

8.4

Whose Scores Are Better?

Calculating and Interpreting Standard Deviation

LEARNING GOALS

In this lesson, you will:

- Calculate and interpret the standard deviation of a data set.
- Compare the standard deviation of data sets.

KEY TERMS

- standard deviation
- normal distribution

How many times this year have you asked about your grade in a class? Most students who are serious about their learning and the future are interested in their progress in classes. Some students may even keep track of their own grades throughout the semester. But did you know that every country in the world has its own grading system?

Most likely your school uses letter grades from A to E or F which represent a percent of the points you earned in a class. However, if you went to school in Tunisia, your grades would range from 0 (worst) to 20 (best) and any score below a 10 is a fail. In Denmark, a 7-step-scale is used which ranges from 12 (excellent) to -3 (unacceptable). The grading in Denmark is also very strict with very few students receiving a 12 grade. In some schools in Italy, grades vary from 2 to 8 and each teacher can apply his or her own grading customs. The grades between 5 and 6 could range from 5+, 5++, $5\frac{1}{2}$, 5/6, 6--, 6-. The symbols on these grades have no real mathematical meaning so calculating grades is somewhat arbitrary. Lately though there has been some push to try to get these schools to use a more uniform system like 1 through 10.

Are you familiar with any other grading scales or techniques teachers use in the classroom? Do you think some grading scales are easier or harder than others? Do you think anything else other than earned points can be used to determine a grade?

8

PROBLEM 1 Spelling S U C C E S S

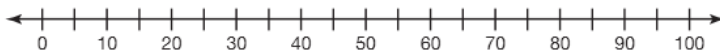
Ms. Webb is determining which student she should add to the spelling bee roster that will represent Tyler High School. The chart shows the 10 most recent scores for three students.

Jack	Aleah	Tymar
33	20	5
32	42	10
30	45	12
50	51	40
49	49	45
50	47	55
35	58	88
73	53	60
71	55	90
77	80	95

1. Determine the mean and median for each student's spelling bee scores.

2. What conclusions can you draw about the data from the mean and median scores?

3. Construct box-and-whisker plots of each student's spelling bee scores using the same number line.



4. Interpret the test scores of each student.

- Jack

- Aleah

- Tymer



5. Do you think these three students performed about the same on all the tests? Why or why not?

8

PROBLEM 2 So . . . Who Did Better?



You have learned about the spread of data values from the IQR, which is based on the median. However, is there a way to measure the spread of data from another measure of central tendency? **Standard deviation** is a measure of how spread out the data is from the mean. A formula can be used to determine the standard deviation of a data set. A lower standard deviation represents data that are more tightly clustered. A higher standard deviation represents data that are more spread out from the mean.

So, if the IQR is the measure of how spread out data is from the median, and standard deviation is the measure of how spread out data is from the mean, I wonder which will be affected by outliers?



The symbol to the left of the equals sign is a lower case sigma. This represents the standard deviation.

The formula to determine standard deviation of a population is represented as:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

where σ is the standard deviation, x_i represents each individual data value, \bar{x} represents the mean of the data set, and n is the number of data points.



Let's look at each part of the standard deviation formula separately.

Follow the steps to determine the standard deviation. Let's use the data set 6, 4, 10, 8 where $\bar{x} = 7$.

First, think of each data value as its own term labeled as $x_1, x_2,$ and so on.	$x_1 = 6$
	$x_2 = 4$
	$x_3 = 10$
	$x_4 = 8$

The first part of the formula identifies the terms to be added. Since n represents the total number of values and $i = 1$, add all the values that result from substituting in the first term to the fourth term.

$$\sum_{i=1}^n$$

This part of the formula just gives you information. You will not sum anything until after the next step.

© 2012 Carnegie Learning

8

Next, evaluate the expressions to be added. Take each term and subtract it from \bar{x} and then square each difference.

$$(x_i - \bar{x})^2$$

$$(6 - 7)^2 = 1$$

$$(4 - 7)^2 = 9$$

$$(10 - 7)^2 = 9$$

$$(8 - 7)^2 = 1$$

Now determine the sum of the squared values and divide the sum by the number of data values.

$$\frac{1 + 9 + 9 + 1}{4} = \frac{20}{4} = 5$$

Finally, calculate the square root of the quotient.

$$\sigma = \sqrt{5}$$

$$\sigma \approx 2.24$$

The standard deviation is approximately 2.24.

So the standard deviation for the given data set is approximately 2.24. It is important to note that if the data values have a unit of measure, the standard deviation of the data set also uses the same unit of measure.

1. Do you think the standard deviation for each student's spelling bee scores will be the same? If yes, explain your reasoning. If no, predict who will have a higher or lower standard deviation.

8



2. Now, let's use the standard deviation formula to determine the standard deviation of Jack's spelling bee scores.

a. Identify the data values you will use to determine the standard deviation. Explain your reasoning.

b. Determine the \bar{x} value.


c. Complete the table to represent each part of the formula. The data values have been put in ascending order.

x_i	$(x_i - \bar{x})^2$
30	$(30 - 50)^2 = 400$
32	
33	
35	
49	
50	
50	
71	
73	
77	
$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$	
$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$	



3. Determine the standard deviation for Jack's spelling bee scores and interpret the meaning.

© 2012 Carnegie Learning



You can use a graphing calculator to determine the standard deviation of a data set.

8

Step 1: Press **STAT** and then **ENTER** to select **1:Edit**. Enter each data set into its own List.

Step 2: Press **STAT** then scroll to the right to highlight **CALC**. Press **ENTER** to select **1:Var-Stats**. Press **ENTER**.

Step 3: Your screen should display 1-Var Stats. Press **2ND** then the list you want the calculator to use for these calculations.

Step 4: Your calculator should display the same data values as when you determined the mean. However, this time use the value for σ_x .

These steps are similar to the steps used to determine the mean on the graphing calculator.

You can enter Jack's data in L₁, Aleah's data in L₂, and Tymar's data in L₃.

So to determine the standard deviation of Jack's data you would enter **2nd L₁**.



4. Use a graphing calculator to determine the standard deviation of Aleah's and Tymar's spelling bee scores.

5. Was the prediction you made in Question 1 correct? What do the standard deviations tell you about each student's spelling bee scores?



6. Which student do you think Ms. Webb should add to the spelling bee roster? Use the standard deviation for the student you recommend to add to the roster to justify your answer.

8

PROBLEM 3 Working as a Team

Recall Lesson 8.2, *Does Height Really Matter?* The Mountain View High School basketball team has its first game of the season and Coach Maynard is comparing the heights of the home team's top ten players to the heights of the visiting team's top ten players. The dot plots of the data are given.



- Predict which team has the greatest standard deviation in their heights. Explain how you determined your answer.



- Determine the standard deviation of the heights of each team. Describe what this means in terms of this problem situation. How does this information help Coach Maynard?

PROBLEM 4 68–95–99: The Combination to Standard Deviation**8**

So far, you have determined the standard deviation for different data sets. You have also interpreted the standard deviations to make decisions given a problem situation. Standard deviation can also be represented graphically by graphing a data set.

Recall that Ms. Webb is the spelling bee coach in Problem 1, *Spelling S U C C E S S*. Her class is preparing for their first spelling bee scrimmage. Ms. Webb needs to determine which student should be the spelling bee captain. Ms. Webb believes the captain should have the greatest mean score of the team. The two top spelling bee students' scores are shown.

Maria	Heidi
81	81
73	68
94	60
86	109
70	82
68	88
97	60
93	102
81	78
67	69
85	84
77	103
79	92
103	60
90	108



1. Analyze each student's spelling bee scores.
 - a. Determine the mean spelling bee score for each student.

8

Ms. Webb wants to also use the standard deviation to help her determine which student is a more consistent speller.

- b. Determine the standard deviation of Maria's scores. Then determine the value of the spelling bee scores that are 1 standard deviation from the mean. Explain how you determined her spelling bee point values.

Make sure to use the mean to the nearest hundredths place.



- c. Determine the standard deviation of Heidi's scores. Then determine the value of the spelling bee scores that are 1 standard deviation from the mean. Explain how you determine her spelling bee point values.



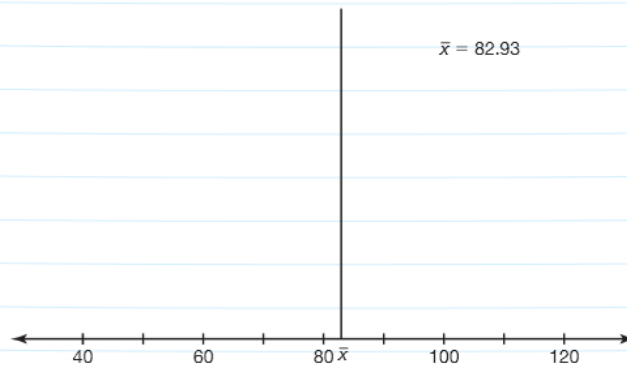
You have calculated 1 standard deviation for the data sets in previous problem situations. However, you can also determine different numbers of standard deviations. For example, 2 standard deviations or greater are calculated by multiplying the standard deviation by the number of standard deviations you are determining. Therefore, if a data set has a standard deviation of 15, then 2 standard deviations would be 30, and 3 standard deviations would be 45.

When you determine the standard deviation of a data set, you can represent it graphically. You can also determine the general percent of data values that are within 1 standard deviation, and the percent of data values that lie within 2 standard deviations in *normal distributions*. A **normal distribution** is a collection of many data points that form a bell-shaped curve.

8

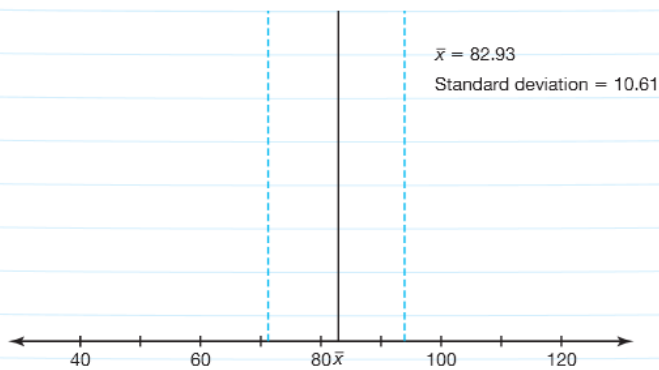
The mean of Maria's spelling bee scores is 82.93 points and 1 standard deviation is 10.61 points.

To graph the normal distribution of Maria's spelling bee scores, first graph the mean on a number line as: $x = 82.93$.



Next, graph 1 standard deviation from the mean. For Maria's spelling bee scores, the standard deviation is 10.61. Therefore, the values of the standard deviation from the mean are 72.32 and 93.54.

Use a dotted lines as $x = 72.32$ and $x = 93.54$.



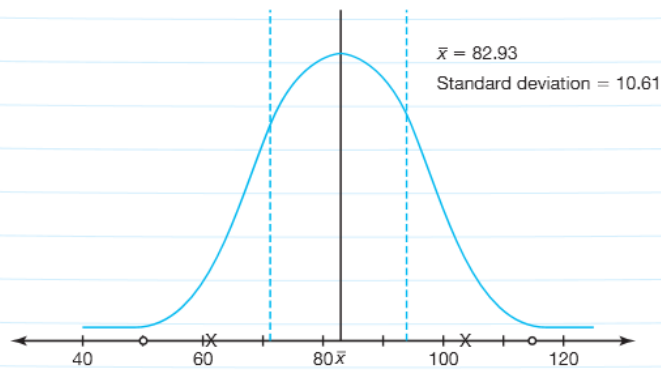
8

Then graph 2 standard deviations and 3 standard deviations from the mean. To determine 2 standard deviations, multiply the standard deviation by 2. To determine 3 standard deviations, multiply the standard deviation by 3.

For Maria's scores, 2 standard deviations would be 21.22 and 3 standard deviations would be 31.83.

Mark an "x" for the two points as 61.71 and 104.15 represent a standard deviation of 2. Mark an "o" for the two points 114.76 and 51.10 for a standard deviation of 3.

Finally draw a smooth curve starting from the far left minimum value. The smooth curve should resemble a bell-shaped curve.



2. Describe some observations you can make about the graph of Maria's spelling bee scores.

3. Plot each of Maria's scores on the graph of the worked example. Mark an "x" for the approximate location on the number line for each score.

To plot the scores in the graph, mark x's like you would with dot plots.

8

- a. Determine how many spelling bee scores are within 1 standard deviation of the mean for Maria's spelling bee scores.
- b. Determine how many spelling bee scores are within 2 standard deviations of the mean for Maria's spelling bee scores.
- c. Determine how many spelling bee scores are within 3 standard deviations of the mean for Maria's spelling bee scores.



Within the graph of a normal distribution, you can predict the percent of data points that are within one, two, or three standard deviations from the mean. Generally, 68% of the data points of a data set will fall within one standard deviation of the mean; while 95% of the data points of a data set will fall within two standard deviations of the mean; and 99% of the data points of a data set will fall within three standard deviations of the mean.



4. Analyze the number of data points you determined lie within 1, 2, or 3 standard deviations.
- a. What percent of data points from Maria's spelling bee scores fall within 1 standard deviation of the mean? Explain how you determined your answer.
- b. What percent of data points from Maria's spelling bee scores fall within 2 standard deviations from the mean? Explain how you determined your answer.

8

- c. What percent of data points from Maria's spelling bee scores fall within 3 standard deviations from the mean? Explain how you determined your answer.



- d. Did the prediction about the percent of data points that fall within 1, 2, or 3 standard deviations match Maria's data set? Why do you think it did or did not?



It is important to note that the guideline regarding 68%, 95% and 99% is simply a guideline. In fact, there may be some data sets in which all of the data points lie within two standard deviations of the mean while other data sets may actually need four or greater standard deviations to encapsulate the entire data set. It is also important to know that because standard deviation is based on the mean of a data set, outliers may affect the standard deviation of the data set.



5. Graph 1, 2, and 3 standard deviations on the number line shown for Heidi's scores using a bell-shaped curve.



6. What similarities and differences do you notice between Maria's spelling score graph and Heidi's spelling score graph?



7. Advise Ms. Webb whom she should choose to captain the spelling bee team given the information about each student's standard deviation.

8

Talk the Talk



Mean and median are both measures of central tendency.

1. Identify which is more resistant to outliers, and which is more sensitive to outliers. Explain your reasoning.

The interquartile range and the standard deviation both measure the spread of data.

2. Identify which is more resistant to outliers, and which is more sensitive to outliers. Explain your reasoning.



Be prepared to share your solutions and methods.