**The Role of Standards in a Design of Experiment (DOE)**

So now that you have gotten a chance to work with the averages and find the three effect values, we will now turn our attention to the standards. Standards are used for a variety of reasons. First, they are often the “optimal value” in the experiment. Scientists frequently want to see how high above or how far below an optimal value they can safely go. Second, they are used as a basis of comparison. In any of your research projects, normal biological variation, researcher bias and environmental variation are all factors that can skew data, so your standards will provide a baseline for you to compare data. Your standards can help to show whether the experimental design you came up with is capable of actually generating “good results”. This will depend on either the range (variability) or any trends in the standards.

The purpose of using standards is to help you determine how much variability there is in an experiment. If there is a lot of variability or if there are trends in the standards that is a signal that something is wrong with the experimental design and you should investigate and fix/adjust the experiment.

A large amount of variability in your standards might indicate confounding or the presence of lurking variables in your experiment. While you won’t be able to identify with certainty all of them, you can and should offer a conjecture as to which ones might have been present in your experiment. **One way to do that is to find the range of standards (maximum – minimum).** Let’s do this with our diffusion lab:

*Data of Standards*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Standard (Time in Seconds)* | | | | | | | | |
| *1379.49* | *1683.02* | *1230.1* | *1121* | *1197* | *1336* | *1298.37* | *932* | *579.34* |

9A Standards 9B Standards 9C Standards

Now compare this to the overall variability (range) of the ENTIRE experiment. Our minimum value was 579.34 seconds and our maximum value was 5,064.22 seconds.

Should one range be larger than the other? Explain your reasoning.

Does this comparison imply anything about our experimental design? What about any confounding? Lurking variables?

**Another way to use the standards is to graph them as a scatter plot and look for trends.** The goal is to have all of the dots line up horizontally.   
*(NOTE: This almost NEVER happens.)* Let graph our standards now:

Do you see any upward or downward trends?

If you are doing your research over several days, be sure to keep those standards blocked when you graph your scatter plot. It will be much easier to spot upward or downward trends that way. In our case, we will keep 9A, 9B, and 9C blocked since they were done at different times, by different experimenters.

Let’s also compare the mean standard and range from each class. What do you notice? Who had the most consistent results? What suggestions do you have to improve the experimental design or execution of the experiment?

The last way that we will use our standards as a basis for comparison is to **check to see which of our three effect values (Two main effects and the interaction effect) is deemed statistically significant.**

Tests of significance are statistical tools that help us make decisions about changes to responses. Without these tools, we might look at a change in a process output and think that it is important, but the change was just part of the common cause variation from the process. Tests of significance give us a statistical basis for determining if a change in factor levels leads to a statistically significant effect on the process response.

But what does it mean to be statistically significant? It means **a result that is not likely to occur randomly, but rather is likely to be attributable to a specific cause.** Statistical significance can be strong or weak, and is important to research in many math- and science-related fields, including medicine, sociology, psychology and biology. **Statistical significance does not always indicate practical significance.** In addition, it can be misinterpreted when researchers do not use language carefully in reporting their results.  
  
 A note on interactions:

**Two Types of Interactions**

**Synergistic** – variables together are good, e.g. teamwork  
  
 **Antagonistic** – Joint effect of variables is bad, e.g. certain drug  
 **Note that when interactions are present: 2 + 2 is not equal to four!**  
The sum is not equal to the total of the parts! Thus it is of utmost importance to measure interactions when testing.

Therefore it is necessary to look at the expected average values for the solid and dashed line segments that you have drawn and compare them to your interaction values. This is an EXTREMELY important part of the interpretation of your data.

One way to test for significance is to set up fences on a dot plot of the three effects.

Dot Plotsare a great way to visually show the reader which effect value(s) played a significant role in the experiment. After your dot plot of all three effects is created, put in a left and a right fence that represents double the range of standards. Let’s do this now:

Are any of the effects outside of the fences? What do you think this means?