

Measures of Dispersion

GOAL Find the range, interquartile range, variance, and standard deviation of sets of data.

The lists below give the average fuel economy (in miles per gallon of gasoline) used in city driving for 2002 models of selected subcompact and midsize cars.

Subcompact: 27, 11, 20, 19, 25, 19, 35, 17, 20, 28, 19, 11, 21, 26, 19, 28, 23

Midsize: 19, 17, 20, 18, 21, 21, 21, 25, 20, 21, 26, 23, 18, 24

The mean for subcompact cars is about 21.6 mi/gal and the mean for midsize cars is 21 mi/gal. The median for subcompact cars is 20 mi/gal and the median for midsize cars is 21 mi/gal. While these measures of central tendency are approximately equal, this is not the complete picture. Additional statistics, called **measures of dispersion**, are used to describe the spread of a set of data. Two measures of dispersion are *range* and *interquartile range*.

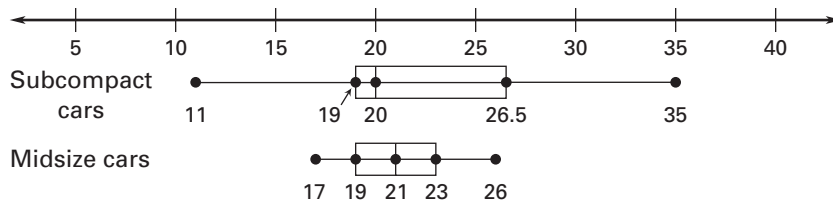
The **range** of a set of data is the difference between the greatest value and the least value. The **interquartile range (IQR)** of a set of data is a measure of the spread of the middle 50% of the data. The IQR is the difference between the upper and lower quartiles.

EXAMPLE 1 Finding Range and Interquartile Range

Use the data listed above to make a double box-and-whisker plot of the average fuel economies for subcompact and midsize cars. Then find the range and interquartile range of each set of data. Use your results to make a conclusion about the data.

SOLUTION

First, make a double box-and-whisker plot of the data.



Subcompact: range = greatest value – least value = $35 - 11 = 24$ mi/gal
 IQR = upper quartile – lower quartile = $26.5 - 19 = 7.5$ mi/gal

Midsize: range = greatest value – least value = $26 - 17 = 9$ mi/gal
 IQR = upper quartile – lower quartile = $23 - 19 = 4$ mi/gal

The range for the subcompact cars is much greater than for the midsize cars, which indicates a greater spread of the data. With an IQR of 7.5 mi/gal, the middle 50% of the subcompact cars is more spread out than the middle 50% of the midsize cars with an IQR of 4 mi/gal.

**CHECK Example 1**

The average fuel economy (in miles per gallon) for selected cars used in highway driving is given. Find the range and interquartile range of the data.

1. Subcompact: 33, 16, 29, 30, 33, 27, 40, 24, 27, 35, 30, 16, 29, 36, 26, 33, 30
2. Midsize: 29, 24, 29, 27, 32, 32, 30, 32, 26, 28, 32, 29, 28, 33

While the range is based on the extremes of the data and the interquartile range is based on quartiles, there are two measures of dispersion that are based on the mean of the data: *variance* and *standard deviation*. The **standard deviation** of a set of data describes the typical difference, or deviation, between the mean and a data value. The standard deviation is the positive square root of the **variance**.

Variance and Standard Deviation

The variance σ^2 (read as “sigma squared”) of a set of data x_1, x_2, \dots, x_n is:

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

The standard deviation σ (read as “sigma”) of a set of data x_1, x_2, \dots, x_n is:

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

EXAMPLE 2 Finding Variance and Standard Deviation

The ages (in weeks) at which 6 infants learned to crawl are shown below. Find the variance and standard deviation of the data.

23, 37, 33, 39, 29, 31

SOLUTION

Step 1 Find the mean. $\bar{x} = \frac{23 + 37 + 33 + 39 + 29 + 31}{6} = 32$ weeks

Step 2 Find the variance.

$$\begin{aligned} \sigma^2 &= \frac{(23 - 32)^2 + (37 - 32)^2 + (33 - 32)^2 + (39 - 32)^2 + (29 - 32)^2 + (31 - 32)^2}{6} \\ &= \frac{(-9)^2 + 5^2 + 1^2 + 7^2 + (-3)^2 + (-1)^2}{6} \\ &= \frac{81 + 25 + 1 + 49 + 9 + 1}{6} = \frac{166}{6} = 27.\bar{6} \end{aligned}$$

Step 3 Find the standard deviation. $\sigma = \sqrt{\text{variance}} = \sqrt{27.\bar{6}} \approx 5.26$ weeks

You can see that calculating the standard deviation of even a small set of data requires a lot of computation. For larger sets of data, calculators or computers can be used to find the standard deviation.

EXAMPLE 3 Using a Calculator to Find Standard Deviation

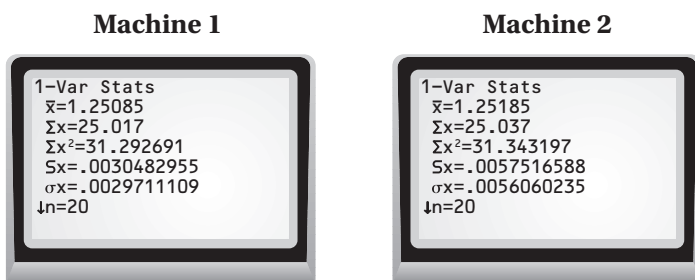
Two machines package nuts in 1.25 pound cans. To test the packaging process, a quality control worker weighs a random sample of 20 cans from each machine. Use a calculator to find the mean and standard deviation for each machine. Use your results to make a conclusion about the data.

Can Weights (lb)	
Machine 1	1.250, 1.252, 1.248, 1.252, 1.251, 1.256, 1.253, 1.248, 1.246, 1.257, 1.250, 1.251, 1.250, 1.255, 1.251, 1.253, 1.249, 1.249, 1.245, 1.251
Machine 2	1.254, 1.245, 1.253, 1.250, 1.261, 1.253, 1.254, 1.244, 1.250, 1.252, 1.260, 1.257, 1.248, 1.242, 1.253, 1.243, 1.260, 1.248, 1.251, 1.259

SOLUTION

Step 1 Use the *Stat Edit* feature to enter the data for each machine. Use L1 for Machine 1 and L2 for Machine 2.

Step 1 Use the *Stat Calc* menu to calculate the 1-variable statistics for each list. The mean is \bar{x} and the standard deviation is σ_x . By scrolling, you can also find the least value (minX), lower quartile (Q1), median (Med), upper quartile (Q3), and greatest value (maxX).



The means for the two machines (1.25085 lb and 1.25185 lb) are about equal. However, the standard deviation for Machine 1 (about 0.00297 lb) is less than that for Machine 2 (about 0.00561 lb), so Machine 1 has less variability.

CHECK Examples 2 and 3

Find the variance and standard deviation of the data. Round your answers to the nearest hundredth.

3. 78, 83, 85, 89, 92, 95
4. 30, 10, 15, 25, 10, 30, 20, 5, 25, 10

Use a calculator to find the mean and standard deviation of the data. Round your answers to the nearest hundredth.

5. 13, 17, 34, 16, 12, 18, 20, 21, 23
6. 12.4, 13.5, 12.1, 12.0, 11.9, 13.1, 12.7

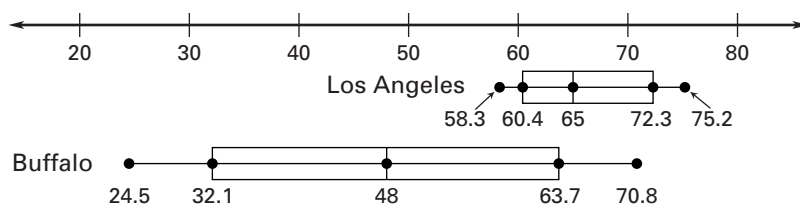
EXERCISES

Find the range and interquartile range of the data.

- Waiting times (in minutes) of 8 students waiting for a school bus:
1, 4, 9, 2, 8, 5, 5, 12
- Number of players on each of 6 teams in a soccer league:
15, 20, 13, 18, 15, 17
- Weekend gross (in millions of dollars) of 10 movies:
16.5, 14.0, 11.1, 10.3, 7.9, 7.0, 5.0, 4.5, 2.9, 2.7
- Money (in dollars) spent by 9 consecutive shoppers at a grocery store:
6.61, 8.04, 11.64, 14.35, 20.58, 23.85, 30.54, 33.80, 43.97

In Exercises 5 and 6, use the following information.

The double box-and-whisker plot shows the average monthly temperatures (in degrees Fahrenheit) for Los Angeles, California, and Buffalo, New York.

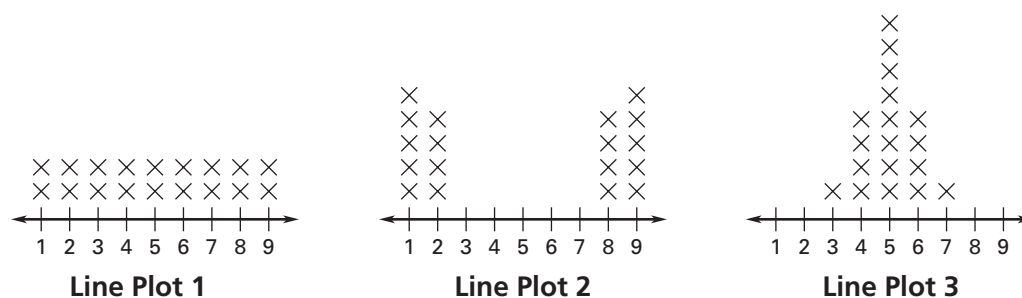


- Find the range and interquartile range for each city.
- Use your results from Exercise 5 to make a conclusion about the data.

Find the variance and standard deviation of the data. Round your answers to the nearest hundredth.

- 125, 240, 175, 200
- 10, 20, 13, 17, 15
- 55, 58, 57, 54, 54, 59, 70, 65
- 85, 76, 93, 82, 67, 59, 82, 83, 87, 86

In Exercises 11 and 12, use the line plots below.



- Find the mean and standard deviation of each set of data. Round your answers to the nearest hundredth.
- How does the standard deviation of each set of data help to describe the spread of the data?

In Exercises 13–15, use the following information.

The table below shows the average precipitation (in inches) for each month in Seattle, Washington, and Galveston, Texas.

	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Seattle	5.4	4.0	3.8	2.5	1.8	1.6	0.9	1.2	1.9	3.3	5.7	6.0
Galveston	3.3	2.3	2.2	2.4	3.6	4.4	4.0	4.5	5.9	2.8	3.4	3.5

13. Find the mean and standard deviation of the Seattle precipitation data.
14. Find the mean and standard deviation of the Galveston precipitation data.
15. Compare the results for each city. Make a conclusion about the data.

In Exercises 16–19, use the following information.

The number of receptions for the top 25 receivers in the National Football Conference (NFC) during the 2002 regular season are listed below.

143, 112, 100, 97, 94, 92, 91, 89,
 89, 81, 80, 79, 79, 78, 70, 69, 69,
 68, 67, 63, 63, 61, 61, 61

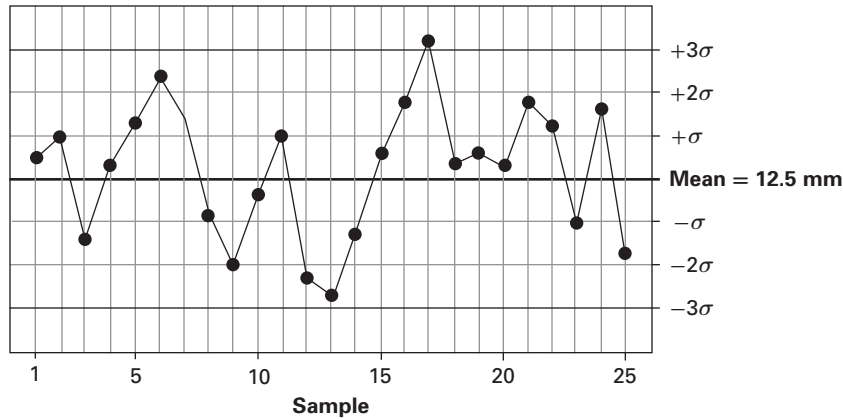
16. Make a line plot of the data.
17. Find the mean of the data. Mark the mean on your line plot.
18. Use a calculator to find the standard deviation of the data. Add one standard deviation to the mean. Subtract one standard deviation from the mean. Mark these values on your line plot.
19. What percent of the data is within one standard deviation of the mean?
20. The tables below show the percent of the population under 15 years of age in various countries in 2000. Find the range, interquartile range, mean, and standard deviation of each set of data. Compare your results for North America and Europe to your results for South America and the Caribbean. Make a conclusion about the data.

North America and Europe	
Belgium	17.5%
Canada	19.2%
Czech Republic	16.5%
France	18.8%
Germany	15.7%
Hungary	16.9%
Netherlands	18.4%
Poland	19.0%
Portugal	17.1%
Spain	14.8%
United States	21.2%

South America and the Caribbean	
Argentina	26.8%
Brazil	29.1%
Chile	27.6%
Colombia	32.2%
Cuba	21.4%
Ecuador	36.2%
Mexico	33.8%
Peru	34.8%
Venezuela	32.7%

In Exercises 21–24, use the following information.

A *control chart* is a tool used to monitor the consistency of a process, such as manufacturing ball bearings. The horizontal line in the center of a control chart represents the mean. In the control chart shown below, the mean diameter of a ball bearing is 12.5 mm. Data from successive samples are plotted above or below the mean depending on how many standard deviations the data are from the mean.



21. Suppose the standard deviation for the above process is 0.04 mm. What is the approximate diameter of sample 2? of sample 9?
22. How many of the samples lie no more than one standard deviation from the mean? How many lie no more than two standard deviations from the mean?
23. A sample is said to send a signal when it meets established criteria that indicate that the process is out of control.
 - a. A sample that is more than three standard deviations from the mean sends a signal. Which sample sends this signal?
 - b. A sample that is the second in a row on the same side of the mean line and more than two standard deviations from the mean sends a signal. Which sample sends this signal?
 - c. A sample that is the eighth in a row on the same side of the mean line sends a signal. Which sample sends this signal?
24. Construct a control chart for the ball bearings in which there are 10 samples, every sample is no more than two standard deviations from the mean, and there are never more than two samples in a row on the same side of the mean line.
25. The *absolute deviation* of a number n from a given value is the absolute value of the difference of n and the given value: Absolute deviation = $|n - \text{given value}|$.
 - a. Find the mean of the following data: 10, 12, 30, 8, 7, 15, 16, 19, 23, 30.
 - b. Find the absolute deviation from the mean for each of the data values in part (a).
 - c. The *mean absolute deviation* is the mean of the absolute deviations you found in part (b). Find the mean absolute deviation for the data values.
 - d. Predict what will happen to the mean absolute deviation of the data in part (a) if a data value that is greater or less than all the other data values is added to the data set. Predict what will happen to the mean absolute deviation if a data value that is equal to the mean is added to the data set. Then check your predictions.