SDS315_HW9

Juhi Malwade

2024-04-17

Github Link

Problem 1: Get out to vote

Part A: How much more likely are GOTV call recipients to have voted in 1998?

The proportion of those receiving a GOTV call who voted in 1998 is 0.444. The proportion of those not receiving a GOTV call who voted in 1998 is 0.648. The 95% confidence interval for the difference in these two proportions is 0.143 to 0.264.

Part B: Considering confounders

Evidence voting in 1996 is a confounding variable

Table 1: 95% Confidence Interval

name	lower	upper	level	method	estimate
diffprop	0.3933262	0.4273586	0.95	percentile	0.4103889

The 95% confidence interval displayed above represents the difference in proportions of 1998 voters who did and did not vote in the 1996 election. The interval is entirely positive and does not contain zero, indicating we are 95% confident that the proportion of 1998 voters among those who voted in 1996 is greater than the proportion of 1998 voters among those who did not vote in 1996. This indicates that voting in 1996 is a confounding variable that can make someone more likely to have voted in 1998.

Table 2: 95% Confidence Interval

name	lower	upper	level	method	estimate
diffprop	0.0107695	0.0216859	0.95	percentile	0.016283

The second 95% confidence interval above displays the difference in proportion of receiving a call among those who did and did not vote in the 1996 election. Because the confidence interval is entirely positive and does not contain zero, it indicates that prior 1996 voters were more likely to receive a call in 1998 encouraging them to vote again. Thus, voting in 1998 may not be solely caused by receiving a call, since receiving a call is also correlated to whether or not someone voted in 1996. This strengthens the claim that voting in 1996 is a confounding variable.

Evidence age is a confounding variable

Table 3: 95% Confidence Interval

name	lower	upper	level	method	estimate
diffmean	9.815439	11.17123	0.95	percentile	10.50131

The 95% confidence interval above displays the mean difference in age among those who did and did not vote in the 1998 election. Because the confidence interval is entirely positive and does not contain zero, it indicates that we are 95% confidence that the mean age of voters in the 1998 election is greater than the mean age of non-voters in the 1998 election. This indicates age is a confounding variable that can make someone more likely to have voted in 1998.

Table 4: 95% Confidence Interval

name	lower	upper	level	method	estimate
diffmean	6.397361	11.32118	0.95	percentile	8.882347

In addition, the 95% confidence interval above displays the mean difference in age among those who did and did not receive a call prior to the 1998 election. Because the confidence interval is entirely positive and does not contain zero, it indicates we are 95% confident that older people are more likely to receive a call than younger people. Thus, voting in 1998 may not be caused solely by receiving a call, since receiving a call is also correlated to how old the person is. This strengthens the claim that age is confounding variable.

Evidence major party affiliation is a confounding variable

Table 5: 95% Confidence Interval

name	lower	upper	level	method	estimate
diffprop	0.1112036	0.1525256	0.95	percentile	0.1323036

The 95% confidence interval above represents the difference in proportion of 1998 voters who were and were not affiliated with a major political party. Because the confidence interval is entirely positive and does not contain zero, it indicates we are 95% confident that the proportion of 1998 voters among those affiliated with a major political party is greater than the proportion of 1998 voters among those not affiliated with a major political party. This indicates that affiliation with a major political party is a confounding variable that can make someone more likely to have voted in 1998.

Table 6: 95% Confidence Interval

name	lower	upper	level	method	estimate
diffprop	0.0053579	0.1071338	0.95	percentile	0.0568642

The 95% confidence interval above represents the difference in proportion of being registered for a major party among those who did and did not receive a call. Because the confidence interval is entirely positive and does not contain zero, it indicates that we are 95% confident that major party affiliates were more likely to receive a call encouraging them to vote compared to non-major party affiliates. Thus, voting in 1998 may not be solely caused by receiving a call, since receiving a call is also correlated with whether or not someone is registered as a member of a major party.

Part C: Matching

```
##  prop_1.0  prop_1.1
## 0.7125506 0.7125506
```

The table above shows that voting in 1996 is no longer a confounder after matching because the proportion of those receiving a GOTV call who voted in 1996 is the same as the proportion of those not receiving a GOTV call who also voted in 1996. Thus, major party is no longer a confounder.

```
##        0        1
## 58.26640 58.30769
```

Similarly, after matching the mean age among those who did and did not receive a call were approximately the same. Thus, age is no longer a confounder.

```
##  prop_1.0  prop_1.1
## 0.8072874 0.8016194
```

And finally, after matching, the proportion of those receiving a GOTV call who registered for a major party was the same as the proportion of those not receiving a GOTV call who also registered for a major party. Thus, major party is no longer a confounder.

Table 7: 95% Confidence Interval

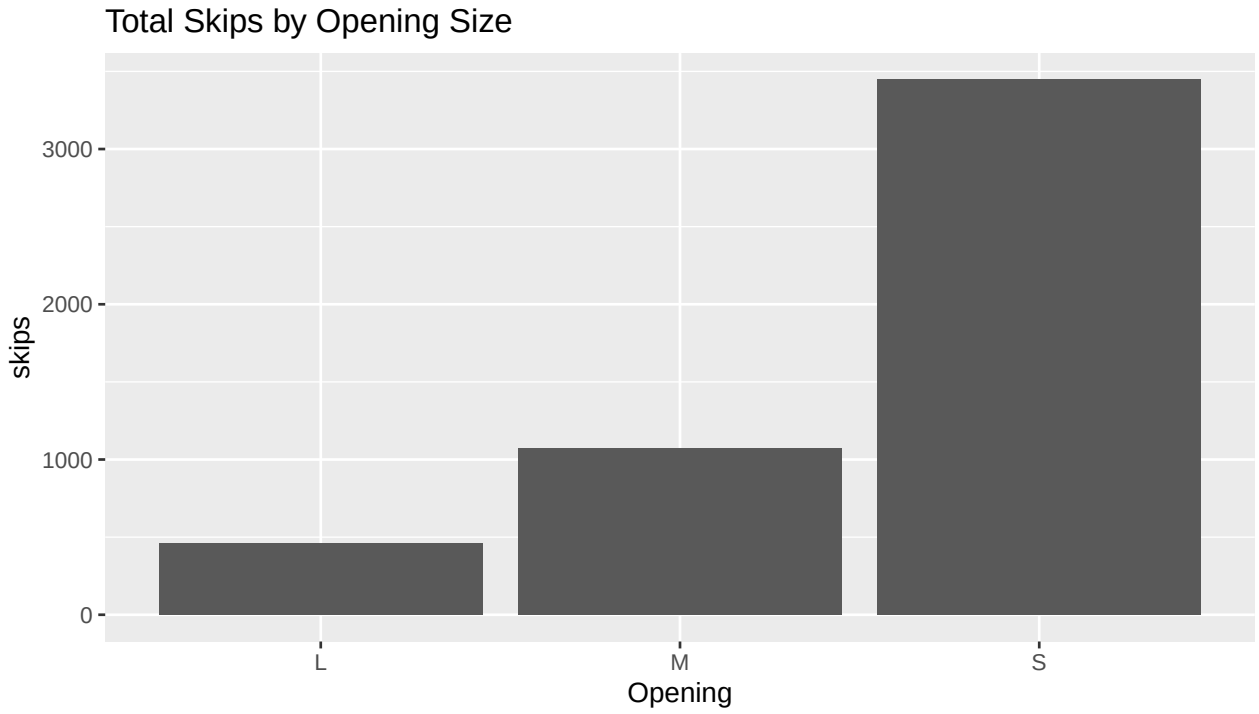
name	lower	upper	level	method	estimate
diffprop	0.007869	0.1400462	0.95	percentile	0.0736842

After matching and accounting for confounders, the proportion of those receiving a GOTV call who voted in 1998 is 0.574. The proportion of those not receiving a GOTV call who voted in 1998 is 0.648. The 95% confidence interval for the difference in these two proportions is 0.008 to 0.14.

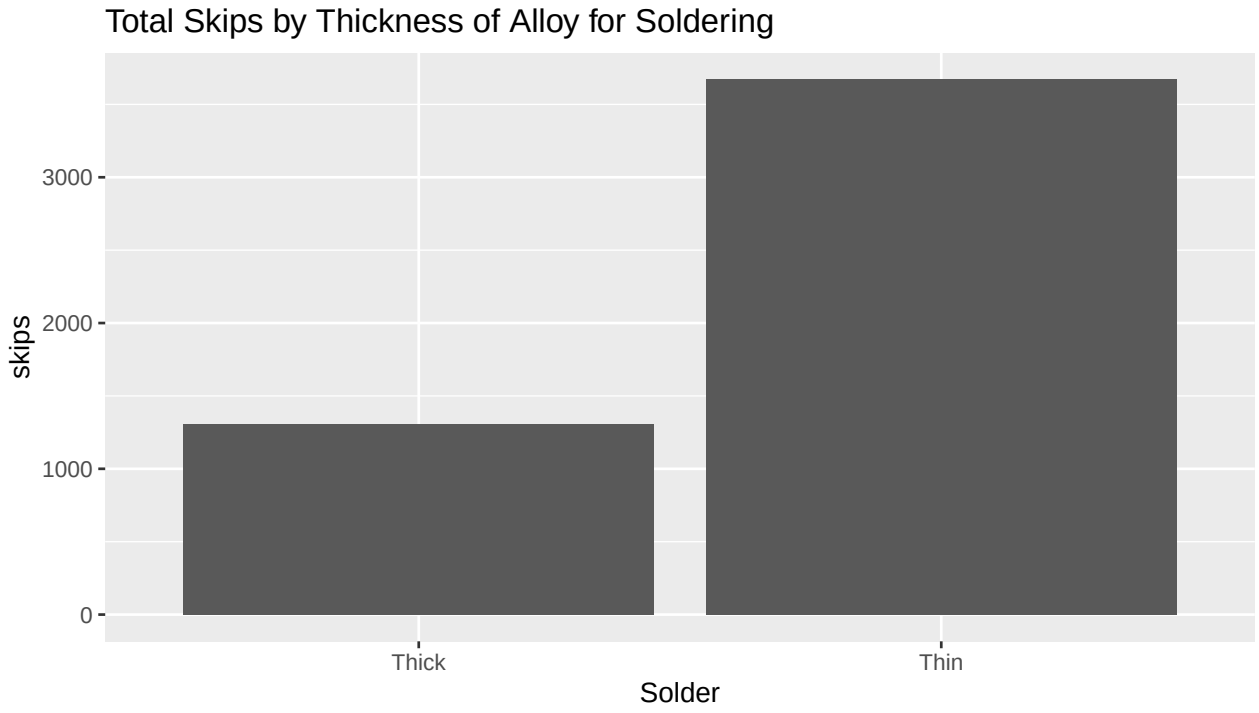
In conclusion, the GOTV call has a very small effect on the likelihood of voting in the 1998 election. This effect is significantly smaller than initially estimated after accounting for confounding variables.

Problem 2: Manufacturing flaws in circuit boards

Part A:



The plot above shows the relationship between opening size and total number of skips. The small opening size had the greatest number of skips (3446 skips), the medium opening size had 1071 skips, and the large opening had the least number of skips (460 skips). All three opening types are equally represented in the dataset (300 of each type). Thus, there is a relation between opening size and total number of skips.



The plot above shows the relationship between thickness of the alloy for soldering and total number of skips. The thin alloy had the more skips (3670 skips) than the thick alloy (1307 skips). Both thicknesses of allow are equally represented in the dataset (450 of each type). Thus, there is a relation between thickness of alloy and total number of skips.

Part B:

```
## # A tibble: 6 x 7
##   term                  estimate std_error statistic p_value lower_ci upper_ci
##   <chr>                    <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept                 0.39      0.52      0.76    0.45    -0.63     1.41
## 2 Opening: M                2.41      0.74      3.27    0        0.96     3.85
## 3 Opening: S                5.13      0.74      6.97    0        3.68     6.57
## 4 Solder: Thin              2.28      0.74      3.1     0        0.84     3.72
## 5 Opening: M:SolderThin    -0.74      1.04     -0.71    0.48    -2.78     1.3
## 6 Opening: S:SolderThin     9.65      1.04      9.28    0        7.61    11.7
```

Part C:

- The baseline number of skips for a large opening and thick alloy is 0.39.
- The main effect for having a medium opening is 2.41 skips, which is the effect a medium opening has in isolation.
- The main effect for having a small opening is 5.13 skips, which is the effect a small opening has in isolation.
- The main effect for using thin alloy is 2.28 skips, which is the effect using thin alloy has in isolation.
- The interaction effect for a medium opening and thin alloy is -0.74 skips. This means that soldering with a medium opening and thin alloy yields average skips that are -0.74 skips less than what you would expect from summing the individual "isolated" effects of the two variables

- The interaction effect for a small opening and thin alloy is 9.65 skips. This means that soldering with a small opening and thin alloy yields average skips that are 9.65 skips more than what you would expect from summing the individual "isolated" effects of the two variables

Part D:

The best combination is using a medium opening and thin alloy. The goal is to minimize the number of skips, so the best combination is the one with the smallest estimated coefficient, which in this case is a coefficient of -0.74.