2022Q10 STANDARD DEVIATION

Employees at One Energy deal with data on a daily basis. Whether it is using wind data to calculate average wind speeds and project power production, working with financial data to provide an accurate project lifetime financial model, or using energy data to assess a customer's energy usage, employees in every department in One Energy work with data frequently. This is why the field of statistics – the practice of collecting, analyzing, and interpreting data – is essential to the operation of One Energy.

This Wind Study will focus on one important statistical metric: standard deviation. Standard deviation is the measure of the dispersion of a set of data relative to its mean. There are two different types of standard deviation, one for a sample and one for a population. The difference between the two is that a sample is a subset of a larger group of data to be analyzed and a population is the whole of the data to be analyzed. We'll focus on the standard deviation of a population for the purposes of this wind study. See the equation below:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}},$$

Where σ is the standard deviation, x_i is a particular data point, μ is the mean of the data, and N is the size of the data set. Let's not forget about that funky "E". That funky "E" ($\sum_{i=1}^{N}$) is the symbol for a sum. In the case of this equation that means we sum the square of the difference from the mean for each point, like this:

$$\sum_{i=1}^{N}(x_i - \mu)^2 = (x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 \ldots (x_N - \mu)^2.$$

Now that we have a better understanding of the standard deviation equation, let's talk about what the standard deviation metric tells us about our data. As we stated above, the standard deviation is a measure of the dispersion of a set of data relative to its mean. This is shown in Figure 1.

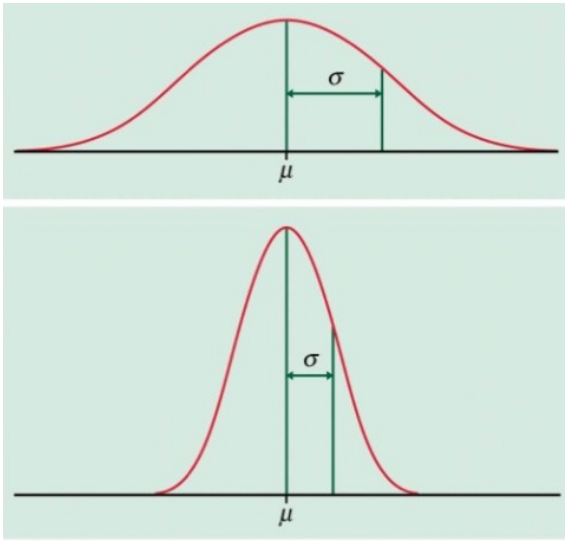

Figure 1: High standard deviation from the mean vs. low standard deviation from the mean[1]

A low standard deviation, in terms of the data set being investigated, indicates the data is less disperse and centers around the mean. On the other hand, a high standard deviation indicates the data is more dispersed. Standard deviation can tell us whether the mean is an accurate representation of the data set as a whole. With this knowledge in mind, let's dive into our questions!

[1] The University of North Carolina at Chapel Hill "Density Curves and Normal Distributions" 9/12/06. Web.

WIND STUDY

Wind Study is intended for grades 5-8 and 8-11
Questions posted on: Monday Answers posted on: Friday
Find downloadable one pagers at www.oneenergy.com/one-energy-feed

Level 1: Below is a set of wind speed data. Find the mean of wind speed and power production.

WIND DATA		
Time Stamp	**Wind Speed (m/s)**	**Power Production (kW)**
00:00	5.24	32.7
00:10	5.46	44.0
00:20	5.32	44.0
00:30	4.93	32.7
00:40	4.82	32.7
00:50	4.65	23.7
01:00	4.72	23.7
01:10	4.80	32.7
01:20	4.86	32.7
01:30	4.58	23.7
01:40	4.16	15.8
01:50	3.91	15.8

Level 2: Find the standard deviation of the wind speed and power production. What do you think the standard deviation tells us about these sets of data?