2022A11 VARIANCE

Level 1: To calculate the variance of a set of data, we must first calculate the mean of the dataset. For the energy production data, this would be:

$$\mu_{EP} = \frac{763 + 1{,}113 + 1{,}329 + 931 + 723 + 668 + 406 + 402 + 927 + 726 + 1{,}191 + 1{,}163}{12}$$

$$\mu_{EP} = 862\ MWh.$$

Now we can use the variance equation discussed earlier. For the energy produced, this would be:

$$\sigma_{EP}^2 = \frac{\sum_{i=1}^{N}(x_i - \mu_{EP})^2}{N}$$

$$\sigma_{EP}^2 = \frac{(763 - 862)^2 + (1{,}113 - 862)^2 + \cdots + (1{,}163 - 862)^2}{12}$$

$$\sigma_{EP}^2 = 82{,}764\ MWh^2$$

We can follow this same process to calculate the variance of the capacity factor data.

$$\mu_{CF} = \frac{22.80\% + 36.79\% + 39.69\% + \cdots + 34.74\%}{12}$$

$$\mu_{CF} = 26.35\%.$$

Using the mean of the capacity factor, we can solve for variance.

$$\sigma_{CF}^2 = \frac{\sum_{i=1}^{N}(x_i - \mu_{CF})^2}{N}$$

$$\sigma_{CF}^2 = \frac{(22.80\% - 26.35\%)^2 + (36.79\% - 26.35\%)^2 + \cdots + (34.74\% - 26.35\%)^2}{12}$$

$$\sigma_{CF}^2 = 0.81\%^2$$

Level 2: To find the standard deviation of the energy produced and capacity factor, we can compare the standard deviation and variance equations:

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

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

Comparing these two equations, we can see that the standard deviation is equal to the square root of variance. So, to find the standard deviation of the data sets, we can say:

$$\sigma = \sqrt{\sigma^2}.$$

For the energy production data set this becomes:

$$\sigma_{EP} = \sqrt{0.81\%^2}$$

$$\sigma_{EP} = 9.01\%.$$

For the capacity factor data set this becomes:

$$\sigma_{CF} = \sqrt{82{,}764\ MWh^2}$$

$$\sigma_{CF} = 287.67\ MWh.$$

The biggest difference between the standard deviation and the variance is the units. As we can see in this solution, variance has squared units. This means that the variance is generally larger than the standard deviation, unless the standard deviation is below 1, as we can see in this case. It also means that for a mean of 862 MWh, a variance of 82,764 MWh^2 tells us the same thing as a standard deviation of 287.67 MWh. Standard deviation and variance are just different ways of interpreting a data set.

For the energy production data, the standard deviation of the data set is 287.67 MWh. This means that most of the data points are within plus or minus 287.67 MWh of the mean, which makes sense when we take the annual variation of wind speeds into account. A variance of 82,764 MWh^2 tells us the same thing.

For the capacity factor data, the standard deviation of 9.01% tells us most of the data points are within 9.01% of above or below the mean. This tracks with the annual variation of wind and also tells us that turbine power production varies throughout the year.