

# SDS315\_HW9

Neil Sebastian  
2024-04-16

## HW9

UT EID: ns36965

[GitHub Link](#)

### Problem 1: Get out the vote

#### Part A

How much more likely are GOTV call recipients to have voted in 1998? As a preliminary analysis, calculate the following quantities.

- The proportion of those receiving a GOTV call who voted in 1998.
- The sample proportion of those not receiving a GOTV call who voted in 1998.
- A large-sample 95% confidence interval for the difference in these two proportions: that is, the proportions of voting in 1998 ( $\text{voted1998}=1$ ) for those who received a GOTV call versus those who didn't.

```
## n_1
## 0.648

## n_0
## 0.444
```

The numbers above represent that the proportion of those who recieved a call and then voted is about 65%. Whereas those who did not recieve a call and voted was about 45%.

Next using bootstrapping we have calculated the difference in proportions on a 95% confidence interval.

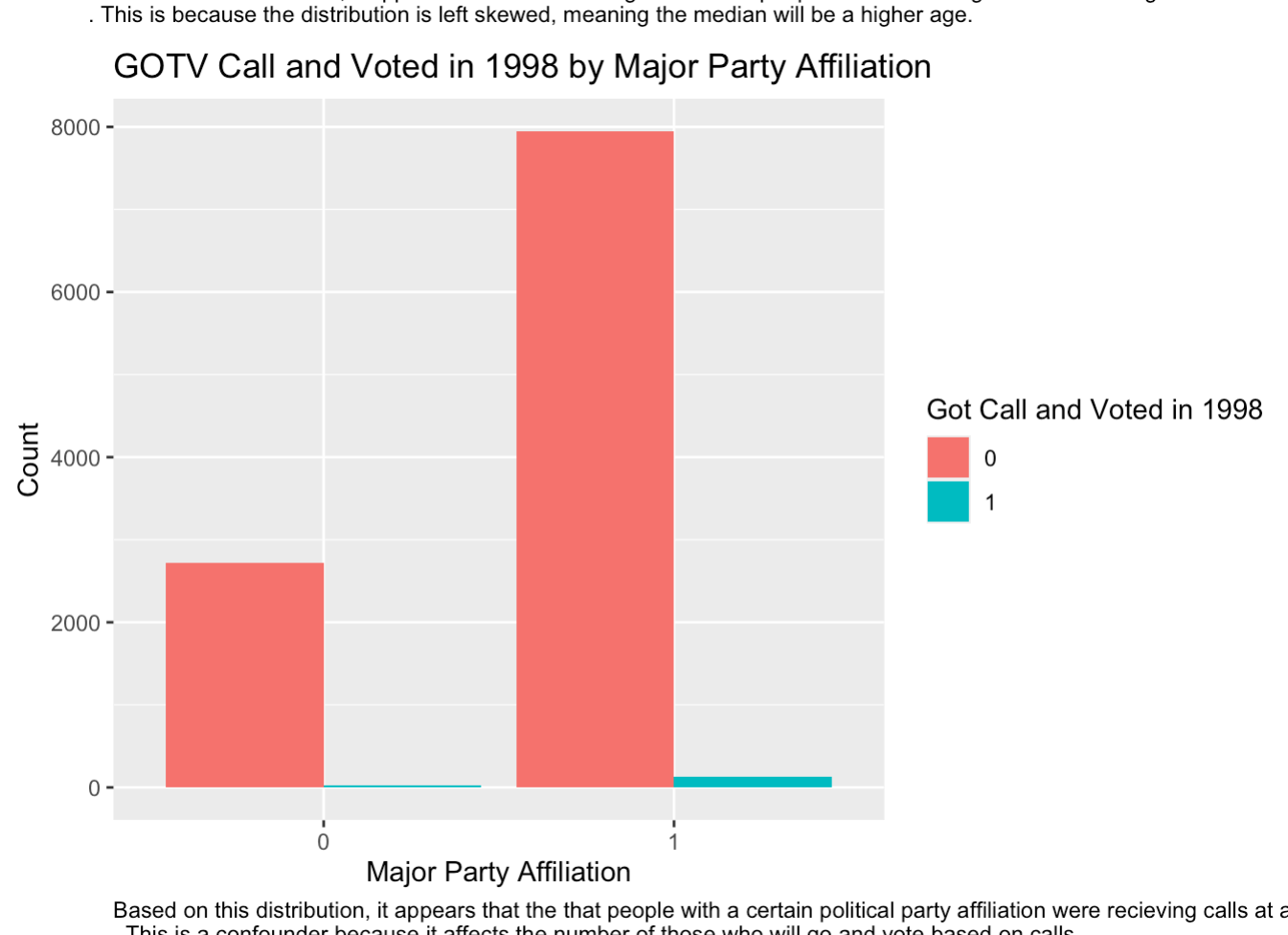
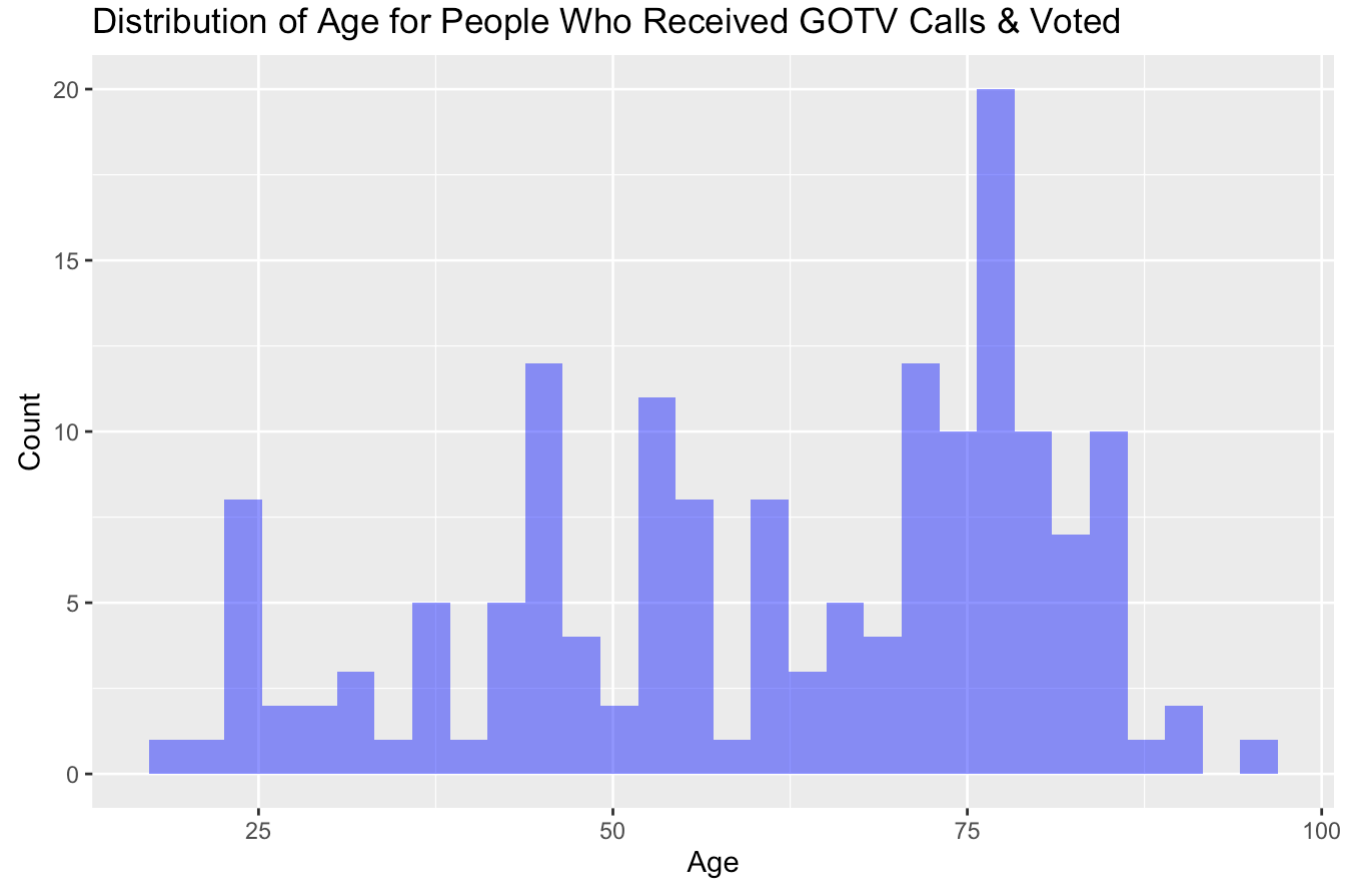
```
## Estimate for the Difference in Proportions: 0.2035 - 0.2035
```

#### Part B

Part B. Consider the `voted1996`, `AGE`, and `MAJORPTY` variables. Provide evidence that at all three of these variables are confounders that prevent the difference you observed in Part A from representing the true causal effect of the GOTV call on the likelihood that a person voted in 1998. Confounders here would be factors that make someone more likely to receive a GOTV call and to have voted in 1998. Your evidence here can consist of any appropriate plot, table, or set of summary statistics.

```
## # A tibble: 2 x 2
##   voted1996 proportion_voted_1998
##   <int>         <dbl>
## 1       0         0.09457
## 2       1         0.0236

## [1] "Based on these results we can see that people who voted in 1996 also voted in 1998 at a higher proportion
than those who did not vote in 1996 but voted in 1998."
```



#### Part C

```
##
## Call:
## matchit(formula = GOTV_call ~ voted1996 + AGE + MAJORPTY, data = turnout,
##   ratio = 5)
##
## Summary of Balance for All Data:
##      Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance      0.0297      0.0226      0.5130      1.3026      0.1572
## voted1996      0.7126      0.5308      0.4016      .      0.1817
## AGE      58.3077      49.4253      0.4475      1.1228      0.1114
## MAJORPTY      0.8016      0.7448      0.1426      .      0.0569
##
##      eCDF Max
## distance      0.2499
## voted1996      0.1817
## AGE      0.2229
## MAJORPTY      0.0569
##
## Summary of Balance for Matched Data:
##      Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance      0.0297      0.0297      0.0001      1.004      0.0000
## voted1996      0.7126      0.7126      -0.0000      .      0.0000
## AGE      58.3077      58.2664      0.0021      1.008      0.0006
## MAJORPTY      0.8016      0.8073      -0.0142      .      0.0057
##
##      eCDF Max Std. Pair Dist.
## distance      0.0057      0.0001
## voted1996      0.0000      0.0000
## AGE      0.0057      0.0027
## MAJORPTY      0.0057      0.0183
##
## Sample Sizes:
##      Control Treated
## All      10582      247
## Matched      1235      247
## Unmatched      9347      0
## Discarded      0      0
```

The summary above proves that the matching was effective because the differences between the means of the previously confounding variables are now approximately the same. This means that the matching was effective, and the variables have been balanced.

```
## n_1
## 0.6477733

## n_0
## 0.5740891
```

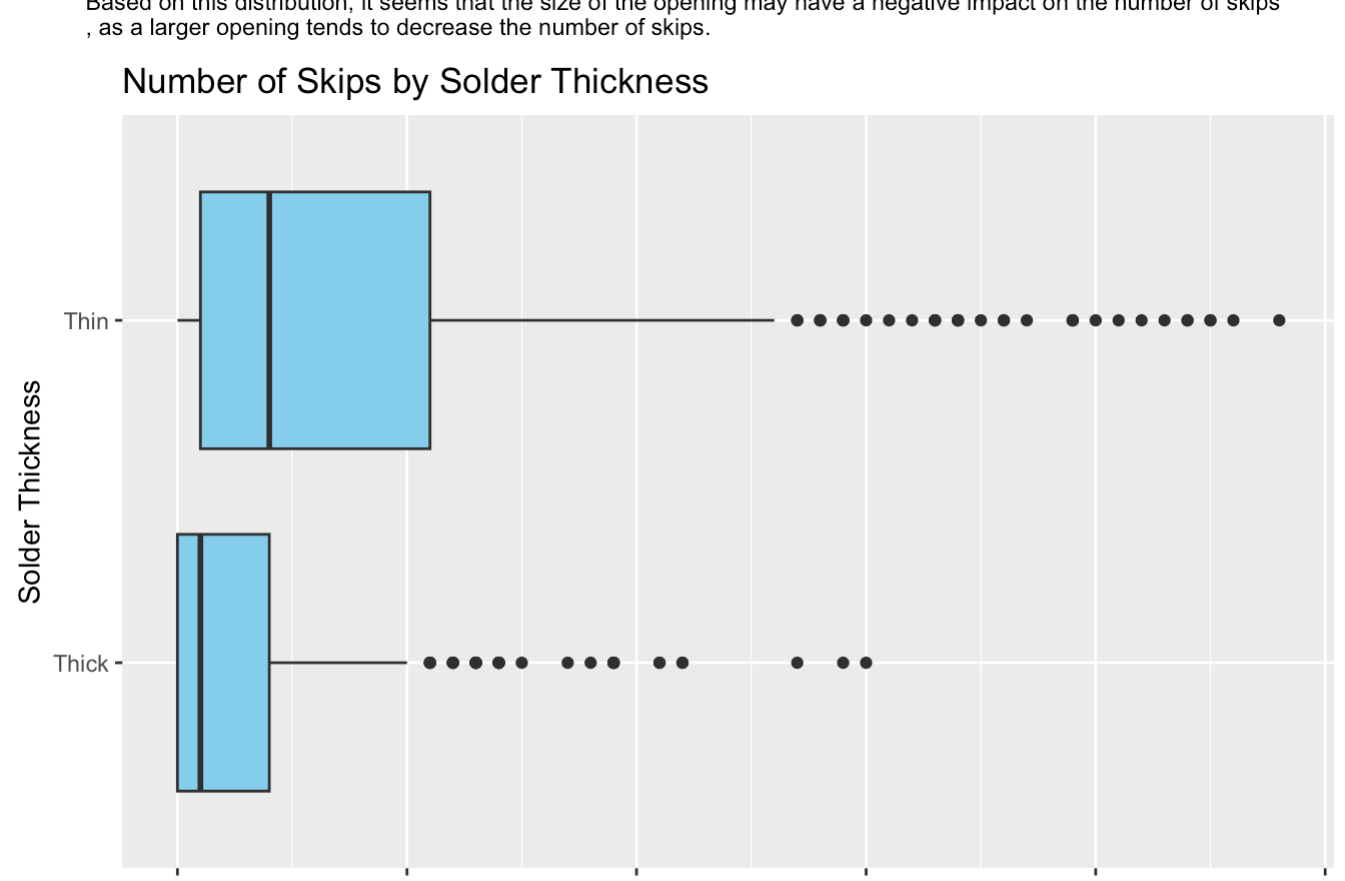
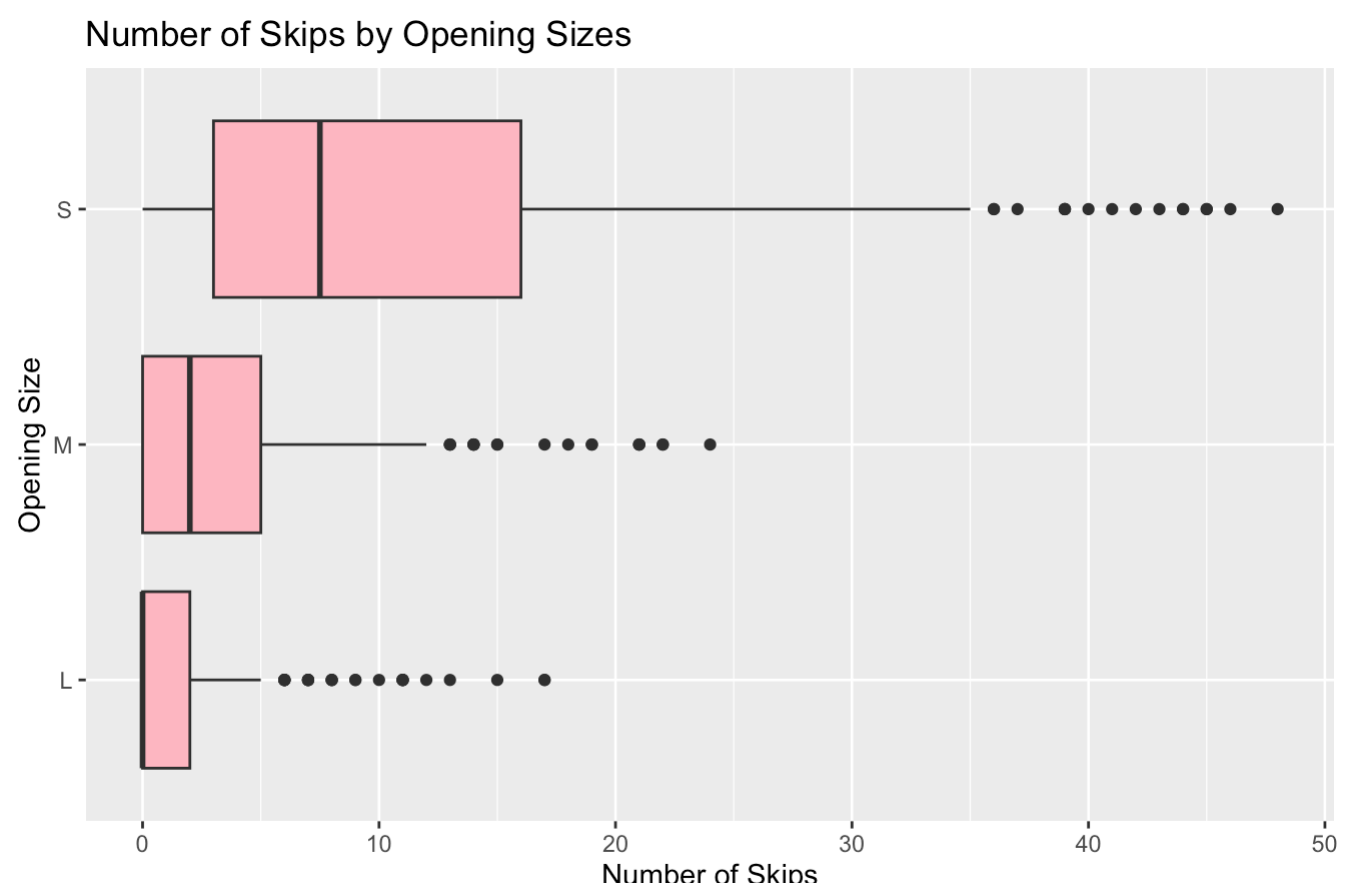
```
## Estimate for the Difference in Matched Proportions: 0.0737 - 0.0737
```

Based on the data, it appears that the GOTV call did indeed play a role in boosting voter turnout during the 1998 election. However, it's important to note that the significant difference in voter turnout initially observed was largely influenced by other factors, such as age, party affiliation, and past participation in elections, which were accounted for through matching. The initial estimate suggested a difference of 20.35% between those who received the call and those who didn't. However, after accounting for confounding variables through matching, the difference attributable solely to the GOTV call reduced to around 7.37%. This indicates that while the GOTV call was effective, its impact was influenced by the other factors.

### Problem 2: Manufacturing flaws in circuit boards

#### Part A:

Make two plots. The first plot should provide evidence that the size of the opening on the solder gun is related to the number of skips. The second should provide evidence that the thickness of the alloy used for soldering is related to the number of skips. Give each plot an informative caption describing what is shown in the plot.



#### Part B

Consider the `voted1996`, `AGE`, and `MAJORPTY` variables. Provide evidence that at all three of these variables are confounders that prevent the difference you observed in Part A from representing the true causal effect of the GOTV call on the likelihood that a person voted in 1998. Confounders here would be factors that make someone more likely to receive a GOTV call and to have voted in 1998. Your evidence here can consist of any appropriate plot, table, or set of summary statistics.

```
##      name      lower      upper level      method      estimate
## 1 Thick + L (Intercept) 0.2301480 0.5862184 0.95 percentile 0.3933333
## 2 Thin + L      1.7181875 2.8812848 0.95 percentile 2.2800000
## 3 Thick + M      1.7276824 3.1695801 0.95 percentile 2.4066667
## 4 Thick + S      4.1574177 6.1585321 0.95 percentile 5.1266667
## 5 Thin + M      -1.9493932 0.4711663 0.95 percentile -0.7400000
## 6 Thin + S      7.4132892 11.9054109 0.95 percentile 9.6533333
## 7 Sigma      5.8192213 6.8691987 0.95 percentile 6.3729241
## 8 R^2      0.3794641 0.4919322 0.95 percentile 0.4339639
## 9 F      109.3380611 173.1215183 0.95 percentile 137.0809061
```

#### Part C

Interpret each estimated coefficient in your model in no more than 1-2 sentences. A good template here is provided in the course packet, when we fit a model for the video games data that had an interaction in it and interpreted each coefficient in a sentence or two.

- When soldering is Thick and the opening is Large, there is an estimated rate of change of .39 skips.
- When soldering is Thin and the opening Large, there is an estimated 2.28 MORE skips as compared to the intercept of .39 skips. This indicates a change of +2.67 skips in this case.
- When soldering is Thick and the opening Medium, there is an estimated 2.41 MORE skips as compared to the intercept of .39 skips. This indicates a change of +2.80 skips in this case.
- When soldering is Thick and the opening Small, there is an estimated 5.13 MORE skips as compared to the intercept of .39 skips. This indicates a change of +5.52 skips in this case.
- When soldering is Thin and the opening Medium, there is an estimated 0.74 LESS skips as compared to the intercept of .39 skips. This indicates a change of -.35 skips in this case.
- When soldering is Thin and the opening Small, there is an estimated 9.65 MORE skips as compared to the intercept of .39 skips. This indicates a change of +10.04 skips in this case.

#### Part D

If you had to recommend a combination of Opening size and Solder thickness to AT&T based on this analysis, which one would it be, and why? (Remember, the goal is to minimize the number of skips in the manufacturing process.)

With the goal in mind to minimize skips in the manufacturing process, it would be most effective to use a thin solder with a medium opening. This is because it is the only case where there is an estimated LESS skips compared to the intercept (.74 skips less). Because it has this negative effect it is likely that you will minimize skips in the manufacturing process.