2.2: One-Way ANOVA

The one-way ANOVA is sometimes also called a between-subjects ANOVA, or an independent factor ANOVA. The critical ingredient for a one-way between-subjects ANOVA, is that you have one independent variable, with at least two-levels. When you have one IV with two levels, you can run a t-test. You can also run an ANOVA. Interestingly, they give you almost the exact same results. You will get a p-value from both tests that is identical (they are really doing the same thing under the hood). The t-test gives a t-value as the important sample statistic. The ANOVA gives you the F-value (for Fisher, the inventor of the test, who is one of the most important statisticians in the history of the field) as the important sample statistic. It turns out that t^2 equals F, when there are only two groups in the design.

F is computed directly from the data. In fact, the idea behind F is the same basic idea that goes into making t. Here is the general idea behind the formula, it is again a ratio of the effect we are measuring (in the numerator), and the variation associated with the effect (in the denominator).

$$\text{name of statistic} = \frac{\text{measure of effect}}{\text{measure of error}}$$

$$\text{F} = \frac{\text{measure of effect}}{\text{measure of error}}$$

The difference with F, is that we use variances to describe both the measure of the effect and the measure of error. So, F is a ratio of two variances.

When the variance associated with the effect is the same size as the variance associated with sampling error, we will get two of the same numbers, this will result in an F-value of 1. When the variance due to the effect is larger than the variance associated with sampling error, then F will be greater than 1. When the variance associated with the effect is smaller than the variance associated with sampling error, F will be less than one.

Let’s rewrite in plainer English. We are talking about two concepts that we would like to measure from our data. 1) A measure of what we can explain, and 2) a measure of error, or stuff about our data we can’t explain. So, the F formula looks like this:

$$\text{F} = \frac{\text{Can Explain}}{\text{Can't Explain}}$$

When we can explain as much as we can’t explain, F = 1. This isn’t that great of a situation for us to be in. It means we have a lot of uncertainty. When we can explain much more than we can’t we are doing a good job, F will be greater than 1. When we can explain less than what we can’t, we really can’t explain very much, F will be less than 1. That’s the concept behind making F.

If you saw an F in the wild, and it was .6. Then you would automatically know the researchers couldn’t explain much of their data. If you saw an F of 5, then you would know the researchers could explain 5 times more than the couldn’t, that’s pretty good. And the point of this is to give you an intuition about the meaning of an F-value, even before you know how to compute it.