Box-and-whisker plots provide a visual representation of how varied, or spread out, a set of data is. Just as there are different measures that can be used to describe the average of a data set, so there are also different measures that can be used to describe the spread of a set of data. You are already familiar with one measure of spread called the range, or the difference between the greatest and least data values.

Another measure of spread that is commonly used in statistics is called the standard deviation. The standard deviation measures the average distance of a data value from the mean.

EXAMPLE 1 Understanding Standard Deviation

Two students’ test scores on five tests are shown below.

Student A: 81, 84, 85, 87, 88
Student B: 62, 76, 88, 95, 99

a. Find the mean test score for each student.

To find the mean, find the sum of the test scores and then divide by the number of test scores:

Mean for Student A: \[
\frac{81 + 84 + 85 + 87 + 88}{5} = \frac{425}{5} = 85
\]

Mean for Student B: \[
\frac{62 + 76 + 88 + 95 + 99}{5} = \frac{420}{5} = 84
\]

b. Which set of scores would you expect to have the greater standard deviation? Why?

Standard deviation is a measure of average distance from the mean.

Student A’s test scores are all clustered very close to the mean of 85. In fact, the furthest score from the mean, 81, is only 4 points away.

Student B’s scores are much more spread out, with all scores at least 4 points away from the mean of 84. The furthest score from the mean, 62, is 22 points away.

So, you can expect the standard deviation of Student B’s test scores to be greater than that for Student A.

Checkpoint for Example 1

The numbers of runs scored in six games by two baseball teams is shown below.

Red Team: 2, 3, 12, 0, 10, 0
Blue Team: 4, 4, 2, 3, 6, 2

1. Find the mean number of runs scored for each team.

2. Which set of scores can you expect to have the greater standard deviation? Explain.
Finding the standard deviation for a set of data requires several steps. The reasoning behind some of the steps can be quite complex. These reasons will be discussed in a later course.

**KEY CONCEPT**

To find the standard deviation of a set of data:

- **Step 1:** Find the difference between each value and the mean.
- **Step 2:** Square each difference.
- **Step 3:** Find the mean of the squared differences.
- **Step 4:** Take the square root of the mean found in Step 3.

### Example 2

**Finding the Standard Deviation of a Data Set**

Find the standard deviation for Student A’s test scores from Example 1. Round your answer to the nearest hundredth.

#### Step 1

Find the difference between each value and the mean.

In Example 1, the mean of Student A’s test scores was found to be 85. Subtract 85 from each data value.

<table>
<thead>
<tr>
<th>Data Value</th>
<th>81</th>
<th>84</th>
<th>85</th>
<th>87</th>
<th>88</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference from the Mean</td>
<td>−4</td>
<td>−1</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

#### Step 2

Square each difference.

<table>
<thead>
<tr>
<th>Data Value</th>
<th>81</th>
<th>84</th>
<th>85</th>
<th>87</th>
<th>88</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference from the Mean</td>
<td>−4</td>
<td>−1</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Square of the Difference</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>9</td>
</tr>
</tbody>
</table>

#### Step 3

Find the mean of the squared differences.

\[ \text{Mean} = \frac{16 + 1 + 0 + 4 + 9}{5} = \frac{30}{5} = 6 \]

#### Step 4

Take the square root of the mean found in Step 3.

\[ \sqrt{6} \approx 2.45 \]

So, the standard deviation of Student A’s test scores is approximately 2.45.

### Checkpoint for Example 2

3. Find the standard deviation for Student B’s test scores from Example 1.

4. How do the actual standard deviations for each students’ test scores compare to your prediction from Example 1, part (b)?

5. Find the standard deviation for the runs scored for each of the following teams. Round your answers to the nearest hundredth.
   - Red Team: 2, 3, 12, 0, 10, 0
   - Blue Team: 4, 4, 2, 3, 6, 2
Introduction to Standard Deviation

Using Standard Deviation

Dana and Steve each conducted an experiment to find the average respiration rate (in breaths per minute) of goldfish at a certain temperature. The results of their trials are shown in the table.

<table>
<thead>
<tr>
<th>Dana’s Trials</th>
<th>Steve’s Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>85, 90, 91, 91, 91, 91, 91, 91, 92, 93, 93, 93, 93, 95, 96, 96, 97</td>
<td>85, 85, 86, 86, 87, 90, 91, 92, 93, 94, 95, 95, 96, 96, 97</td>
</tr>
</tbody>
</table>

a. Which measure of spread would be better to use to compare the data sets to find how they are different: range or standard deviation?

The minimum and maximum values in each set are the same, so the sets have the same range: 97 – 85 = 12.

Since the values in the data sets differ, it is likely that the standard deviations are different, so it would be better to use the standard deviation to compare the data sets.

b. Find the mean of each data set.

Mean for Dana’s trials: \( \frac{1386}{15} = 92.4 \)  
Mean for Steve’s trials: \( \frac{1368}{15} = 91.2 \)

c. The standard deviation for Dana’s trials is approximately 2.87. How many of Dana’s trials are within one standard deviation of the mean?

A value that is within one standard deviation of the mean falls within the range of values from one standard deviation below the mean, to one standard deviation above the mean. For Dana’s trials, these are any values between 92.4 – 2.87 = 89.53 and 92.4 + 2.87 = 95.27. Of her fifteen trials, 11 of them fall within one standard deviation of the mean.

d. The standard deviation for Steve’s trials is about 4.25. How many of Steve’s trials are within one standard deviation of the mean?

For Steve’s trials, these are any values between 91.2 – 4.25 = 86.95 and 91.2 + 4.25 = 95.45. Of his fifteen trials, 8 of them fall within one standard deviation of the mean.

e. Which set of data is more tightly clustered about the mean?

Dana’s data set has a lower standard deviation and a greater number of data values falling within one standard deviation of the mean. So, the results from Dana’s trials are more tightly clustered about the mean than those from Steve’s trials.

Checkpoint for Example 3

6. Amy repeated Dana and Steve’s experiment and obtained the data shown.

<table>
<thead>
<tr>
<th>Amy’s Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>86, 89, 90, 90, 91, 91, 91, 92, 93, 94, 97, 98, 98, 99, 99</td>
</tr>
</tbody>
</table>

a. What is the mean of this data set?

b. The standard deviation for Amy’s trials is approximately 3.95. How many of Amy’s trials are within one standard deviation of the mean?

c. How do Amy’s results compare to those of Dana and Steve?
**Introduction to Standard Deviation continued**

**Practice**

For each set of data, calculate the mean and standard deviation. Round your answers to the nearest hundredth.

1. 10, 10, 11, 12, 12, 13
2. 3, 5, 7, 7, 7, 7, 7, 7, 7
3. 24, 30, 31, 34, 29, 23
4. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
5. 90, 95, 96, 92, 90, 93, 92
6. Challenge $x, 2x, x, 3x, x, 4x, x, 5x$
7. Writing Calculate the mean and standard deviation of the data set: 0, 0, 0, 0, 0, 0, 0, 0. Now calculate the mean and standard deviation of the data set: 1, 1, 1, 1, 1, 1, 1, 1, 1. Can you generate a data set whose standard deviation is zero even though not every entry is the same number? Explain.

8. **Multiple Choice** Which of the following sets of data do you expect to have the largest standard deviation?
   - A. 1, 3, 5
   - B. 24, 25, 26
   - C. 100, 100, 100
   - D. 10, 20, 30

**Problem Solving**

9. A college lacrosse team scored 9, 9, 11, 12, 9, 17, 9, 11, 10, and 8 times in their first 10 games. What is the mean and standard deviation of this team’s goals?

10. A woman’s lacrosse team scored 12, 8, 17, 17, 15, 21, 16, 12, 10, and 9 times in their first 10 games. What is the mean and standard deviation of this team’s goals? How does it compare with your answer to Exercise 9?

11. Using a chemical assay, a scientist found that the concentration of proteins in a series of test tubes was 30, 35, 34, 36, 34, 33, 34, 35, 34 mg/mL. Using a light assay, the concentrations were read as 30, 34, 39, 26, 31, 32, 35, 37, 40 mg/mL. Calculate the mean concentration of the test tubes for each of the assay methods. Calculate the standard deviation for each data set. Which method was a more precise measurement technique?

12. **Open-Ended Math** Toss a coin 20 times. Record 0 if it lands on heads and 1 if it lands on tails. Calculate the mean and standard deviation of your data set.
Using statistics, it can be shown that the average of all sets of data tends to a normal distribution as the experimenter makes more and more observations. The normal distribution is a mathematical function that describes the probability of observing any given event during the experiment. An especially nice property of normally distributed data sets is that approximately 68% of the observed trials will fall within one standard deviation of the mean, approximately 95% of the observed trials will fall within two standard deviations of the mean and approximately 99.7% of the observed trials will fall within three standard deviations of the mean.

For each set of data, determine the mean and standard deviation, and then determine how many values fall within the first standard deviation.

13. 15, 16, 18, 21, 25, 29, 31
14. 10, 20, 30, 40, 50, 60
15. Most statisticians consider any trial that falls outside the third standard deviation to be an outlier.
   a. Calculate the mean and standard deviation for this set of data: 9, 5, 6, 8, 8, 6.
   b. Draw a box-and-whisker plot for the data.
   c. An extreme outlier is defined to be greater than 3 standard deviations from the mean. Use your answer from part (a) and the given definition to find an extreme outlier larger than the mean.
   d. Now recalculate the standard deviation of the data set, including the extreme outlier from part (c). What happens to the mean and standard deviation?