The Bonferroni Correction

Whenever we choose a p-value cutoff we are also limiting the frequency of *false positive* tests. A commonly used significance cutoff is $\alpha = 0.05$. This cutoff implies that 1 out of 20 of our tests will be false positives.

Even a stringent cutoff like 1/20 can be problematic with high dimensional data. Imagine we have a model that uses expression levels of all 25,000 human genes. When looking for significant effects, a cutoff of $\alpha = 0.05$ would give $0.05 \times 25,000 = 1250$ false positive genes! This issue is called the *multiple testing problem*. As the number of significance tests increases, so does the number of false positives we observe. One solution to the multiple testing problem is *multiple testing correction*. Rather than choose a single significance cutoff for all our models, we can adjust the cutoff based on the number of parameters in the model to limit the false positive rate.

A commonly used strategy for multiple testing correction is the Bonferroni correction. If our standard significance cutoff for individual tests is α, we use a cutoff of α/n for a model with n parameters. Notice that the Bonferroni correction depends on the number of parameters, not the number of samples or data points. We are testing the parameters for significance, so n must be the number of parameters. In Homework 5, we built a model with $10 + 45 = 55$ parameters, so a Bonferroni-corrected p-value is $0.05/55 \approx 0.0009$.

The Bonferroni correction limits our expected number of false positive to one per model. It is also the simplest method for multiple testing correction, but others exist. We should also note that the statistics community is split as to whether multiple testing correction should be performed at all. (Your instructor, for example, is not the world's biggest fan of multiple testing correction.) Regardless, you should be aware of the false discovery rate for your models, especially when using high dimensional data.