# NULL DISTRIBUTIONS

Null hypothesis is a statement that is the skeptical viewpoint of your research question. In other words, the pattern you are interested in doesn't exist.

Alternative hypothesis is a statement that is the positive viewpoint of your research question. In other words, the pattern you are interested in exists.

Crafting the correct null and alternative hypotheses is critical to hypothesis testing. If they are incorrect, then the test of statistical significance will be meaningless because it won't be aligned correctly to the research question. Here are some important points to keep in mind.

* Mutually exclusive The alternative hypothesis contains everything not in the null hypothesis, which means the null and alternative are mutually exclusive.
* Exhaustive The null hypothesis and alternative hypothesis describe all possible outcomes.
* Equality The null hypothesis always includes the equality statement.

The null distribution is the core machinery of any statistical test. The null distribution is the sampling distribution that you would get if you repeatedly sampled from a fictitious statistical population where the null hypothesis was true.

The statistical decision that needs to be made is whether we believe our data came from the null distribution or not. There are two possible outcomes:

1. If it is likely that your data could come from the null distribution, then "*we fail to reject the null hypothesis*."
2. If it is unlikely that your data could come from the null distribution, then "*we reject the null hypothesis*."

There are some important points about this statistical decision. The first is that it is a statement about the null hypothesis and only the hypothesis.

The second point is that we *reject* or *fail to reject* the null hypothesis, but never accept the null hypothesis. The reason for this is that while we can evaluate whether our data is consistent with the null distribution, we can never be sure it actually came from the null hypothesis.

A formal hypothesis test is done using two probabilities. The first probability is the Type I error rate, which is often called alpha(⍺). The Type I error rate is the probability of rejecting the null hypothesis when it is true. We use the Type I error rate to decide the point where the data are sufficiently far from the null hypothesis that it should be rejected. Importantly, the Type I error rate is set by the researcher without any reference to the data. It is common to set the Type I error rate at 5% (⍺= 0.05) but other values are possible as well.

The second probability is called the p*-value* (p). The p*-value* is the probability of seeing your data, or something more extreme, under the null hypothesis. Like the Type I *error* rate, it is a probability under the null distribution. However, unlike the Type I error rate the p*-value* is determined by the data. Specifically, the p*-value* is the area under the curve from the data to more extreme values.

So why study the Type II error rateif it's not possible to calculate it? We study it because there is a tradeoff between the two error rates. As the Type I error rate decreases, the Type II increases, and vice versa. To see how the two are related, we need to create a rather abstract figure that shows a null distribution alongside an imaginary alternative distribution as shown below. The figure has a vertical dashed line (at *t*=*-10*) that denotes the mean of a hypothetical sample. Since the mean falls under both distributions in this illustration, there is no way to know whether it was generated from the null or the alternative distribution. The Type I error rate is the area under the null distribution from the data point to something more extreme (shaded blue). The Type II error rate would be the area under the alternative distribution from the data point to something more extreme (shaded red).

