Permutation Test Overview and Results

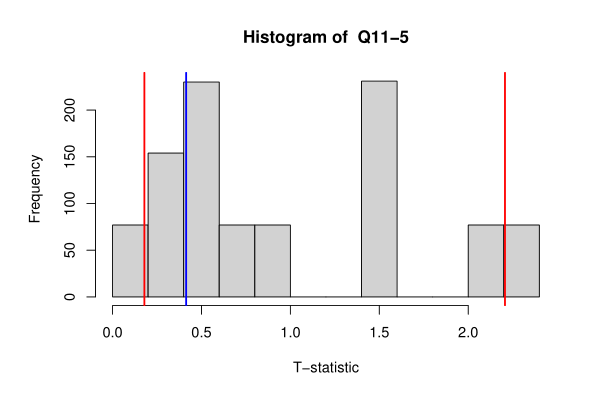
**What is a permutation test?**

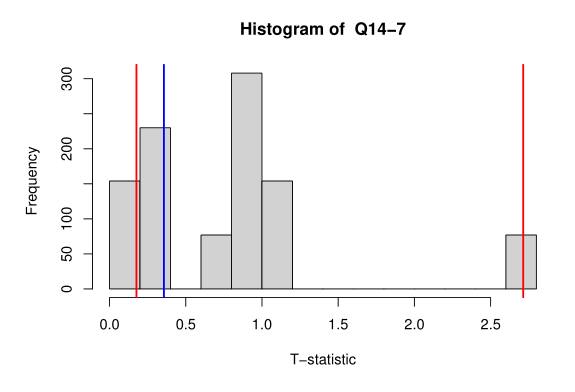
* Also known as the randomization test or exact test, the permutation test allows us to test for common hypotheses (like t-tests) without any assumptions about the distribution of the data. Additionally, the permutation test is valid for small sample sizes, including samples that are < 30. This is significant for the teaching and learning field since these two attributes are often applicable to survey data from small classrooms.

**How does it work?**

* In our case, we are interested in the difference between the student and supervisor responses for each survey question. For our hypothesis test we assume that the null or base hypothesis is that the average survey response is the same between the students and supervisors. Then, for the alternative or contrasting hypothesis we assume that the average survey response is different between the students and supervisor. To start, we will calculate the t-statistic (which is used when comparing averages) and this will be our initial statistic.
* Now here is the intuition behind the mind-blowing part: under the null hypothesis we say that the survey responses are the same, *so there is hypothetically no distinction between the student and supervisor responses*. Therefore, if there really is no distinction between the student and supervisor responses, we can **permute** the responses between the two labels: student and supervisor. If the null hypothesis really is true, the shuffling won’t matter, and we will get the same differences and t-statistics as our initial test statistic. If our initial test statistic is significantly different than the permuted values (or simulations) then we have reason to believe that the survey responses are significantly different (accept alternative hypothesis.

(If you are still confused about this, here is another great visual explanation: <https://www.jwilber.me/permutationtest/>)

In our case, we permuted the responses 1000 times and plotted the results onto a histogram. Here are two examples from the data below, let’s break this down.



Red line on the left = 10th quantile

Red line on the right = 90th quantile

Blue Line = Initial T-statistic (without any permutation)

In essence, we permuted the response and calculated the t-statistic 1000 times. The histograms above show the t-statistics in the 1000 permutations for Q11-5 and Q14-7. In both graphs, we see that our initial statistic falls within the 10th and 90th quantiles which means that it is like the permutated simulations. This means that the average response between students and supervisors is not significantly different for this question. We can also view this from a p-value perspective, which is the percent of simulated responses that fall above our initial statistic. (see table below)

**In the end, none of the questions had significant differences between average response. But now we have statistical backing for this statement.**

I’ll note that Q14-8 has the smallest p-value, and I can agree with considering that as an acceptable value. However, looking at the raw data, there are a lot of blanks and I suspect the p-value is being affected by the only 2 in the responses. Therefore, I would not consider this variable important.