

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

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.

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

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

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?

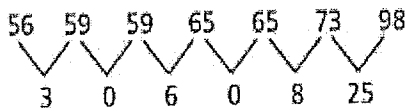
Trees	3	4	5	6	7	8
Yards	1	5	7	4	1	2

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



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

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:

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

Key West:

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

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

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:

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

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.

- 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

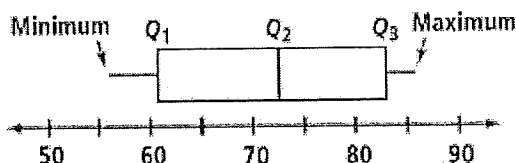
Key Concept Box-and-Whisker Plot

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



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

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

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?
- 55th percentile
 - 95th percentile

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

Identify the outlier of each set of values.

See Problem 2.

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 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.

See Problem 5.

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

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

