Text: Introduction to the Standard Deviation and Variance SEND FEEDBACK

Other Measures of Spread

5 Number Summary

In the previous sections, we have seen how to calculate the values associated with the **five number summary** (**min**, Q_1, Q_2, Q_3, **max**), as well as the measures of spread associated with these values (**range** and **IQR**).

For datasets that are **not symmetric**, the five number summary and a corresponding box plot are a great way to get started with understanding the spread of your data. **Although I still prefer a histogram in most cases, box plots can be easier to compare two or more groups.** You will see this in the quizzes towards the end of this lesson.

Variance and Standard Deviation

Two additional **measures of spread** that are used all the time are the **variance** and **standard deviation**. At first glance, the variance and standard deviation can seem overwhelming. If you do not understand the expressions below, don't panic! In this section, I just want to give you an overview of what the next sections will cover. We will walk through each of these parts thoroughly in the next few sections, but the big picture goal is to generally understand the following:

1. How the mean, variance and standard deviation are calculated.
2. Why the measures of variance and standard deviation make sense to capture the spread of our data.
3. Fields where you might see these values used.
4. Why we might use the standard deviation or variance as opposed to the values associated with the 5 number summary for a particular dataset.

Calculation

We calculate the variance in the following way:

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

The variance is **the average squared difference of each observation from the mean**.

To calculate the variance of a set of 10 values in a spreadsheet application, with our 10 data points in column A, we would create a new column B by typing in something like **=A1-AVERAGE(A$1:A$10)** and copying this down for all 10 rows. This would find us the difference between each data point and the mean average of all the data. Then we create a new column C having the square of these differences, using the formula **=B1^2** in cell C1, and copying that down for all rows. Then in the cell below this new column, cell C11, type in **=SUM(C1:C10)**. This adds up all these values in column C. Finally in cell C12, we divide this sum by the number of data points we have, in this case ten: **=C11/10**. This cell C12 now contains the variance for our 10 data points.