Standard Deviation
Video Script

Scene		Full Transcript
1	**Emily:**	Hey there. Have you ever planned a trip only to have it spoiled by bad weather? If so, you know the importance of good weather data. Today, I'm at the weather center with Dr. Alan Zarnetski learning about the systems meteorologists use to provide up-to-the-minute forecasts.
2	**Voice-Over Emily:**	They use all sorts of statistics when making their predictions including a measure of variation called standard deviation. Accurate predictions of the weather are important in all sorts of professions and in our lives.
3	**Emily:**	But even if you don't care whether it rains or shines, understanding standard deviation can be important. Stick around. I predict we'll make sense of standard deviation and get another *Problem Solved*.
4	**Emily:**	With a click of the mouse, Dr. Zarnetski can access and compare temperature data from anywhere in the world. Oh, I know, let's check in on some of our friends stationed at military bases.
5	**Voice-Over Emily:**	Here are the high temperatures last week from Elmendorf Air Force Base in Alaska, Thule Air Force Base in Greenland, and Osan Air Force Base in South Korea. Brrrrrr! We can see from the data that the temperatures were different at each base. A quick calculation shows that last week the mean temperature and the median temperature were both 5° at Elmendorf. Surprisingly, Thule and Osan have the same mean and median temperatures.
6	**Emily:**	Even though the three bases have the same mean and median, notice that they are experiencing very different temperatures. The mean and median alone do not tell the whole story. We need to have a measure of how much the data varies. One measure of this variation is the standard deviation.
7	**Voice-Over Emily:**	Here are dot plots of the data for each of the air bases. Think of standard deviation as the spread of the data points. Does the data cluster closely about the mean or are there many data points far away from the mean?
8	**Emily:**	We want a single number that represents this information. In our example, the mean is 5°, and we need to express how far, on average, the individual temperatures are from five. So, let's find the average of these distances from the mean.
9	**Voice-Over**	We'll start with the Elmendorf data. Subtract the mean from each temperature. Then, find the average of those differences. Notice that the

	Emily:	sum of the differences below the mean is -11, and the sum of the differences above the mean is +11. If we compute the average, we get zero. This isn't a coincidence. Look at the average of the differences for Thule and Osan, also zero.
10	**Emily:**	So, we need a different approach. If we square each difference, then we'll have a collection of positive values to average. Watch this.
11	**Voice-Over Emily:**	This is what we get when we square the differences. The sum of the squared differences is 88. Since there were seven total values, we divide 88 by 7 to find the average square difference. This rounds to 12.57. However, since we squared the values, we take the square root of 12.57, which rounds to 3.5. This value is our standard deviation.
12	**Emily:**	Remember we're working with degrees, but when we square the differences our units became squared degrees. Squared degrees? That's not going to work.
13	**Emily:**	When we found the average, we still wound up with square degrees, but by taking the square root, our standard deviation goes back to degrees.
14	**Emily:**	The standard deviation provides us a single value that indicates how much the data in the set varies from the mean. I wonder if the Thule Air Force Base temperatures have a smaller or larger standard deviation.
15	**Voice-Over Emily**	Just by looking at the data, we can tell the daily temperatures are closer to the mean. That indicates it would have a smaller deviation. The data in the set varies less. Let's go through the steps of finding the standard deviation but this time using mathematical notation.
16	**Voice-Over Emily:**	Here is a quick look at the symbols we will be using: x_1 is the temperature on the first day, x_2 is the temperature on the second day, etc. x_i is the temperature on the ith day. In our case i runs from one through seven. μ is the mean of the data. n is the total number of pieces of data. $\sum_{i=1}^{n}$ is the sum of an expression from the first piece of data to the last piece of data. Now, it's time to compute the standard deviation. First, compute the mean, 5. Next, find each value's difference from the mean. Next, square the differences. Then, find the sum of the squared differences; this is 2. Now, divide by the number of pieces of data, n; this is 2/7 or

		approximately 0.286. Finally, take the square root of the result; this is approximately 0.535. Just as we suspected, the standard deviation for Thule is quite a bit smaller than the one for Elmendorf.
17	**Emily:**	What about the standard deviation for Osan?
18	**Voice-Over Emily:**	Since it has the data that varies the most, it should have the largest standard deviation. Calculating, we get about 17.05 for the standard deviation.
19	**Emily:**	In our temperature example, we used all of the data. Most of the time, statisticians calculate the standard deviation for a sample of the data instead of all of the data. The entire set of data is called the population.
20	**Voice-Over Emily:**	When a sample is used, dividing by n could under-estimate the true standard deviation. To compensate for that, statisticians divide by (n – 1) when using a sample of a population. Many calculators and computer software programs calculate both the standard deviation for a population and the standard deviation for a sample of the population.
21	**Emily:**	My forecast was correct; the weather center is the perfect place for making sense of standard deviation. The standard deviation (a measure of spread) along with the mean (a measure of center) give us a good understanding of our data. *Problem Solved*.