*: $12,500 it seems clear that the median is most representative. The mean is inflated by the two highest salaries.
Choosing a Measure of Central Tendency

Which of the three measures of central tendency is the most representative? The answer is that it depends on the distribution of the data and the way in which you plan to use the data.

For instance, in Example 1, the mean salary of $60,849 does not seem very representative to a potential employee. To a city income tax collector who wants to estimate 1% of the total income of the 25 employees, however, the mean is precisely the right measure.

Example 2

Which measure of central tendency is the most representative of the data shown in each frequency distribution?

<table>
<thead>
<tr>
<th>a. Number</th>
<th>Tally</th>
<th>b. Number</th>
<th>Tally</th>
<th>c. Number</th>
<th>Tally</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>1</td>
<td>9</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>0</td>
</tr>
</tbody>
</table>

Solution

a. For this data, the mean is 4.23, the median is 3, and the mode is 2. Of these, the mode is probably the most representative.
b. For this data, the mean and median are each 5 and the modes are 1 and 9 (the distribution is bimodal). Of these, the mean or median is the most representative.
c. For this data, the mean is 4.59, the median is 5, and the mode is 1. Of these, the mean or median is the most representative.

Variance and Standard Deviation

Very different sets of numbers can have the same mean. You will now study two measures of dispersion, which give you an idea of how much the numbers in a set differ from the mean of the set. These two measures are called the variance of the set and the standard deviation of the set.

Definitions of Variance and Standard Deviation

Consider a set of numbers \( \{x_1, x_2, \ldots, x_n\} \) with a mean of \( \bar{x} \). The variance of the set is

\[
\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}
\]

and the standard deviation of the set is

\[
\sigma = \sqrt{\sigma^2}
\]

(\( \sigma \) is the lowercase Greek letter sigma).
The standard deviation of a set is a measure of how much a typical number in the set differs from the mean. The greater the standard deviation, the more the numbers in the set vary from the mean. For instance, each of the following sets has a mean of 5.

\{5, 5, 5\}, \{4, 4, 6, 6\}, and \{3, 3, 7, 7\}

The standard deviations of the sets are 0, 1, and 2.

\[\sigma_1 = \sqrt{\frac{(5-5)^2 + (5-5)^2 + (5-5)^2 + (5-5)^2}{4}} = 0\]

\[\sigma_2 = \sqrt{\frac{(4-5)^2 + (4-5)^2 + (6-5)^2 + (6-5)^2}{4}} = 1\]

\[\sigma_3 = \sqrt{\frac{(3-5)^2 + (3-5)^2 + (7-5)^2 + (7-5)^2}{4}} = 2\]

**Example 3 ➤ Estimations of Standard Deviation**

Consider the three sets of data represented by the bar graphs in Figure A.4. Which set has the smallest standard deviation? Which has the largest?

**Solution**

Of the three sets, the numbers in set A are grouped most closely to the center and the numbers in set C are the most dispersed. So, set A has the smallest standard deviation and set C has the largest standard deviation.
Example 4  
Finding Standard Deviation

Find the standard deviation of each set shown in Example 3.

Solution

Because of the symmetry of each bar graph, you can conclude that each has a mean of \( \bar{x} = 4 \). The standard deviation of set \( A \) is

\[
\sigma = \sqrt{\frac{(-3)^2 + 2(-2)^2 + 3(-1)^2 + 5(0)^2 + 3(1)^2 + 2(2)^2 + (3)^2}{17}}
\]

\[
\approx 1.53.
\]

The standard deviation of set \( B \) is

\[
\sigma = \sqrt{\frac{2(-3)^2 + 2(-2)^2 + 2(-1)^2 + 2(0)^2 + 2(1)^2 + 2(2)^2 + 2(3)^2}{14}}
\]

\[
= 2.
\]

The standard deviation of set \( C \) is

\[
\sigma = \sqrt{\frac{5(-3)^2 + 4(-2)^2 + 3(-1)^2 + 2(0)^2 + 3(1)^2 + 4(2)^2 + 5(3)^2}{26}}
\]

\[
\approx 2.22.
\]

These values confirm the results of Example 3. That is, set \( A \) has the smallest standard deviation and set \( C \) has the largest.

The following alternative formula provides a more efficient way to compute the standard deviation.

Alternative Formula for Standard Deviation

The standard deviation of \( \{x_1, x_2, \ldots, x_n\} \) is

\[
\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n} - \bar{x}^2}.
\]

Because of messy computations, this formula is difficult to verify. Conceptually, however, the process is straightforward. It consists of showing that the expressions

\[
\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}
\]

and

\[
\sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n} - \bar{x}^2}
\]

are equivalent. Try verifying this equivalence for the set \( \{x_1, x_2, x_3\} \) with \( \bar{x} = (x_1 + x_2 + x_3)/3 \).
Example 5  Using the Alternative Formula

Use the alternative formula for standard deviation to find the standard deviation of the following set of numbers.

5, 6, 6, 7, 7, 8, 8, 8, 9, 10

Solution

Begin by finding the mean of the set, which is 7.4. So, the standard deviation is

\[
\sigma = \sqrt{\frac{5^2 + 2(6^2) + 2(7^2) + 3(8^2) + 9^2 + 10^2}{10} - (7.4)^2} \\
= \sqrt{\frac{568}{10}} - 54.76 \\
= \sqrt{2.04} \\
\approx 1.43.
\]

You can use the statistical features of a graphing utility to check this result.

A well-known theorem in statistics, called Chebychev’s Theorem, states that at least

\[
1 - \frac{1}{k^2}
\]

of the numbers in a distribution must lie within \(k\) standard deviations of the mean. So, 75% of the numbers in a set must lie within two standard deviations of the mean, and at least 88.9% of the numbers must lie within three standard deviations of the mean. For most distributions, these percentages are low. For instance, in all three distributions shown in Example 3, 100% of the numbers lie within two standard deviations of the mean.

Example 6  Describing a Distribution

The table at the left above shows the number of hospitals (in thousands) in each state and the District of Columbia in 1999. Find the mean and standard deviation of the numbers. What percent of the numbers lie within two standard deviations of the mean? (Source: Health Forum)

Solution

Begin by entering the numbers into a graphing utility that has a standard deviation program. After running the program, you should obtain

\[\bar{x} \approx 97.18 \quad \text{and} \quad \sigma = 81.99.\]

The interval that contains all numbers that lie within two standard deviations of the mean is

\[[97.18 - 2(81.99), 97.18 + 2(81.99)] \quad \text{or} \quad [-66.80, 261.16].\]

From the histogram in Figure A.5, you can see that all but two of the numbers (96%) lie in this interval—all but the numbers that correspond to the number of hospitals (in thousands) in California and Texas.
Box-and-Whisker Plots

Standard deviation is the measure of dispersion that is associated with the mean. **Quartiles** measure dispersion associated with the median.

**Definition of Quartiles**

Consider an ordered set of numbers whose median is \( m \). The **lower quartile** is the median of the numbers that occur before \( m \). The **upper quartile** is the median of the numbers that occur after \( m \).

**Example 7 ► Finding Quartiles of a Set**

Find the lower and upper quartiles for the set.

\[ 34, 14, 24, 16, 12, 18, 20, 24, 16, 26, 13, 27 \]

**Solution**

Begin by ordering the set.

\[ 12, 13, 14, 16, 16, 18, 20, 24, 24, 26, 27, 34 \]

1st 25% 2nd 25% 3rd 25% 4th 25%

The median of the entire set is 19. The median of the six numbers that are less than 19 is 15. So, the lower quartile is 15. The median of the six numbers that are greater than 19 is 25. So, the upper quartile is 25.

Quartiles are represented graphically by a **box-and-whisker plot**, as shown in Figure A.6. In the plot, notice that five numbers are listed: the smallest number, the lower quartile, the median, the upper quartile, and the largest number. Also notice that the numbers are spaced proportionally, as though they were on a real number line.

![Box-and-Whisker Plot](image)

The next example shows how to find quartiles when the number of elements in a set is not divisible by 4.
**Example 8  Sketching Box-and-Whisker Plots**

Sketch a box-and-whisker plot for each set.

a. 27, 28, 30, 42, 45, 50, 50, 61, 62, 64, 66  
b. 82, 82, 83, 85, 87, 89, 90, 94, 95, 95, 96, 98, 99  
c. 11, 13, 13, 15, 17, 18, 20, 24, 24, 27

**Solution**  

a. This set has 11 numbers. The median is 50 (the sixth number). The lower quartile is 30 (the median of the first five numbers). The upper quartile is 62 (the median of the last five numbers). See Figure A.7.

![Figure A.7](image1)

b. This set has 13 numbers. The median is 90 (the seventh number). The lower quartile is 84 (the median of the first six numbers). The upper quartile is 95.5 (the median of the last six numbers). See Figure A.8.

![Figure A.8](image2)

c. This set has 10 numbers. The median is 17.5 (the average of the fifth and sixth numbers). The lower quartile is 13 (the median of the first five numbers). The upper quartile is 24 (the median of the last five numbers). See Figure A.9.

![Figure A.9](image3)

**A.2 Exercises**

In Exercises 1–6, find the mean, median, and mode of the set of measurements.

1. 5, 12, 7, 14, 8, 9, 7  
2. 30, 37, 32, 39, 33, 34, 32  
3. 5, 12, 7, 24, 8, 9, 7  
4. 20, 37, 32, 39, 33, 34, 32  
5. 5, 12, 7, 14, 9, 7  
6. 30, 37, 32, 39, 34, 32

7. **Reasoning** Compare your answers for Exercises 1 and 3 with those for Exercises 2 and 4. Which of the measures of central tendency is sensitive to extreme measurements? Explain your reasoning.

8. **Reasoning**  
   (a) Add 6 to each measurement in Exercise 1 and calculate the mean, median, and mode of the revised measurements. How are the measures of central tendency changed?  
   (b) If a constant $k$ is added to each measurement in a set of data, how will the measures of central tendency change?
9. **Electric Bills** A person had the following monthly bills for electricity. What are the mean and median of the collection of bills?

<table>
<thead>
<tr>
<th>Month</th>
<th>Bill</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>$67.92</td>
</tr>
<tr>
<td>February</td>
<td>$59.84</td>
</tr>
<tr>
<td>March</td>
<td>$52.00</td>
</tr>
<tr>
<td>April</td>
<td>$52.50</td>
</tr>
<tr>
<td>May</td>
<td>$57.99</td>
</tr>
<tr>
<td>June</td>
<td>$65.35</td>
</tr>
<tr>
<td>July</td>
<td>$81.76</td>
</tr>
<tr>
<td>August</td>
<td>$74.98</td>
</tr>
<tr>
<td>September</td>
<td>$87.82</td>
</tr>
<tr>
<td>October</td>
<td>$83.18</td>
</tr>
<tr>
<td>November</td>
<td>$65.35</td>
</tr>
</tbody>
</table>

10. **Car Rental** A car rental company kept the following record of the numbers of miles a rental car was driven. What are the mean, median, and mode of this data?

<table>
<thead>
<tr>
<th>Day</th>
<th>Miles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>410</td>
</tr>
<tr>
<td>Tuesday</td>
<td>260</td>
</tr>
<tr>
<td>Wednesday</td>
<td>320</td>
</tr>
<tr>
<td>Thursday</td>
<td>320</td>
</tr>
<tr>
<td>Friday</td>
<td>460</td>
</tr>
<tr>
<td>Saturday</td>
<td>150</td>
</tr>
</tbody>
</table>

11. **Six-Child Families** A study was done on families having six children. The table shows the numbers of families in the study with the indicated numbers of girls. Determine the mean, median, and mode of this set of data.

<table>
<thead>
<tr>
<th>Number of girls</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

12. **Sports** A baseball fan examined the records of a favorite baseball player’s performance during his last 50 games. The numbers of games in which the player had 0, 1, 2, 3, and 4 hits are recorded in the table.

<table>
<thead>
<tr>
<th>Number of hits</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

(a) Determine the average number of hits per game.
(b) Determine the player’s batting average if he had 200 at-bats during the 50-game series.

13. **Think About It** Construct a collection of numbers that has the following properties. If this is not possible, explain why it is not.
Mean = 6, median = 4, mode = 4

14. **Think About It** Construct a collection of numbers that has the following properties. If this is not possible, explain why it is not.
Mean = 6, median = 6, mode = 4

15. **Test Scores** A professor records the following scores for a 100-point exam.
99, 64, 80, 77, 59, 72, 87, 79, 92, 88, 90, 42, 20, 89, 42, 100, 98, 84, 78, 91
Which measure of central tendency best describes these test scores?

16. **Shoe Sales** A salesman sold eight pairs of men’s black dress shoes. The sizes of the eight pairs were as follows: 9 1/2, 8, 12, 10, 9 1/2, 7, 11, and 10 1/2. Which measure (or measures) of central tendency best describes the typical shoe size for this data?

In Exercises 17–24, find the mean ($\bar{x}$), variance ($\nu$), and standard deviation ($\sigma$) of the set.

17. 4, 10, 8, 2
18. 3, 15, 6, 9, 2
19. 0, 1, 1, 2, 2, 3, 3, 4
20. 2, 2, 2, 2, 2, 2
21. 1, 2, 3, 4, 5, 6, 7
22. 1, 1, 1, 5, 5, 5
23. 49, 62, 40, 29, 32, 70
24. 1.5, 0.4, 2.1, 0.7, 0.8

In Exercises 25–30, use the alternative formula to find the standard deviation of the set.

25. 2, 4, 6, 6, 13, 5
26. 10, 25, 50, 26, 15, 33, 29, 4
27. 246, 336, 473, 167, 219, 359
28. 6.0, 9.1, 4.4, 8.7, 10.4
29. 8.1, 6.9, 3.7, 4.2, 6.1
30. 9.0, 7.5, 3.3, 7.4, 6.0

In Exercises 31 and 32, line plots of sets of data are given. Determine the mean and standard deviation of each set.

31. (a) 
(b) 
(c) 
(d)
32. (a) [Diagram of data points]
(b) [Diagram of data points]
(c) [Diagram of data points]
(d) [Diagram of data points]

33. **Reasoning** Without calculating the standard deviation, explain why the set \{4, 4, 20, 20\} has a standard deviation of 8.

34. **Reasoning** If the standard deviation of a set of numbers is 0, what does this imply about the set?

35. **Test Scores** An instructor adds five points to each student’s exam score. Will this change the mean or standard deviation of the exam scores? Explain.

36. **Price of Gold** The following data represents the average prices of gold (in dollars per fine ounce) for the years 1981 to 2000. Use a computer or graphing utility to find the mean, variance, and standard deviation of the data. What percent of the data lies within two standard deviations of the mean? (Source: U.S. Bureau of Mines and U.S. Geological Survey)

   460, 376, 424, 361, 318,
   368, 478, 438, 383, 385,
   363, 345, 361, 385, 386,
   389, 332, 295, 280, 280

37. **Think About It** The histograms represent the test scores of two classes of a college course in mathematics. Which histogram has the smaller standard deviation?

38. **Test Scores** The scores of a mathematics exam given to 600 science and engineering students at a college had a mean and standard deviation of 235 and 28, respectively. Use Chebychev’s Theorem to determine the intervals containing at least \(\frac{1}{2}\) and at least \(\frac{3}{4}\) of the scores. How would the intervals change if the standard deviation were 16?

In Exercises 39–42, sketch a box-and-whisker plot for the data without the aid of a graphing utility.

39. 23, 15, 14, 23, 13, 14, 13, 20, 12
40. 11, 10, 11, 14, 17, 16, 14, 11, 8, 14, 20
41. 46, 48, 48, 52, 47, 51, 47, 49, 53
42. 25, 20, 22, 28, 24, 28, 25, 19, 27, 29, 28, 21

In Exercises 43–46, use a graphing utility to create a box-and-whisker plot for the data.

43. 19, 12, 14, 9, 14, 15, 17, 13, 19, 11, 10, 19
44. 9, 5, 5, 5, 6, 5, 4, 12, 7, 10, 7, 11, 8, 9, 9
45. 20.1, 43.4, 34.9, 23.9, 33.5, 24.1, 22.5, 42.4, 25.7, 17.4, 23.8, 33.3, 17.3, 36.4, 21.8
46. 78.4, 76.3, 107.5, 78.5, 93.2, 90.3, 77.8, 37.1, 97.1, 75.5, 58.8, 65.6

47. **Product Lifetime** A company has redesigned a product in an attempt to increase the lifetime of the product. The two sets of data list the lifetimes (in months) of 20 units with the original design and 20 units with the new design. Create a box-and-whisker plot for each set of data, and then comment on the differences between the plots.

   **Original Design**
   15.1 78.3 56.3 68.9 30.6
   27.2 12.5 42.7 72.7 20.2
   53.0 13.5 11.0 18.4 85.2
   10.8 38.3 85.1 10.0 12.6

   **New Design**
   55.8 71.5 25.6 19.0 23.1
   37.2 60.0 35.3 18.9 80.5
   46.7 31.1 67.9 23.5 99.5
   54.0 23.2 45.5 24.8 87.8