Multiple testing
(aka multiple comparisons)

Stat 340

When a null hypothesis is true,

what is the probability that you reject at level .05?

- For simplicity, throughout this lecture “rejecting a null hypothesis” means the p-value is less than .05.
- If the null hypothesis is true, what is the probability that you reject?
- If the null hypothesis is false, should the probability of rejection be larger or smaller?

What happens when you test 100 null hypotheses?

- Suppose you test 100 null hypotheses, all of which are true, all tests independent.
- Let X be the number of null hypotheses with p-values less than .05.
- What is the distribution of X?

Why is this a problem?

- What if you only talked about the p-values that are less than .05?

Here are some good practices…

- Good practice: specify all the hypothesis tests that you are going to perform **before you actually look at the p-values**
- Good practice: tell people all the tests you've performed
- Good practice: don't study data that lots of people have studied
- Good practice: multiply your p-values by the number of tests you've performed. This has a fancy name. So you can be fancy. It's called the "Bonferroni correction". https://en.wikipedia.org/wiki/Bonferroni_correction
    - What if your p-value is .01, that sounds great… but you tested 10 null hypotheses, what is your new Bonferroni p-value?
- Good practice: If you want to get serious about multiple testing, you can learn about "false discovery rates." It isn't as mean as Bonferroni.

What are some cases where multiple testing occurs,

… but it might not be so obvious.

Let’s do a simulation.