# Overview

Hypothesis testing is what we do when we have an assumption about some data. We will then test this hypothesis and either accept or reject some null hypothesis depending on the results of our statistical test.

# Statistical Power

## Definition

Statistical power is the probability that we will successfully reject the null hypothesis. A lot of times in stats we want to reject some null hypothesis that two samples come from the same distribution (think placebo vs drug medical results). Power is the probability that if these two distributions are in fact different, that we will correctly classify them as so.

## Power Analysis

This is a method to determine the effective sample size we should use for our hypothesis test. We typically try and have a power of 0.8 (the range is 0 – 1). This means we want an 80% probability we will correctly reject the null hypothesis.

One method to write this is like so:

# Central Limit Theorem

## Definition

The central limit theorem says that sample means from a distribution will be normally distributed. This would allow us to not know what the underlying population distribution is. A common rule of thumb is that the sample size must be at least 30 for the central limit theorem to work.

# Standard Error

## Definition

Standard error is the standard deviation of the mean of means. If we take a measurement to be a sample of say 10, we will take measurements 10 times. If we take the mean of each measurement then we have 10 means, we can show the mean of these means and the standard error which is the standard deviation of these means.

Confidence intervals encapsules x% of the means. This means a 90% CI would encapsulate 9 of 10 of the given measurement means.

# Bootstrapping

## Definition

This is the idea that if we have a dataset of N measurements, we can sample with replacement N samples from the original dataset to get the bootstrapped dataset. We can take the means of the samples and get a histogram of these means. This helps us estimate the full distribution using Central Limit Theorem. This gives an idea for what would happen if we redid the experiment.

## p-values

We can shift the values to have a mean of zero and bootstrap the data from this point. This would likely only work for normally distributed data I think (this is likely not the case). We can then get the histogram of the bootstrapped data which has been scaled to have a mean of zero.

This would give us the histogram and probabilities of observed mean if something had on average no effect, and the null hypothesis was true. If our actual mean was something like 0.5 on the non-scaled data, we could take the probability that the bootstrapped samples from the null hypothesis were >= 0.5 + <= -0.5 (the most extreme cases). If the p value (total of those probabilities) is less than 0.05 we could reject the null hypothesis.

We can do the same with medians meaning we can use non-parametric data. Or data with outliers which makes this easier.

You could also probably do a bootstrapped mean or medians to get the normally distributed data. We could then do a t-test or a z-test depending on the number of samples (likely a t-test) (use t-test if the variance is unknown or the sample size is smaller). I think the other method might be more rigorous because we transform the data to the null hypothesis.

# Sample Size

## Definition

The number of data points in each sample of the data. Not all samples are created equal though. Say we have correlated samples then we use. This means that we don’t over value correlated samples that do not really add new data points. Samples that are highly correlated don’t count as individual samples.

# Two-Tailed P-Values

## Definition

The two tailed P-value analyzes above, lower, or no statistical difference. Since we pick our test and what we will do before seeing the data we will normally use the two tailed P-test. This lets us compare the data in a more reasonable manner because there are three possibilities. We would only use the 1-tailed p-value when it only makes sense to have one side.

We can use the 1-sided test if we are completely uninterested in the case that group A is lower than group B. This works for some types of analysis because it will test if one distribution is above another. However, in most cases we would still want to consider the possibility that the A is lower.

The two-tailed test can be used to look at the probability overlap on both directions. This splits the alpha onto both sides.

# Binomial Distribution and Test

## Distribution