LESSON 18 Outliers

GOAL Determine whether a set of data contains outliers.

In a set of data, values that are widely separated from the rest of the data are called outliers. **Outliers** are not just the greatest and least values, but values that are very different from the rest of the data.

One way to determine whether a set of data contains outliers is to use quartiles. If a data value is more than 1.5 interquartile ranges below the lower quartile or more than 1.5 interquartile ranges above the upper quartile, then it is an outlier.

EXAMPLE 1 Finding Outliers Using Quartiles

In 2001, a Major League Baseball player set a record by hitting 73 home runs in one season. The list below shows the number of home runs the player hit each year from 1986 to 2001. Use quartiles to determine whether 73 is an outlier. If so, make a modified box-and-whisker plot of the data.

16, 25, 24, 19, 33, 25, 34, 46, 37, 33, 42, 40, 37, 34, 49, 73

SOLUTION

Step 1 List the data in increasing order. Then find the lower and upper quartiles.

Median: $\frac{34 + 34}{2} = 34$

16 19 24 25 25 33 33 34 34 37 37 40 42 46 49 73

Lower quartile: $\frac{25 + 25}{2} = 25$ **Upper quartile:** $\frac{40 + 42}{2} = 41$

Step 2 Find the interquartile range.

IQR = Upper quartile − Lower quartile = 41 − 25 = 16

Step 3 Subtract 1.5 interquartile ranges from the lower quartile to find the lower outlier boundary.

Lower quartile − 1.5(IQR) = 25 − 1.5(16) = 25 − 24 = 1

Because there are no values less than 1, there are no outliers on the low end.

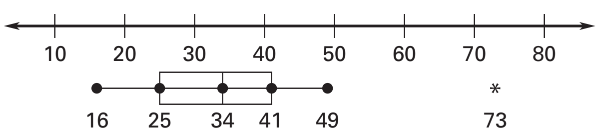
Step 4 Add 1.5 interquartile ranges to the upper quartile to find the upper outlier boundary.

Upper quartile + 1.5(IQR) = 41 + 1.5(16) = 41 + 24 = 65

Because 73 is greater than 65, 73 is considered an outlier.

Step 5 Make a modified box-and-whisker plot of the data.

The box-and-whisker plot is *modified*. The upper whisker stops at 49, the greatest value that is not an outlier. Then, an asterisk is placed at 73, the outlier. The lower extreme, median, and quartiles are not recalculated.



CHECK **Example 1**

In Exercises 1 and 2, use the stem-and-leaf plot.

1. Find the interquartile range of the data.

2. Use quartiles to identify any outliers. If there are any outliers, make a modified box-and-whisker plot of the data.

0	1 9
1	8 8 9 9
2	0 1 2 3 3 6 7 9
3	5 8 9

Key: 1|8 = 18

Another way to determine whether a set of data contains outliers is to use the mean and standard deviation. If a data value is more than 3 standard deviations above or below the mean, then it is an outlier.

EXAMPLE 2 Finding Outliers Using Mean and Standard Deviation

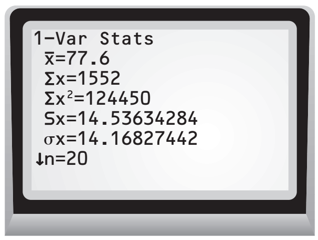
The list below shows 20 scores on a math test. Use the mean and standard deviation to identify any outliers.

61, 92, 83, 63, 93, 69, 72, 74, 75, 32, 80, 83, 86, 88, 75, 78, 88, 89, 97, 74

SOLUTION

Step 1 Use a calculator to find the mean and standard deviation of the data.

$\bar{x} = 77.6 \qquad \sigma \approx 14.2$



Step 2 Subtract 3 standard deviations from the mean to find the lower outlier boundary.

$$\bar{x} - 3\sigma \approx 77.6 - 3(14.2)$$
$$= 77.6 - 42.6 = 35$$

Because 32 is less than 35, 32 is considered an outlier.

Step 3 Add 3 standard deviations to the mean to find the upper outlier boundary.

$$\bar{x} + 3\sigma \approx 77.6 + 3(14.2)$$
$$= 77.6 + 42.6 = 120.2$$

Because there are no scores greater than 120.2, there are no outliers on the high end.

The test score 32 is an outlier.

CHECK **Example 2**

Use the mean and standard deviation to identify any outliers in the set of data.

3. Prices (in dollars) of bicycles listed in newspaper advertisements:

300, 500, 350, 450, 225, 350, 300, 575, 450, 325, 425, 250, 350, 75, 200, 350, 650, 400

4. Salaries (in millions of dollars) of members of a professional sports team:

13.2, 5.7, 4.5, 3.5, 3.4, 2.5, 2.0, 1.5, 1.3, 1.2, 1.1, 0.9, 0.7, 0.7, 0.5, 0.4, 0.4, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2

EXERCISES

Use quartiles to identify any outliers in the set of data.

1. 37, 59, 60, 60, 61, 62, 64, 66, 67, 68, 68, 70, 71, 72, 72, 72, 75, 90
2. −21, −15, −12, −11, −11, −8, −5, −2, 0, 1, 1, 4, 7, 10, 12, 13, 13
3. 250, 230, 270, 360, 250, 10, 230, 260, 250, 230, 330, 340, 220, 350, 220
4. −5, −3, 9, 5, 10, 57, 27, −7, −15, 14, 18, 25, 28, 3, 0, 23, 21, − 1

In Exercises 5–7, use quartiles to identify any outliers in the set of data. If there are any outliers, make a modified box-and-whisker plot of the data.

5. The stem-and-leaf plot below shows the winning margin in the first 36 Super Bowls in the National Football League.

0	1 3 3 4 4 4 5 7 7 7 9
1	0 0 0 2 3 4 5 6 7 7 7 7 8 9 9
2	1 2 3 5 7 9
3	2 5 6
4	5

Key: $1|9 = 19$

6. The list below shows the ages of the winners of the Academy Award for Actress in a Supporting Role for 1982 through 2001.

   11, 24, 28, 28, 31, 33, 33, 33, 34, 35, 38, 39, 41, 44, 44, 45, 46, 56, 64, 77

7. The list below shows the heights (in inches) of singers in the alto section of a chorus.

   65, 62, 68, 67, 67, 63, 67, 66, 63, 72, 62, 61, 66, 64, 60, 61, 66, 66, 66, 62, 70, 65, 64, 63, 65, 69, 61, 66, 65, 61, 63, 64, 67, 66, 68

8. A set of data has the following characteristics: The lower extreme is 25, the lower quartile is 38, the median is 45, the upper quartile is 53, and the upper extreme is 70. Can you determine whether there are any outliers in the set of data? Explain your reasoning.

In Exercises 9–12, use the following information.

The average acorn size (in cubic centimeters) for 28 species of oak trees in the Atlantic region and 11 species of oak trees in California are shown below.

Atlantic: 1.4, 3.4, 9.1, 1.6, 10.5, 2.5, 0.9, 6.8, 1.8, 0.3, 0.9, 0.8, 2.0, 1.1, 0.6, 1.8, 4.8, 1.1, 3.6, 1.1, 1.1, 3.6, 8.1, 3.6, 1.8, 0.4, 1.1, 1.2

California: 4.1, 1.6, 2.0, 5.5, 5.9, 2.6, 6.0, 1.0, 17.1, 0.4, 7.1

9. Use quartiles to identify any outliers in the Atlantic region data.
10. Use quartiles to identify any outliers in the California data.
11. Make a modified double box-and-whisker plot of the data.
12. Use your modified double box-and-whisker plot to make a conclusion about the data.

In Exercises 13 and 14, use a calculator to find the mean and standard deviation. Then use them to identify any outliers in the set of data.

13. 21, 35, 23, 38, 62, 36, 37, 21, 30, 32, 34, 29, 26, 23, 36, 27, 29, 31, 9

14. 3.025, 3.125, 3.015, 3.000, 3.030, 4.005, 3.125, 3.100, 3.025, 2.995, 3.015, 3.115

In Exercises 15–17, use the double stem-and-leaf plot of the ages of a sample of listeners of two local radio stations.

Station A		Station B
2 4 4 6 8 8 9	1	8 9
0 1 2 3 3 4 5 7 7 8 8 9 9 9	2	0 1 1 3 4 8 9
1 1 1 2 4 5 6 7	3	0 2 4 4 5 7 7 9 9
0 1 2 3 7 7	4	0 2 3 4 5 5 6 6 7 8 9
3 5 5	5	0 2 3 4 5 5 5 5 6 7
1 2	6	3 3 5 6 7 7 9
	7	1 3 6 6 7 8
0	8	0 2 4

Key: 9|1|8 = 19, 18

15. Use the mean and standard deviation to identify any outliers for Station A.

16. Use the mean and standard deviation to identify any outliers for Station B.

17. Make a conclusion about the data.

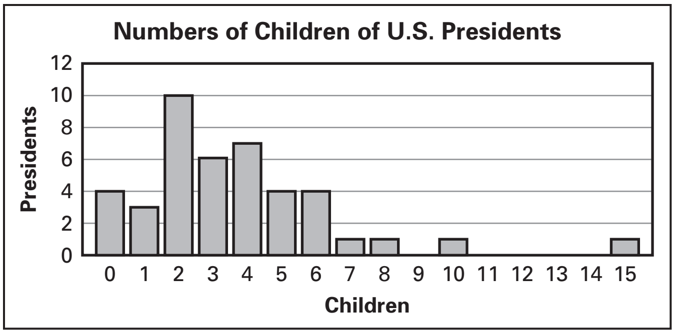
In Exercises 18–21, use the bar graph below, which shows the number of children of each U.S. president.

18. Without doing any computations, name a data value that might be considered an outlier.

19. Use quartiles to identify any outliers in the set of data.

20. Use the mean and standard deviation to identify any outliers in the set of data.

21. Compare your answers for Exercises 19 and 20.



In Exercises 22–24, use the following set of data. 100 is an outlier.

40, 40, 45, 45, 47, 50, 50, 50, 55, 57, 60, 60, 62, 62, 65, 70, 100

22. Find the mean, median, and mode of the set of data. Then omit the outlier from the set of data and find the mean, median, and mode. What effect does the outlier have on these measures of central tendency?

23. Find the range, interquartile range, and standard deviation of the set of data. Then omit the outlier from the set of data and find the range, interquartile range, and standard deviation. What effect does the outlier have on these measures of dispersion?

24. Make a box-and-whisker plot of the data. Then omit the outlier from the set of data and make a new box-and-whisker plot. Compare the graphs.