11.5: Hypotheses in ANOVA

So far we have seen what ANOVA is used for, why we use it, and how we use it. Now we can turn to the formal hypotheses we will be testing. As with before, we have a null and an alternative hypothesis to lay out. Our null hypothesis is still the idea of "no difference" in our data. Because we have multiple group means, we simply list them out as equal to each other:

$$\mathrm{H}_0: \text{ There is no difference in the group means}$$
$$\mathrm{H}_0: \mu_1 = \mu_2 = \mu_3$$

We list as many μ parameters as groups we have. In the example above, we have three groups to test, so we have three parameters in our null hypothesis. If we had more groups, say, four, we would simply add another μ to the list and give it the appropriate subscript, giving us:

$$\mathrm{H}_0: \text{ There is no difference in the group means}$$
$$\mathrm{H}_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

Notice that we do not say that the means are all equal to zero, we only say that they are equal to one another; it does not matter what the actual value is, so long as it holds for all groups equally.

Our alternative hypothesis for ANOVA is a little bit different. Let's take a look at it and then dive deeper into what it means:

$$\mathrm{H}_A: \text{ At least one mean is different}$$

The first difference in obvious: there is no mathematical statement of the alternative hypothesis in ANOVA. This is due to the second difference: we are not saying which group is going to be different, only that at least one will be. Because we do not hypothesize about which mean will be different, there is no way to write it mathematically. Related to this, we do not have directional hypotheses (greater than or less than) like we did in Unit 2. Due to this, our alternative hypothesis is always exactly the same: at least one mean is different.

In Unit 2, we saw that, if we reject the null hypothesis, we can adopt the alternative, and this made it easy to understand what the differences looked like. In ANOVA, we will still adopt the alternative hypothesis as the best explanation of our data if we reject the null hypothesis. However, when we look at the alternative hypothesis, we can see that it does not give us much information. We will know that a difference exists somewhere, but we will not know where that difference is. Is only group 1 different but groups 2 and 3 the same? Is it only group 2? Are all three of them different? Based on just our alternative hypothesis, there is no way to be sure. We will come back to this issue later and see how to find out specific differences. For now, just remember that we are testing for any difference in group means, and it does not matter where that difference occurs.

Now that we have our hypotheses for ANOVA, let's work through an example. We will continue to use the data from Figures 11.1.1 through 11.1.3 for continuity.