Problem Set 01

Aaron Graybill
Collaborators: John Cho, Thomas Li, Blair Moore

2025-09-28

Contents

If you're having trouble running my code, see https://github.com/aarongraybill/MTGECON_603

1 Problem 1:

1.1 Problem 1.a:

The test statistic from the 100,000 draws of the randomization distribution is 1.14. This result is extremely significant as none of the 100,000 draws produced a value of the test statistic as extreme as the test statistic from the true assignments. Numerically, this means that the p-value is 0, although with limitless computation, we would expect to see an extremely small, but positive value of the test statistic if we could fully compute each possible assignment vector. We can firmly reject the null hypothesis that the treatment has no effect.

Figure 1 summarizes this graphically. The value of the test statistic from the actual assignments (the vertical line) is well outside the values from the randomization distribution. I decided on 100,000 draws from the randomization distribution because it is a large enough number to explore many of the possible assignment vectors while still being small enough to be easily computed without requiring additional computational infrastructure.

1.2 Problem 1.b

The previous exercise uses the difference in means. Figure 2 reports the test statistic and randomization distribution for the difference in medians between the treatment and control groups. Here, the p-value is non-zero because there are a few (negative) values from the randomization distribution that are more extreme than the observed test statistic of 0.02. The p-value is still 0.01649, which is still strongly significant, so the treatment does appear to have some impact on the difference in medians.

That said, I feel the difference in medians is an inferior test statistic in this case due to the prevalence of zeros in the distribution of incomes. In the data, the fraction of zero incomes after one year is 0.5274036. This means that in many of the draws from the randomization have a median of zero for both the treatment and control groups. While there's nothing inherently wrong about that, this choice of test statistics does not capture the interesting variation in the data.

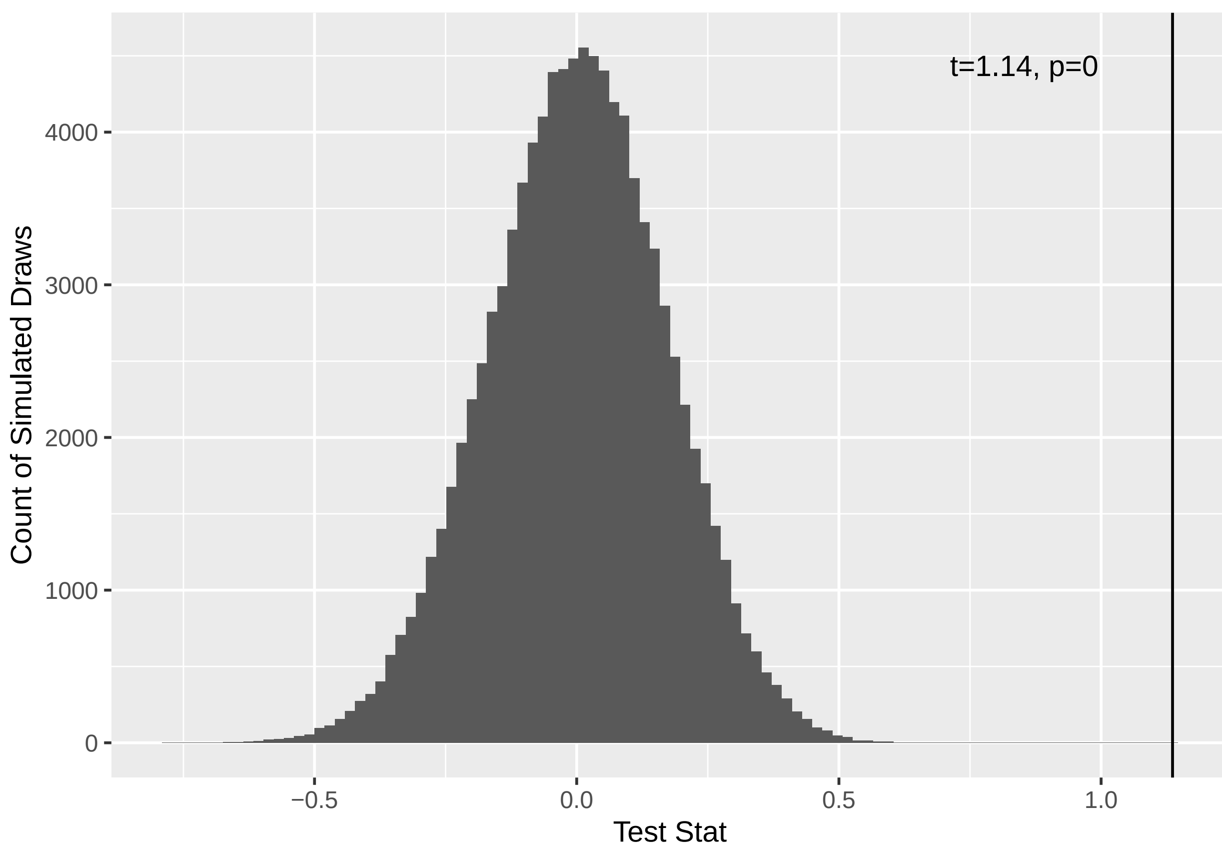

Figure 1: Randomization distribution and test statistic for difference in means.

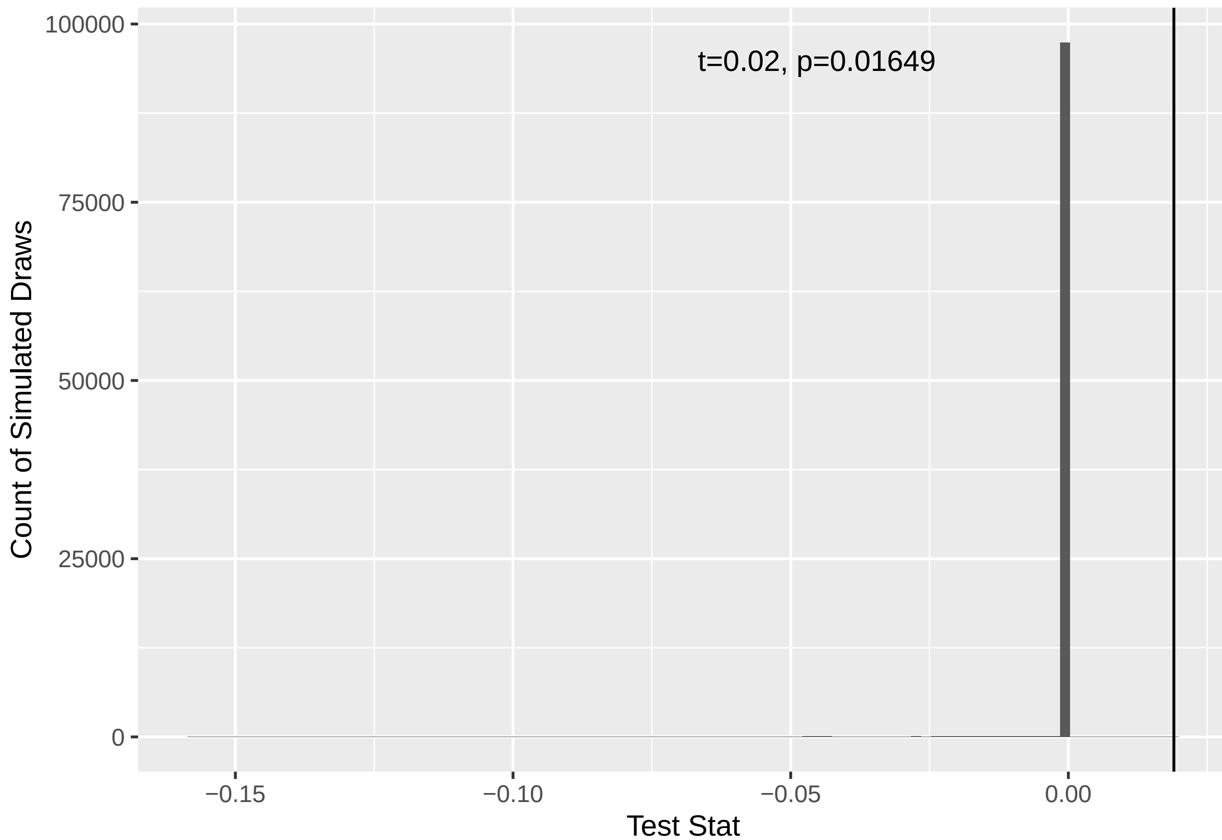

Figure 2: Randomization distribution and test statistic for difference in medians.

1.3 Problem 1.c:

An additional test statistic I found personally interesting was the difference in the share of zero-income individuals between treated and untreated groups. Because of the aforementioned mass at zero, one might worry that the treatment doesn't actually change the share of individuals making positive income and that it only changes income for those who would be employed either way. In particular, I define my test statistic to be:

$$\frac{\sum_{Y_i=T} \mathbf{1}(earnings > 0)}{\sum_{Y_i=T} 1} - \frac{\sum_{Y_i=C} \mathbf{1}(earnings > 0)}{\sum_{Y_i=C} 1}$$

Once again the test statistic is extremely significant, with $t = 0.158$ and a p -value equal to approximately zero. Again there were no draws from the randomization distribution in which the value was more extreme that the observed value of the test statistic. This is readily apparent in figure 3

This provides strong evidence that the treatment had an effect on the fraction of individuals participating in the labor force.

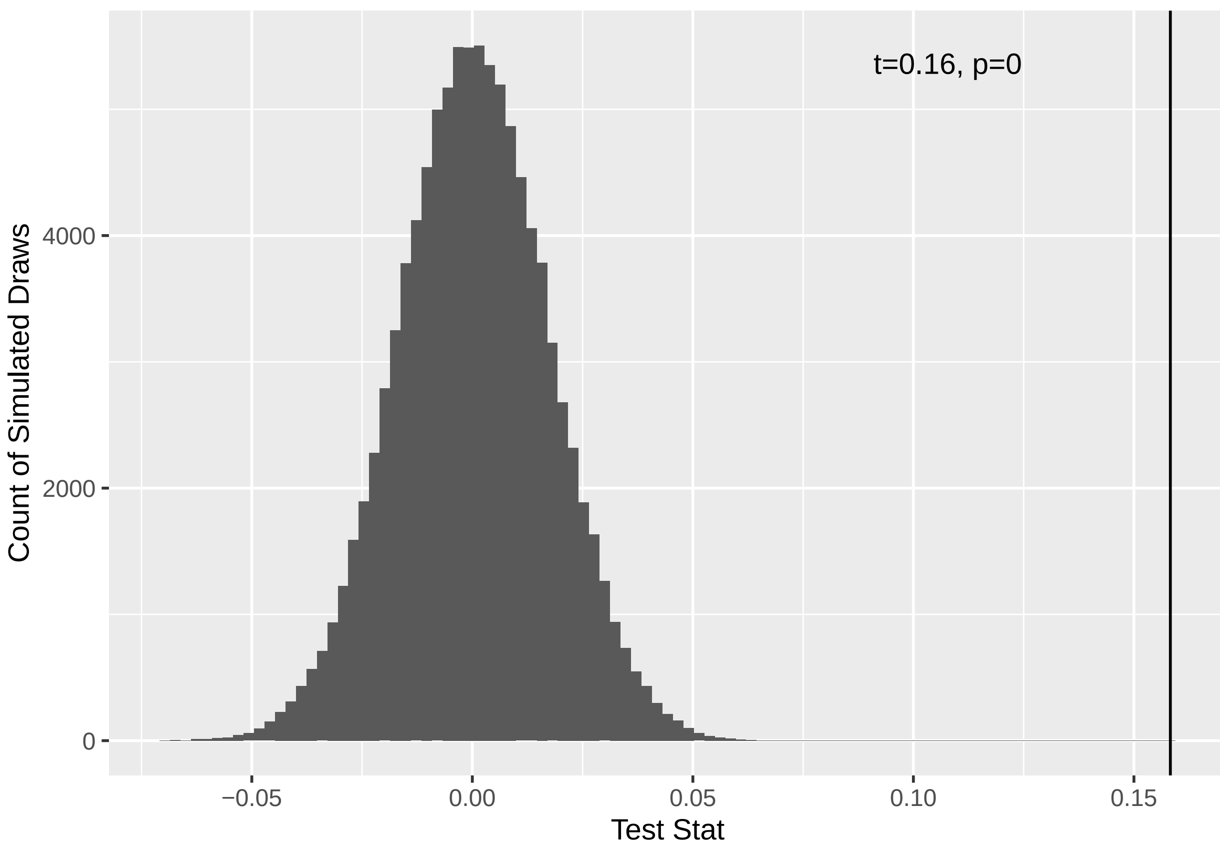

Figure 3: Randomization distribution and test statistic for difference in fractions greater than zero.

1.4 Problem 1.d

Table 1 summarizes the p-values from the fixed-M versus various bernoulli assignment probabilities. The question asks about $p = .8$, but to better illustrate the analysis I also report the results from additional simulations where $p = .5$ and $p = .99$. In each of those cases, I take p to be the probability of assignment to the *treament* group. I choose this because the empirical fraction of treated is 0.81, so $p = .8$ closely matches

Table 1: Table of p-Values from various Fisher Exact Tests. Fixed M corresponds to the p-values from a fixed number of treated units. Columns with p= correspond to bernoulli simulations where the number of treated units varies randomly according to the stated probabilities. Rows correspond to the test statistic in question.

Test Stat	Fixed M	p=0.5	p=0.8	p=0.99
mean_diff	0.00000	0	0.00000	0.08716
median_diff	0.01649	0	0.01376	0.34656
frac_above_zero	0.00000	0	0.00000	0.02101

the empirical fraction of treated units. As the table shows, the p-values from the fixed-M versus $p = .8$ are extremely similar. That's because the bernoulli assignments with similar balance closely mirror the fixed-M case when there is such a large number of units.

That said, the p-values are weakly lower (more signifcant) in the $p = .8$ column than under the fixed-M computation. I believe $p = .8$ being slightly less than the empirical 0.81 means that the bernoulli trial is slightly more balanced (closer to 0.5), and therefore has greater power, so a lower p-value.

To further illustrate that point, I included the other columns for bernoulli assignments when $p = .5$ and $p = .99$. $p = .5$ maximizes power and therefore has the lowest p-values. $p = .99$ is extremely unbalanced and is unable to provide enough power to reject the null under most of the test statistics. This shows that it may matter to use the bernoulli randomization distribution when the bernoulli assignment probability is quite different from the fixed-M empirical fraction or when the number of units is quite low. Neither of those cases apply when $p = .8$ with more than 5,000 units, so there is not a strong (quanitative) reason to use one over another in our case.

2 Problem 2:

In this case, I would suggest treating the experiment as though the 49 treated units was non-random, despite the coin flipping. The outcome of 49 treated units was extremely typical after flipping a coin 100 times, and there are many possible assignment vectors with 49 units treated, so the randomization distribution should be able to capture lots of possible scenarios. While the Bernoulli $p = .5$ is slightly more balanced than the observed $M = 49/100$ and may result in slightly lower p-values, this result is likely to be neglible given how close 49 is to the maximally balanced 50. Furthermore, with 100 total units, there are probably enough units for Bernoulli assignments to not play a meaningful role in the total number of assignments.

One might consider using the Bernoulli assignments if the number of units was smaller and number of treated individuals is extremely unbalanced, then the bernoulli assignments might increase the power and more accurately capture the variation.