

Study Guide and Intervention

Statistical Measures

Measures of Central Tendency

| Measures of Central Tendency | Use | When |
|------------------------------|--------|---|
| | mean | the data are spread out and you want an average of values |
| | median | the data contain outliers |
| | mode | the data are tightly clustered around one or two values |

Example

Find the mean, median, and mode of the following set of data: {42, 39, 35, 40, 38, 35, 45}.

To find the mean, add the values and divide by the number of values.

$$\text{mean} = \frac{42 + 39 + 35 + 40 + 38 + 35 + 45}{7} \approx 39.14.$$

To find the median, arrange the values in ascending or descending order and choose the middle value. (If there is an even number of values, find the mean of the two middle values.) In this case, the median is 39.

To find the mode, take the most common value. In this case, the mode is 35.

Exercises

Find the mean, median, and mode of each set of data. Round to the nearest hundredth, if necessary.

- {238, 261, 245, 249, 255, 262, 241, 245}
- {9, 13, 8, 10, 11, 9, 12, 16, 10, 9}
- {120, 108, 145, 129, 102, 132, 134, 118, 108, 142}
- {68, 54, 73, 58, 63, 72, 65, 70, 61}
- {34, 49, 42, 38, 40, 45, 34, 28, 43, 30}

- The table at the right shows the populations of the six New England capitals. Which would be the most appropriate measure of central tendency to represent the data? Explain why and find that value.

Source: www.factfinder.census.gov

| City | Population (rounded to the nearest 1000) |
|----------------|--|
| Augusta, ME | 19,000 |
| Boston, MA | 589,000 |
| Concord, NH | 37,000 |
| Hartford, CT | 122,000 |
| Montpelier, VT | 8,000 |
| Providence, RI | 174,000 |

Study Guide and Intervention *(continued)*

Statistical Measures

Measures of Variation The *range* and the **standard deviation** measure how scattered a set of data is.

| | |
|---------------------------|---|
| Standard Deviation | If a set of data consists of the n values x_1, x_2, \dots, x_n and has mean \bar{x} , then the standard deviation is given by $\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$. |
|---------------------------|---|

The square of the standard deviation is called the **variance**.

Example

Find the variance and standard deviation of the data set

{10, 9, 6, 9, 18, 4, 8, 20}.

Step 1 Find the mean.

$$\bar{x} = \frac{10 + 9 + 6 + 9 + 18 + 4 + 8 + 20}{8} = 10.5$$

Step 2 Find the variance.

$$\begin{aligned} \sigma^2 &= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} && \text{Variance formula} \\ &= \frac{(10 - 10.5)^2 + (9 - 10.5)^2 + \dots + (20 - 10.5)^2}{8} \\ &= \frac{220}{8} \text{ or } 27.5 \end{aligned}$$

Step 3 Find the standard deviation.

$$\begin{aligned} \sigma &= \sqrt{27.5} \\ &\approx 5.2 \end{aligned}$$

The variance is 27.5 and the standard deviation is about 5.2.

Exercises

Find the variance and standard deviation of each set of data. Round to the nearest tenth.

1. {100, 89, 112, 104, 96, 108, 93}
2. {62, 54, 49, 62, 48, 53, 50}
3. {8, 9, 8, 8, 9, 7, 8, 9, 6}
4. {4.2, 5.0, 4.7, 4.5, 5.2, 4.8, 4.6, 5.1}

5. The table at the right lists the prices of ten brands of breakfast cereal. What is the standard deviation of the values to the nearest penny?

6. Conduct a small survey among ten of your friends. You'll want to:
- Choose a topic. Make sure that the responses can be measured.
 - Ask ten of your friends for their input.
 - Find the mean, median, and mode of the responses.
 - Find the variance and standard deviation of the responses.
- What conclusions can you draw about the results of your survey?

| Price of 10 Brands of Breakfast Cereal | |
|--|--------|
| \$2.29 | \$3.19 |
| \$3.39 | \$2.79 |
| \$2.99 | \$3.09 |
| \$3.19 | \$2.59 |
| \$2.79 | \$3.29 |

Skills Practice

Statistical Measures

Find the variance and standard deviation of each set of data to the nearest tenth.

- {32, 41, 35, 35, 46, 42}
- {13, 62, 77, 24, 38, 19, 88}
- {89, 99, 42, 16, 42, 71, 16}
- {450, 400, 625, 225, 300, 750, 650, 625}
- {17, 23, 65, 94, 33, 33, 33, 8, 57, 75, 44, 12, 11, 68, 39}
- {7.2, 3.1, 3.8, 9.5, 8.3, 8.4}
- {1.5, 2.5, 3.5, 4.5, 4.5, 5.5, 6.5, 7.5}

For Exercises 8 and 9, use the table that shows the profit in billions of dollars reported by U.S. manufacturers for the first quarter of the years from 1997 through 2001.

| Year | 1997 | 1998 | 1999 | 2000 | 2001 |
|--|--------|--------|--------|--------|--------|
| Seasonally-Adjusted Profit (\$ billions) | \$61.4 | \$75.6 | \$60.9 | \$78.5 | \$45.3 |

Source: U. S. Census Bureau

- Find the mean and median of the data to the nearest tenth.
- Which measure of central tendency best represents the data? Explain.

For Exercises 10 and 11, use the table that shows the percent of fourth grade students reading at or above the proficiency level in a nationally-administered reading assessment.

| Year | 1992 | 1994 | 1998 | 2000 |
|---------------------------------------|------|------|------|------|
| Percent at or above proficiency level | 29% | 30% | 31% | 32% |

Source: National Center for Education Statistics

- Find the mean, median, and standard deviation of the data to the nearest tenth.
- What do the statistics from Exercise 11 tell you about the data?

Practice

Statistical Measures

Find the variance and standard deviation of each set of data to the nearest tenth.

1. {47, 61, 93, 22, 82, 22, 37}
2. {10, 10, 54, 39, 96, 91, 91, 18}
3. {1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5}
4. {1100, 725, 850, 335, 700, 800, 950}
5. {3.4, 7.1, 8.5, 5.1, 4.7, 6.3, 9.9, 8.4, 3.6}
6. {2.8, 0.5, 1.9, 0.8, 1.9, 1.5, 3.3, 2.6, 0.7, 2.5}

7. HEALTH CARE Eight physicians with 15 patients on a hospital floor see these patients an average of 18 minutes a day. The 22 nurses on the same floor see the patients an average of 3 hours a day. As a hospital administrator, would you quote the mean, median, or mode as an indicator of the amount of daily medical attention the patients on this floor receive? Explain.

For Exercises 8–10, use the frequency table that shows the percent of public school teachers in the U. S. in 1999 who used computers or the Internet at school for various administrative and teaching activities.

| Activity | Percent Using Computer or Internet |
|---|------------------------------------|
| Create instructional materials | 39 |
| Administrative record keeping | 34 |
| Communicate with colleagues | 23 |
| Gather information for planning lessons | 16 |
| Multimedia classroom presentations | 8 |
| Access research and best practices for teaching | 8 |
| Communicate with parents or students | 8 |
| Access model lesson plans | 6 |

Source: National Assessment of Educational Progress

8. Find the mean, median, and mode of the data.
9. Suppose you believe teachers use computers or the Internet too infrequently. Which measure would you quote as the “average?” Explain.
10. Suppose you believe teachers use computers or the Internet too often. Which measure would you quote as the “average?” Explain.

For Exercises 11 and 12, use the frequency table that shows the number of games played by 24 American League baseball players between opening day, 2001 and September 8, 2001.

| No. of Games | Frequency |
|--------------|-----------|
| 141 | 4 |
| 140 | 3 |
| 139 | 4 |
| 138 | 5 |
| 137 | 2 |
| 136 | 3 |
| 135 | 3 |

11. Find the mean, median, mode, and standard deviation of the number of games played to the nearest tenth.
12. For how many players is the number of games within one standard deviation of the mean?

Source: Major League Baseball

Reading to Learn Mathematics

Statistical Measures

Reading the Lesson

1. Suppose you had to design a small survey and then collect, organize, and interpret the data.
 - a. How would you choose a topic?
 - b. What does it mean if the responses are said to be *measurable*?
 - c. What can you learn by finding the mean, median, and mode of the responses?
 - d. What can you learn by finding the mean, median, and mode of the responses?
2. Match each measure with one of the six descriptions of how to find measures of central tendency and variation.

| | | |
|-------------|---------|-----------------------|
| a. median | b. mode | c. range |
| d. variance | e. mean | f. standard deviation |

 - i. Find the most commonly occurring values or values in a set of data.
 - ii. Add the data and divide by the number of items.
 - iii. Find the mean of the squares of the differences between each value in the set of data and the mean.
 - iv. Find the difference between the largest and smallest values in the set of data.
 - v. Take the positive square root of the variance.
 - vi. If there is an odd number of items in a set of data, take the middle one. If there is an even number of items, add the two middle items and divide by 2.

Helping You Remember

3. It is usually easier to remember a complicated procedure if you break it down into steps. Write the procedure for finding the standard deviation for a set of data in a series of brief, numbered steps.

Enrichment

Probabilities in Genetics

Genes are the units which transmit hereditary traits. The possible forms which a gene may take, **dominant** and **recessive**, are called **alleles**. A particular trait is determined by two alleles, one from the female parent and one from the male parent. If an organism has the trait which is dominant, it may have either two dominant alleles or one dominant and one recessive allele. If the organism has the trait which is recessive, it must have two recessive alleles.

Example

Consider a plant in which tall stems, T , are dominant to short stems, t . What is the probability of obtaining a long-stemmed plant if two long-stemmed plants both with the genetic formula Tt are crossed?

| | | |
|-----|------|------|
| | T | t |
| T | TT | Tt |
| t | Tt | tt |

A *Punnett square* is a chart used to determine the possible combinations of characteristics among offspring.

$$\begin{array}{r} 3 \text{ tall-stemmed} \\ + 1 \text{ short-stemmed} \\ \hline 4 \text{ total} \end{array}$$

Thus, the probability is $\frac{3}{4}$.

In a certain plant, red flowers, R , are dominant to white flowers, r . If a white-flowered plant, rr is crossed with a red-flowered plant, Rr , find the probability of each of the following.

- white-flowered plant
- red-flowered plant

In a certain plant, tall, T , is dominant to short, t , and green pods, G , are dominant to yellow pods, g . Plants with the genetic formulas $TtGg$ and $TTGg$ are crossed. Find the probability of each of the following.

- tall plant with green pods
- tall plant with yellow pods

NAME _____ DATE _____ PERIOD _____

Study Guide and Intervention *(continued)*

Statistical Measures

Measures of Variation

The *range* and the **standard deviation** measure how scattered a set of data is.

| Standard Deviation | When |
|---|---|
| is given by $\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$ | If a set of data consists of the n values x_1, x_2, \dots, x_n and has mean \bar{x} , then the standard deviation |

The square of the standard deviation is called the **variance**.

Example Find the variance and standard deviation of the data set {10, 9, 6, 9, 18, 4, 8, 20}.

Step 1 Find the mean.

$$\bar{x} = \frac{10 + 9 + 6 + 9 + 18 + 4 + 8 + 20}{8} = 10.5$$

Step 2 Find the variance.

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

$$= \frac{(10 - 10.5)^2 + (9 - 10.5)^2 + \dots + (20 - 10.5)^2}{8}$$

$$= \frac{220}{8} \text{ or } 27.5$$

Step 3 Find the standard deviation.

$$\sigma = \sqrt{27.5} \approx 5.2$$

The variance is 27.5 and the standard deviation is about 5.2.

Exercises

Find the variance and standard deviation of each set of data. Round to the nearest tenth.

- {100, 89, 112, 104, 96, 108, 93} 2. {62, 54, 49, 62, 48, 53, 50}
58.5; 7.6 **29.4; 5.4**
- {8, 9, 8, 8, 9, 7, 8, 9, 6} 4. {4.2, 5.0, 4.7, 4.5, 5.2, 4.8, 4.6, 5.1}
0.9; 0.9 **0.1; 0.3**

5. The table at the right lists the prices of ten brands of breakfast cereal. What is the standard deviation of the values to the nearest penny? **\$0.33**

6. Conduct a small survey among ten of your friends. You'll want to:

- Choose a topic. Make sure that the responses can be measured.
- Ask ten of your friends for their input.
- Find the mean, median, and mode of the responses.
- Find the variance and standard deviation of the responses.

What conclusions can you draw about the results of your survey?
See students' work.

| Price of 10 Brands of Breakfast Cereal | Price |
|--|--------|
| | \$2.29 |
| | \$3.39 |
| | \$2.99 |
| | \$3.19 |
| | \$2.79 |
| | \$3.09 |
| | \$2.59 |
| | \$3.29 |

NAME _____ DATE _____ PERIOD _____

Study Guide and Intervention

Statistical Measures

Measures of Central Tendency

| Measures of Central Tendency | Use | When |
|------------------------------|---|------|
| mean | the data are spread out and you want an average of values | |
| median | the data contain outliers | |
| mode | the data are tightly clustered around one or two values | |

Example Find the mean, median, and mode of the following set of data: {42, 39, 35, 40, 38, 35, 45}.

To find the mean, add the values and divide by the number of values.

$$\text{mean} = \frac{42 + 39 + 35 + 40 + 38 + 35 + 45}{7} \approx 39.14.$$

To find the median, arrange the values in ascending or descending order and choose the middle value. (If there is an even number of values, find the mean of the two middle values.) In this case, the median is 39.

To find the mode, take the most common value. In this case, the mode is 35.

Exercises

Find the mean, median, and mode of each set of data. Round to the nearest hundredth, if necessary.

- {238, 261, 245, 249, 255, 262, 241, 245} **249.5; 247; 245**
- {9, 13, 8, 10, 11, 9, 12, 16, 10, 9} **10.7; 10; 9**
- {120, 108, 145, 129, 102, 132, 134, 118, 108, 142} **123.8; 124.5; 108**
- {68, 54, 73, 58, 63, 72, 65, 70, 61} **64.89; 65; no mode**
- {34, 49, 42, 38, 40, 45, 34, 28, 43, 30} **38.3; 39; 34**

6. The table at the right shows the populations of the six New England capitals. Which would be the most appropriate measure of central tendency to represent the data? Explain why and find that value.

Source: www.factfinder.census.gov. **There is no mode. The population of Boston is an outlier and would raise the mean too high. The median, 79,500, would be the best choice.**

| City | Population (rounded to the nearest 1000) |
|----------------|--|
| Augusta, ME | 19,000 |
| Boston, MA | 589,000 |
| Concord, NH | 37,000 |
| Hartford, CT | 122,000 |
| Montpelier, VT | 8,000 |
| Providence, RI | 174,000 |

NAME _____

DATE _____

PERIOD _____

Skills Practice

Statistical Measures

Find the variance and standard deviation of each set of data to the nearest tenth.

- {32, 41, 35, 35, 46, 42} **23.6, 4.9**
- {13, 62, 77, 24, 38, 19, 88} **763.8, 27.6**
- {89, 99, 42, 16, 42, 71, 16} **959.1, 31.0**
- {450, 400, 625, 225, 300, 750, 650, 625} **30,537.1; 174.7**
- {17, 23, 65, 94, 33, 33, 33, 8, 57, 75, 44, 12, 11, 68, 39} **630.7, 25.1**
- {7.2, 3.1, 3.8, 9.5, 8.3, 8.4} **5.8, 2.4**
- {1.5, 2.5, 3.5, 4.5, 4.5, 5.5, 6.5, 7.5} **3.5, 1.9**

For Exercises 8 and 9, use the table that shows the profit in billions of dollars reported by U.S. manufacturers for the first quarter of the years from 1997 through 2001.

| Year | 1997 | 1998 | 1999 | 2000 | 2001 |
|--|--------|--------|--------|--------|--------|
| Seasonally-Adjusted Profit (\$ billions) | \$61.4 | \$75.6 | \$60.9 | \$78.5 | \$45.3 |

Source: U. S. Census Bureau

- Find the mean and median of the data to the nearest tenth. **\$64.3 billion, \$61.4 billion**
- Which measure of central tendency best represents the data? Explain.
The median is more representative because the value 45.3 is not close to the other data points, and it lowers the mean.

For Exercises 10 and 11, use the table that shows the percent of fourth grade students reading at or above the proficiency level in a nationally-administered reading assessment.

| Year | 1992 | 1994 | 1998 | 2000 |
|---------------------------------------|------|------|------|------|
| Percent at or above proficiency level | 29% | 30% | 31% | 32% |

Source: National Center for Education Statistics

- Find the mean, median, and standard deviation of the data to the nearest tenth.
30.5%, 30.5%, 1.1%
- What do the statistics from Exercise 11 tell you about the data?
Sample answer: Since the median and mean are equal and the standard deviation is small, the percent of students reading at or above the proficiency level has not varied much from 1992 to 2000.

© Glencoe/McGraw-Hill

Algebra: Concepts and Applications

NAME _____

DATE _____

PERIOD _____

Practice (Average)

Statistical Measures

Find the variance and standard deviation of each set of data to the nearest tenth.

- {47, 61, 93, 22, 82, 22, 37} **673.1, 25.9**
- {10, 10, 54, 39, 96, 91, 91, 18} **1228.6, 35.1**
- {1, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5} **49,150.0; 221.7**
- {1100, 725, 850, 335, 700, 800, 950} **1.6, 1.2**
- {3.4, 7.1, 8.5, 5.1, 4.7, 6.3, 9.9, 8.4, 3.6} **0.8, 0.9**
- {2.8, 0.5, 1.9, 0.8, 1.9, 1.5, 3.3, 2.6, 0.7, 2.5} **4.7, 2.2**

7. HEALTH CARE Eight physicians with 15 patients on a hospital floor see these patients an average of 18 minutes a day. The 22 nurses on the same floor see the patients an average of 3 hours a day. As a hospital administrator, would you quote the mean, median, or mode as an indicator of the amount of daily medical attention the patients on this floor receive? Explain. **Either the median or the mode; they are equal and higher than the mean, which is lowered by the smaller amount of time the physicians spend with the patients.**

For Exercises 8–10, use the frequency table that shows the percent of public school teachers in the U. S. in 1999 who used computers or the Internet at school for various administrative and teaching activities.

| Activity | Percent Using Computer or Internet |
|---|------------------------------------|
| Create instructional materials | 39 |
| Administrative record keeping | 34 |
| Communicate with colleagues | 23 |
| Gather information for planning lessons | 16 |
| Multimedia classroom presentations | 8 |
| Access research and best practices for teaching | 8 |
| Communicate with parents or students | 8 |
| Access model lesson plans | 6 |

Source: National Assessment of Educational Progress

- Find the mean, median, and mode of the data. **17.75%, 12%, 8%**
- Suppose you believe teachers use computers or the Internet too infrequently. Which measure would you quote as the “average?” Explain. **Mode; it is lowest.**
- Suppose you believe teachers use computers or the Internet too often. Which measure would you quote as the “average?” Explain. **Mean; it is highest.**

For Exercises 11 and 12, use the frequency table that shows the number of games played by 24 American League baseball players between opening day, 2001 and September 8, 2001.

| No. of Games | Frequency |
|--------------|-----------|
| 141 | 4 |
| 140 | 3 |
| 139 | 4 |
| 138 | 5 |
| 137 | 2 |
| 136 | 3 |
| 135 | 3 |

Source: Major League Baseball

- Find the mean, median, mode, and standard deviation of the number of games played to the nearest tenth.
138.2, 138; 138, 2.0
- For how many players is the number of games within one standard deviation of the mean? **14**

© Glencoe/McGraw-Hill

Algebra: Concepts and Applications

Reading to Learn Mathematics

Statistical Measures

Reading the Lesson

- Suppose you had to design a small survey and then collect, organize, and interpret the data.
 - How would you choose a topic? **Sample answer: I would choose something that interests me.**
 - What does it mean if the responses are said to be *measurable*? **They will need to be numeric, or you will need to use numbers to analyze them.**
 - What can you learn by finding the mean, median, and mode of the responses? **These measures will help you learn what the average response is.**
 - What can you learn by finding the mean, median, and mode of the responses? **These measures will help you see the spread of the responses.**
- Match each measure with one of the six descriptions of how to find measures of central tendency and variation.

| | | |
|------------------------|-------------------|--------------------------------|
| a. median vi | b. mode i | c. range iv |
| d. variance iii | e. mean ii | f. standard deviation v |

 - Find the most commonly occurring values or values in a set of data.
 - Add the data and divide by the number of items.
 - Find the mean of the squares of the differences between each value in the set of data and the mean.
 - Find the difference between the largest and smallest values in the set of data.
 - Take the positive square root of the variance.
 - If there is an odd number of items in a set of data, take the middle one. If there is an even number of items, add the two middle items and divide by 2.

Helping You Remember

- It is usually easier to remember a complicated procedure if you break it down into steps. Write the procedure for finding the standard deviation for a set of data in a series of brief, numbered steps.

Sample answer:

- Find the mean.
- Find the difference between each value and the mean.
- Square each difference.
- Find the mean of the squares.
- Take the positive square root.

Enrichment

Probabilities in Genetics

Genes are the units which transmit hereditary traits. The possible forms which a gene may take, **dominant** and **recessive**, are called **alleles**. A particular trait is determined by two alleles, one from the female parent and one from the male parent. If an organism has the trait which is dominant, it may have either two dominant alleles or one dominant and one recessive allele. If the organism has the trait which is recessive, it must have two recessive alleles.

Example

Consider a plant in which tall stems, T , are dominant to short stems, t . What is the probability of obtaining a long-stemmed plant if two long-stemmed plants both with the genetic formula Tt are crossed?

| | | |
|-----|------|------|
| | T | t |
| T | TT | Tt |
| t | Tt | tt |

A *Punnett square* is a chart used to determine the possible combinations of characteristics among offspring.

3 tall-stemmed
+ 1 short-stemmed
4 total

Thus, the probability is $\frac{3}{4}$.

In a certain plant, red flowers, R , are dominant to white flowers, r . If a white-flowered plant, rr is crossed with a red-flowered plant, Rr , find the probability of each of the following.

- white-flowered plant $\frac{1}{2}$
 - red-flowered plant $\frac{1}{2}$
- In a certain plant, tall, T , is dominant to short, t , and green pods, G , are dominant to yellow pods, g . Plants with the genetic formulas $TtGg$ and $TtGg$ are crossed. Find the probability of each of the following.
- tall plant with green pods $\frac{3}{4}$
 - tall plant with yellow pods $\frac{1}{4}$