# 2022Q11 VARIANCE

In last week's wind study, we talked about how One Energy regularly deals with data. Whether it be energy, financial, or weather, data is essential to the operation of One Energy. And where there's data, there is a need to analyze it. Because of this need, the field of statistics - the practice of collecting, analyzing, and interpreting data – is essential to the operation of One Energy. So, this week we're doubling down and visiting another important statistical metric: variance.

*Figure 1: One Energy employees work with weather, energy, and financial data on a daily basis*

Variance is a measure of the dispersion of a set of data relative to its mean. Like standard deviation, there are two types of variances, one for a sample and one for a population. Recall that a sample is a subset of a larger group of data you wish to draw conclusions from, whereas a population is the whole of the data you wish to draw conclusions from. We'll again focus on the population form of variance for the purposes of this wind study. Let's look at the variance equation below:

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

Where $\sigma^2$ is the variance, $x_i$ is a particular data point, $\mu$ is the mean of the data, $N$ is the size of the data set, and that funky "E" ($\sum_{i=1}^{N}$ ) is the symbol for a sum. By comparing the equation for variance with the equation for standard deviation from last week, we can see that the standard deviation is equal to the square root of the variance. This means that the variance has larger units in terms of the mean.

Like standard deviation, high measures of variance indicate a high dispersion in relation to the mean of a dataset, while low measures of variance indicate low dispersion around the mean of a dataset. The best way to think about standard deviation and variance is two ways to measure the dispersion of data around a mean and receive the same result. As you get deeper into the field of statistics, the applications of both measures diversify, but for now it is important to understand the relationship between standard deviation and variance. That being said, let's dive into our questions for this wind study.

*Figure 2: One Energy uses statistics to help assess project performance*

**Level 1:** Below is a set of data. Find the variance of the power produced and capacity factor.

| Monthly Turbine Energy Production Data | | |
|---|---|---|
| **Month** | **Energy Produced (MWh)** | **Capacity Factor** |
| January | 763 | 22.80% |
| February | 1,113 | 36.79% |
| March | 1,329 | 39.69% |
| April | 931 | 28.73% |
| May | 723 | 21.61% |
| June | 668 | 20.62% |
| July | 406 | 12.13% |
| August | 402 | 12.00% |
| September | 927 | 28.61% |
| October | 726 | 21.69% |
| November | 1,191 | 36.76% |
| December | 1,163 | 34.74% |

**Level 2:** Find the standard deviation of the power produced and capacity factor. How do the standard deviation and variance differ? How do they describe this set of data?