

11-6

Analyzing Data

Content Standards

S.MD.6 Use probabilities to make fair decisions ...

S.MD.7 Analyze decisions and strategies using probability concepts ...

Objectives To calculate measures of central tendency
To draw and interpret box-and-whisker plots



This is not your average average!



Getting Ready!

You travel east and west at the speeds shown. Is your average speed for the round trip 55 mi/h, less than 55 mi/h, or greater than 55 mi/h? Answer without calculating, and explain your reasoning.



Lesson Vocabulary

- measure of central tendency
- mean
- median
- mode
- bimodal
- outlier
- range of a set of data
- quartile
- interquartile range
- box-and-whisker plot
- percentile

People often refer to the mean as the *average*. The mean is only one of the measures considered the average, a measure of the center of a set of data.

Essential Understanding You can describe and compare sets of data using various statistical measures, depending on what characteristics you want to study.

Statistics is the study, analysis, and interpretation of data. One way to analyze data is by finding a *measure of central tendency*. A **measure of central tendency** indicates the “middle” of the data set. The *mean*, *median*, and *mode* are the most common measures of central tendency.



Key Concepts Measures of Central Tendency

Measure	Definition	Example, using 1, 2, 3, 3, 4, 5, 5, 9
Mean	$\frac{\text{sum of the data values}}{\text{number of data values}}$	$\frac{1 + 2 + 3 + 3 + 4 + 5 + 5 + 9}{8} = 4$
Median	for a data set listed in order: the middle value for an odd number of data values; the mean of the two middle values for an even number of data values	For 1, 2, 3, 3, 4, 5, 5, 9, the middle two values are 3 and 4. The median is their mean $\frac{3 + 4}{2} = 3.5$.
Mode	the most frequently occurring value(s)	Two modes: In 1, 2, 3, 3, 4, 5, 5, 9, both 3 and 5 occur twice.

Dynamic Activity
Box-and-Whisker Plots

A **bimodal** data set has two modes. If a data set has more than two modes, then the modes are probably not statistically useful. If no value occurs more frequently than any other, then there is no mode.

Problem 1 Finding Measures of Central Tendency

Career The frequency table shows the number of job offers received by each student within two months of graduating with a mathematics degree from a small college. What are the mean, median, and mode for the job offers per student?

Job Offers	0	1	2	3	4
Students	2	2	4	5	2

Think

How do you find the total number of job offers?

Add the products of each number of job offers and the number of students with that many job offers.

$$\begin{aligned} \text{Mean: } \bar{x} &= \frac{2(0) + 2(1) + 4(2) + 5(3) + 2(4)}{15} \\ &= \frac{33}{15} = 2.2 \end{aligned}$$

The symbol \bar{x} , read "x bar," represents the mean.

The mean is 2.2.

Median: 0, 0, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4
The median is 2.

List each value the number of times it occurs. Arrange them in order. Find the middle value.

Mode: Five students received 3 job offers each.
The mode is 3.

The mode is the number of job offers received by most students.

Got It? 1. The frequency table shows the number of trees in the yard of each house on one street. What are the mean, median, and mode for the trees per yard?

Trees	3	4	5	6	7	8
Yards	1	5	7	4	1	2

An **outlier** is a value that is substantially different from the rest of the data in a set. If the data is in one variable, outliers can occur at the "ends." They can be misleading because they can affect measures of central tendency.

Problem 2 Identifying an Outlier

Multiple Choice Which is an outlier for this data set: 56 65 73 59 98 65 59?

(A) 42

(B) 65

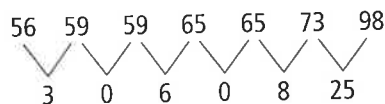
(C) 98

(D) 59

Think

What should you do first?

Put the numbers in order.



Order the data.

Find differences between adjacent values.

98 appears to be substantially different, so 98 is an outlier. The correct answer is C.

Got It? 2. Suppose the values in Problem 2 are the data for the situations below. Would you discard the outlier? Explain.

- water temperature of a lake at seven locations
- the number of customers in a restaurant each night in one week

The **range of a set of data** is the difference between the greatest and least values. If you order data from least value to greatest value, the median divides the data into two parts. The median of each part divides the data further and you have four parts in all. The values separating the four parts are **quartiles**. The **interquartile range** is the difference between the third and first quartiles.

Problem 3 Comparing Data Sets

Temperature The table shows average monthly water temperatures for four locations on the Gulf of Mexico. How can you compare the 12 water temperatures from St. Petersburg with the 12 water temperatures from Key West?

Gulf of Mexico Eastern Coast Water Temperatures (°F)

Location	J	F	M	A	M	J	J	A	S	O	N	D
St. Petersburg, Florida	62	64	68	74	80	84	86	86	84	78	70	64
Key West, Florida	69	70	75	78	82	85	87	87	86	82	76	72
Dauphin Island, Alabama	51	53	60	70	75	82	84	84	80	72	62	56
Grand Isle, Louisiana	61	61	64	70	77	83	85	85	83	77	70	65

SOURCE: National Oceanographic Data Center

Know

Water temperatures near the two cities

Need

The means, medians, modes, ranges, and interquartile ranges

Plan

Order the data. Find the means, medians, modes, minimums, maximums, quartiles, range, and interquartile range.

St. Petersburg:

$$\bar{x} = \frac{62 + 64 + 64 + 68 + 70 + 74 + 78 + 80 + 84 + 84 + 86 + 86}{12}$$

$$= \frac{900}{12} = 75 \text{ (mean water temperature)}$$

Modes: 64, 84, and 86

Min.: 62; Max.: 86; Range: $86 - 62 = 24$

$$\text{Median } (Q_2) = 76$$

62 64 (64 68) 70 (74 78) 80 (84 84) 86 86

$$\text{Median of lower part } (Q_1) = 66$$

$$\text{Median of upper part } (Q_3) = 84$$

Interquartile range:

$$Q_3 - Q_1 = 84 - 66 = 18$$

The range and the interquartile range show the temperatures varying less at Key West than at St. Petersburg. Also, the temperatures at Key West are generally higher.

Key West:

$$\bar{x} = \frac{69 + 70 + 72 + 75 + 76 + 78 + 82 + 82 + 85 + 86 + 87 + 87}{12}$$

$$= \frac{949}{12} \approx 79.1 \text{ (mean water temperature)}$$

Modes: 82 and 87

Min.: 69; Max.: 87; Range: $87 - 69 = 18$

$$\text{Median } (Q_2) = 80$$

69 70 (72 75) 76 (78 82) 82 (85 86) 87 87

$$\text{Median of lower part } (Q_1) = 73.5$$

$$\text{Median of upper part } (Q_3) = 85.5$$

Interquartile range:

$$Q_3 - Q_1 = 85.5 - 73.5 = 12$$

Think

What location has a greater range in water temperature?

The range of water temperatures at St. Petersburg is 6°F greater than the range at Key West.

- Got It?** 3. How can you compare the 12 water temperatures in Problem 3 from Dauphin Island with the 12 water temperatures from Grand Isle?

A *box-and-whisker plot* uses minimum and maximum values, the median, and the first and third quartiles to display the spread, or variability, in a data set.

Take note

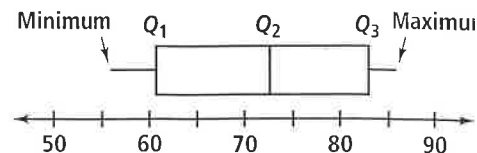
Key Concept Box-and-Whisker Plot

Definition

A **box-and-whisker plot** is a way to display data that uses

- quartiles to bound the center box and
- the minimum and maximum values to form the whiskers.

Graph



Problem 4 Using a Box-and-Whisker Plot

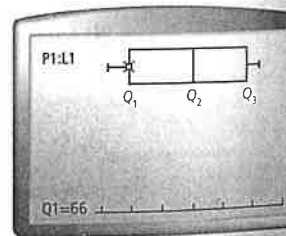
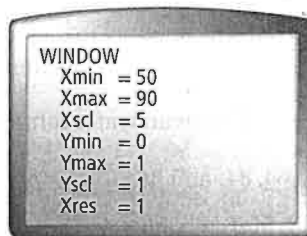
How can you use a graphing calculator box-and-whisker plot to find quartiles for the water temperature data of St. Petersburg from Problem 3?

Think

What about the appearance of a box-and-whisker plot might suggest an outlier?

If a "whisker" is much longer than the box, its endpoint may be an outlier.

- Step 1** For St. Petersburg, use **STAT EDIT** to enter the temperature data in **L1**.
- Step 2** In **STAT PLOT**, select a box-and-whisker plot. Enter **L1** for the St. Petersburg data. Enter the window values. Draw the box-and-whisker plot.



- Step 3** Use **TRACE** to find the quartiles: $Q_1 = 66$, $Q_2 = 76$, and $Q_3 = 84$.

- Got It?** 4. a. How can you use graphing calculator box-and-whisker plots to find water temperature quartiles for other Gulf Coast sites in Problem 3?
- b. **Reasoning** Is a box-and-whisker plot a useful graphical display for data with an outlier? Explain.

A **percentile** is a number from 0 to 100 that you can associate with a value x from a data set. It shows the percent of the data that are less than or equal to x . If x is at the 63rd percentile, then 63% of the data are less than or equal to x .

Problem 5 Finding Percentiles

Testing Here is an ordered list of midterm test scores for a Spanish class. What value is at the 65th percentile?

41	54	61	65	67	73	74
77	77	77	79	80	82	88
89	93	97	98	98	100	

Of the 20 values, 65% fall at or below the value at the 65th percentile.

$$20 \cdot 65\% = 20 \cdot 0.65 = 13$$

13 values fall at or below 82, the value at the 65th percentile.

- Got It?** 5. What are the values at each percentile for the data in Problem 5?
 a. 55th percentile b. 95th percentile

Plan
 What should you do first to find percentiles?
 Put the data in order.

Lesson Check

Do you know HOW?

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

- 16 19 21 18 18 54 20 22 23 17
- 90 100 110 40 98 102 112 90 92
- Using your results from Exercises 1 and 2, explain which measure of central tendency is most affected by an outlier.
- Find the values at the 40th and 80th percentiles for the values below.

58 53 35 60 58 42 57 60 43 44 51 49 58

Do you UNDERSTAND?



5. **Vocabulary** Which measure of central tendency would best represent the values below? Explain your reasoning.

4 1 5 5 6 8 9 5 5 3 2 7 5 5 1

6. **Error Analysis** A student found the median of the data set below. Explain the student's error. What is the median?

Score	80	85	90	95
Frequency	6	4	10	1

~~Median: $\frac{85 + 90}{2} = \frac{175}{2} = 87.5$~~

Practice and Problem-Solving Exercises



Practice

Find the mean, median, and mode of each set of values.

See Problem 1.

7. Time spent on Internet per day (in minutes): 75 68 43 120 65 180 95 225 140

8.

Age (years)	13	14	15	16	17	18	19
Frequency	7	12	18	9	5	4	2

Identify the outlier of each set of values.

9. 3.4 4.5 2.3 5.9 9.8 3.3 2.1 3.0 2.9

10. 17 21 19 10 15 19 14 0 11 16

11. **Weather** The table shows average monthly temperatures of two cities. How can you compare the temperatures?

See Problem 2.

See Problems 3 and 4.

	J	F	M	A	M	J	J	A	S	O	N	D
Jacksonville, Florida	52.4	55.2	61.1	67.0	73.4	79.1	81.6	81.2	78.1	69.8	61.9	55.1
Austin, Texas	48.8	52.8	61.5	69.9	75.6	81.3	84.5	84.8	80.2	71.1	60.9	51.6

Make a box-and-whisker plot for each set of values.

12. 12 11 15 12 19 20 19 14 18 15 16

13. 120 145 133 105 117 150 130 136 128

Find the values at the 30th and 90th percentiles for each data set.

14. 6283 5700 6381 6274 5700 5896 5972 6075 5993 5581

15. 7 12 3 14 17 20 5 3 17 4 13 2 15 9 15 18 16 9 1 6

See Problem 5.

B Apply

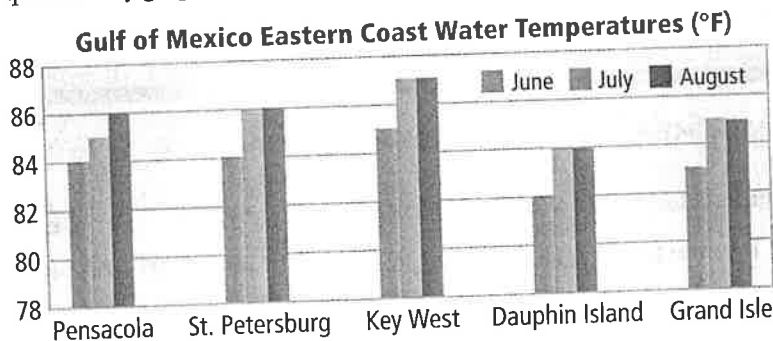
Identify the outlier in each data set. Then find the mean, median, and mode of the data set when the outlier is included and when it is not.

16. 947 757 103 619 661 582 626 900 869 728 1001 596 515

17. 87 104 381 215 174 199 233 186 142 228 9 53 117 129

18. 49 57.5 58 49.2 62 22.2 67 52.1 77 99.9 80 51.7 64

© 19. **Think About a Plan** Use the water temperature data for the eastern coast of the Gulf of Mexico during the summer months, as shown in the graph below. Find the quartiles by graphing a box-and-whisker plot of the data.



- What information can you get from the graph?
- How can you use that information to make a box-and-whisker plot?
- How can you find the quartiles using your box-and-whisker plot?

- STEM** 20. **Meteorology** On May 3, 1999, 59 tornadoes hit Oklahoma in the largest tornado outbreak ever recorded in the state. Sixteen of these were classified as strong (F2 or F3) or violent (F4 or F5).
- Make a box-and-whisker plot of the data for length of path.
 - Identify the outliers. Remove them from the data set and make a revised box-and-whisker plot.
- © 21. **Writing** How does the removal of the outliers affect the box-and-whisker plot? How does it affect the median of the data set?

For Exercises 21–23, use the set of values below.

1 1 1 1 1 2 3 5 8 13 21 34 55 89 89 89 89 89 89

21. At what percentile is 1? 22. At what percentile is 34?
- © 23. **Error Analysis** A student claims that 89 is at the 70th percentile. Explain the student's error.
24. **Advertising** An electronics store placed an ad in the newspaper showing flat-screen TVs for sale. The ad says "Our flat-screen TVs average \$695." The prices of the flat-screen TVs are \$1200, \$999, \$1499, \$895, \$695, \$1100, \$1300, and \$695.
- Find the mean, median, and mode of the prices.
 - Which measure is the store using in its ad? Why did they choose it?
 - As a consumer, which measure would you want to see advertised? Explain your reasoning.
- © 25. **Reasoning** Which measure better represents a data set with several outliers—the mean or the median? Justify your answer.

26. The table displays the frequency of scores for one Calculus class on the Advanced Placement Calculus exam. The mean of the exam scores is 3.5.

Score	1	2	3	4	5
Frequency	1	3	f	12	3

- What is the value of f in the table?
 - What is the mode of all of the exam scores?
 - What is the median of all of the exam scores?
27. **Grades** Some teachers use a *weighted mean* to calculate grades. Each score is assigned a weight based on its importance. To find a weighted mean, multiply each score by its weight and add the results. For example, a student's final chemistry grade is based on four sources: 30% from lab reports, 10% from quizzes, 25% from the midterm exam, and 35% from the final exam. What is the student's weighted mean given the scores shown?

Lab Reports	82
Quizzes	95
Midterm Exam	76
Final Exam	88

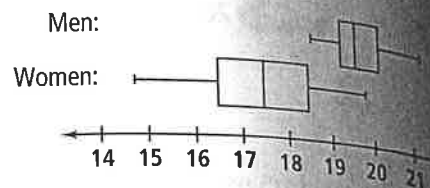
Major Tornadoes in Oklahoma, May 3, 1999

Length of Path (miles)	Intensity
6	F3
9	F3
4	F2
37	F5
7	F2
12	F3
8	F2
7	F2
15	F4
39	F4
1	F2
22	F3
15	F3
8	F2
13	F3
2	F2

Source: National Oceanic & Atmospheric Administration

- © **Challenge** 28. **Reasoning** What effect will adding 10 to every value in a data set have on the mean, median, mode, range, and box-and-whisker plot? What will be the effect if you multiply each value by 10?

- 29. Track and Field** The box-and-whisker plots show the 36 best qualifying distances for the shot put events for men and women during the 2004 Olympics. Compare box-and-whisker plots. Describe any conclusions you can draw about Olympic-level male and female shot-putters.



Standardized Test Prep

SAT/ACT

- 30.** Use a calculator to solve $2x^2 - 7x - 5 = 0$. Round answers to the nearest hundredth.
 (A) $-1.56, -4.44$ (B) $-5.44, 1.56$ (C) $-0.61, 4.11$ (D) $-5.56, -1.44$
- 31.** Which function generates the table of values below?

x	-2	-1	0	1	2
y	$\frac{27}{8}$	$\frac{9}{2}$	6	8	$\frac{32}{3}$

- (F) $27\left(\frac{2}{3}\right)^x$ (G) $6\left(\frac{4}{3}\right)^x$ (H) $\left(\frac{8}{3}\right)^x$ (I) $6\left(\frac{3}{4}\right)^x$
- 32.** A homeroom class consists of 6 boys whose last name begins with S, 8 boys whose last name begins with T, 4 girls whose last name begins with S, and 11 girls whose last name begins with T. A student is chosen at random from the class. What is the probability that the student is a girl or has a last name that begins with S?
- (A) $\frac{18}{29}$ (B) $\frac{21}{29}$ (C) $\frac{23}{29}$ (D) $\frac{25}{29}$
- 33.** In a library, the probability that a book is a hardback, given that it is illustrated, is 0.40. The probability that a book is hardback *and* illustrated is 0.20. Find the probability that a book is illustrated.

Short Response

Mixed Review

Of all the respondents to a survey, 59% are girls. Of the girls, 61% read horror stories. Of the boys, 49% read horror stories. Find each probability. ◀ See Lesson 11-4.

- 34.** $P(\text{boy and reads horror stories})$ **35.** $P(\text{reads horror stories})$

Determine whether each sequence is arithmetic. If it is, identify the common difference. ◀ See Lesson 9-2.

- 36.** 16, 7, -2, ... **37.** 34, 51, 68, ... **38.** 2, 2.2, 2.22, ... **39.** 1, 1, 1, ...

Get Ready! To prepare for Lesson 11-7, do Exercises 40-43.

Find all real square roots of each number. ◀ See Lesson 6-1.

- 40.** 256 **41.** 0.0081 **42.** $\frac{121}{16}$ **43.** $\frac{361}{25}$