The two resampling methods that we have met are bootstrapping and permutation tests. In this handout, we’ll review these two methods and discuss the contexts in which you should be able to use them.

**Bootstrapping**

You can use bootstrapping when you would like an approximation of the distribution of a test statistic. (Remember that test statistics vary over \( n \)-samples, so in bootstrapping you are approximating the distribution of the test statistic over all possible \( n \)-samples.) You can use such an approximation to conduct hypothesis tests, but in this class we will focus on bootstrapping only to compute confidence intervals. The process of bootstrapping is:

- repeatedly resample (\( N \) times) *WITH REPLACEMENT*, meaning repeatedly gather samples of size \( n \) from your sample, with replacement.
- for each resample, compute the value of the test statistic.

The distribution of the collection of test statistic values (the “resampling distribution”) that you assemble this way is an approximation of the actual (theoretical) distribution of the test statistic, and you can now use it to compute a 95% confidence interval. In the simplest approach to this, the percentile method, the endpoints of a 95% confidence interval are given by the endpoints of the middle 95% of the resampling distribution, which are the 0.025-quantile and the 0.975-quantile. (For a level \( c \) confidence interval, use the middle \( c \) of the resampling distribution instead.) However, the percentile method isn’t as accurate as some other methods, such as the BCa (bias-controlled and accelerated) method of computing a bootstrap confidence interval. In this method, the sample is transformed so as to be approximately normally distributed, a percentile bootstrap is computed on the transformed sample, and then the results are transformed back. This can be carried out using the `boot.ci()` function in the `boot` package in R.

If you’re wondering how many resamples to generate (how big \( N \) should be), a good starting point is usually 1000. Then if you see anything that indicates a larger \( N \) would help, you can up it to 10,000 and then 100,000 or more if need be. (There aren’t specific things to look for on this, but in some cases you may notice oddities about the approximate distribution that you calculated. Or you might just try the next power of 10 to see if the results change any.)
Permutation tests

You can use a permutation test when you would like to conduct a hypothesis test whose null hypothesis is that the distribution of the test statistic is the same for two groups that you are comparing. (Of course, if the null hypothesis is that the two groups come from the same population, that is suitable for a permutation test, but it is actually a stronger null hypothesis than is needed for such a test.) In this case, each permutation of the labels on the two groups in your sample is equally probable under the null hypothesis.

We haven’t learned how to use a permutation test to compute a confidence interval; it is for hypothesis tests only.

The process of conducting a permutation test is:

- repeatedly permute your sample \( (N \text{ times}) \), meaning repeatedly gather samples of size \( n \) from your sample WITHOUT REPLACEMENT
- for each permutated sample, compute the value of the test statistic
- the proportion of these values that lie in the rejection region corresponding to the test statistic for your sample is an estimator of the probability that the test statistic would land in that rejection region, which is the \( p \)-value. Using this, compute the estimate of this probability (namely \( p \), the proportion you got) and a 95% confidence interval for the probability, using the usual formula for the confidence interval of a probability, which is

\[
p \pm z_{.05} \sqrt{\frac{p(1-p)}{N}}
\]

Make sure that you choose \( N \) large enough so that your entire confidence interval is on one side or another of 0.05. (If not, try again with a larger \( N \). Also, if you have a \( p \)-value of 0, you need a larger \( N \).) Otherwise, you won’t be able to assert as strongly whether or not you found statistically significant evidence against the null hypothesis.

If you were to compute all possible permutations of your sample, the resulting collection of test statistics would be the exact permutation distribution of the test statistic, meaning the exact distribution of the test statistic as it varies over all possible permutations of your particular sample. Notice that this is a different type of test statistic distribution from what you have considered in other hypothesis tests so far. Here you are not considering a theoretical distribution of a test statistic over “all
possible \( n \)-samples”, but rather an actual distribution of a test statistic over all possible permutations of your \( n \)-sample, which has nothing theoretical about it.

Because it isn’t feasible to compute all permutations of your sample (except when you have a tiny \( n \)), you instead compute a random selection of those permutations and use that to obtain an estimate and a confidence interval for the \( p \)-value.

The alternative hypothesis is the trickiest part of a permutation test, and it is generally better to focus on determining a rejection region within the test statistic’s permutation distribution. If you have an alternative hypothesis in mind, then the rejection region corresponding to a particular test statistic value is the region that is “further from the null hypothesis in the direction of the alternative hypothesis”. This will be different depending on what you choose as your test statistic (such as the median of Group A versus the median of Group B), so there is no formula for it and it requires careful attention.

**How have we used these methods?**
We have used bootstrapping to estimate mean, median, first and third quartiles, and correlation, and you should be prepared to bootstrap any parameter that we have an estimator for.

We have used permutation tests to test null hypotheses that: the means of two populations are equal (a nonparametric alternative to a 2-sample \( t \)-test), the medians of two populations are equal, two categorial random variables are independent (in the guise of Fisher’s exact test, a nonparametric alternative to a chi-square independence test), and two random variables have correlation 0 (a nonparametric alternative to a \( t \)-distribution based test of that). In addition to being able to conduct these permutation tests (all of them manually with your own code except Fisher’s exact test) and their parametric alternatives, you should be prepared to conduct a permutation test whenever the null hypothesis implies that the distribution of the test statistic is the same for the two groups being compared.