Homework 9 SDS-315

Lucas Chiang (lmc4866) UT-Austin

2024-04-14

Github Link:

Problem 1 - Get out the vote

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

Table 1: Proportion Table of those Receiving a GOTV call and Whether They Voted in 1998

Voted?	Proportion
0	0.352
1	0.648

Table 2: Proportion Table of those Not Receiving a GOTV call and Whether They Voted in 1998

Voted?	Proportion
0	0.556
1	0.444

Note: 0 means they didn't vote in 1998, while 1 means they did vote

Above, we can observe that according to the major party's database in the given data set (turnout.csv), the proportion of those that received a GOTV call and voted was around 64.8%. Meanwhile, the proportion of those that didn't receive a GOTV call and voted was around 44.4%.

Table 3: Confidence Interval of the Difference of Proportions

	method	mean.of.x	lower	upper	level
diff_prop	t.test	0.203727	0.2031265	0.2043275	0.95

Above is a calculation of a confidence interval for the difference of proportions of people who received a GOTV call and voted in 1998 and people who didn't receive the call and voted (difference was calculated by call minus no call). Based on the confidence interval above, we can be 95% confident that the difference of proportions is between .203 and .204. Since the confidence interval does not contain 0, it is in favor of suggesting that those who received the call were more likely to vote in the 1998 Congressional midterm elections. We cannot state causation because possible confounders haven't been accounted for.

Part B: Are voted1996, AGE, and MAJORPTY confounders?

Table 4: Table of Voters in 1998

Variable	Received Call	Did Not Receive Call
Voted in 1996	85.6%	75.9%
Part of a Majority Party	83.1%	80.0%

Observe in the table above, that of those who voted in 1998, the proportion of 1996 voters in those that received a GOTV call is higher than the category of those that did not receive the call. Additionally, the category of those that received the GOTV call also had a higher proportion of majority party members.

Table 5: Proportion Table of those Who Voted in 1996 and Whether They Voted in 1998

Voted in 1998?	Proportion
0	0.36
1	0.64

Table 6: Proportion Table of those Who Did Not Vote in 1996 and Whether They Voted in 1998

Voted in 1998?	Proportion
0	0.771
1	0.229

In the tables above, it is observed that it is more likely that those who have voted in 1996 to have also voted in 1998. Therefore, since the variable voted1996 makes someone more likely to receive a GOTV call and also vote in 1998, this variable is a competing casual explanation to the GOTV_call variable which deems it as a confounder.

Table 7: Proportion Table of those Who Were Part of a Majority Party and Whether They Voted in 1998

Voted in 1998?	Proportion
0	0.36
1	0.64

Table 8: Proportion Table of those Who Were Not Part of a Majority Party and Whether They Voted in 1998

| Voted in 1998?	Proportion
0	0.65
1	0.35

Similarly, the tables above show that is more likely for a person part of majority party to have voted in 1998. Since people belonging to a majority party made up a higher proportion in those that recieved a GOTV call than the category that did not, this shows that the variable MAJORPTY is a confounder as well.

Distributions of Ages According to Whether a Person Received a GOTV Call
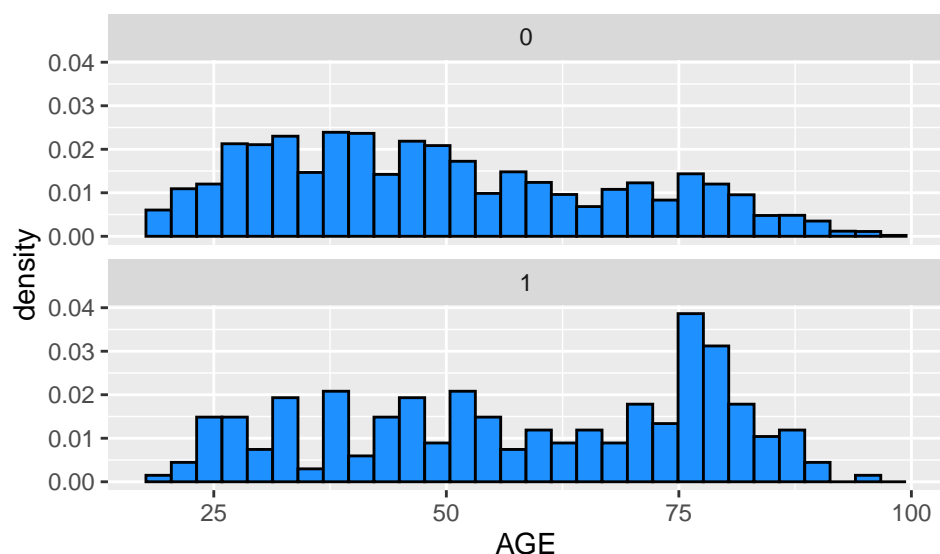
In the faceted histogram above, we can see that the distribution of ages for those that received the GOTV call (indicated by 1) is skewed left which shows it is likely for those that are older to receive the GOTV call. Meanwhile, the distribution of those that didn't receive the GOTV seems to have a slight right skewness which indicates that younger people were less likely to receive the call. This is also supported by the mean ages of these distributions. Those who received a call had a mean age of 58 while those who did not receive a call had a mean age of 49.

Distributions of Ages According to Whether a Person Voted in the 1998 Congressional Midterm
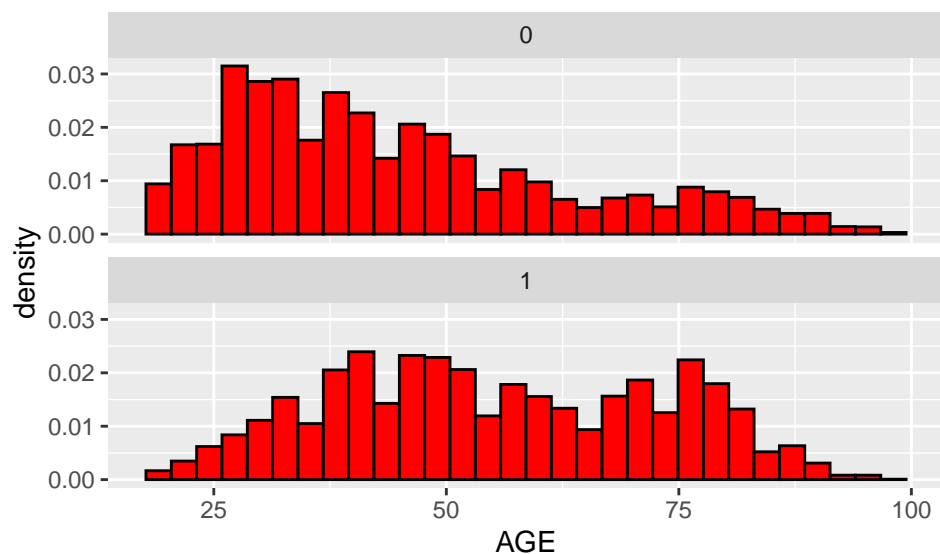
Here is another faceted histogram that allows us to compare the age distributions according to the group that voted in the 1998 Congressional midterm and the group that did not. We can observe that those who didn't vote (indicated by 0) had a right-skewed distribution of age which is consistent with the distribution of ages of those not receiving the GOTV call. Although it is harder in this particular histogram to discern the skewness of the distribution of ages in people that voted, the mean age of people that voted is 55, while the group that didn't had a younger mean with 44.

Overall, since it is more likely that older people were more likely to receive a GOTV call and vote in 1998, the variable AGE is also a confounder.

Part C: Reevaluating the effect of GOTV call on Voting in 1998 with Matching.

Table 9: Summary of Balance for Matched Data With GOTV Call
as The Treatment

	Treatment	Control
AGE mean	58.307	58.266
Proportion Did Not Vote in 1996	.287	.287
Proportion Voted in 1996	.713	.713
Proportion Not in Majority Party	.198	.193
Proportion in a Majority Party	.802	.807

The summary of covariate balance between the treatment and control group has values that are very close to each other, so the matching process will allow us to determine a cause and effect relationship between receiving a GOTV call and voting in the 1998 Congressional midterm elections.

Now, we recalculate proportions and a confidence interval to measure GOTV call's effect using matched data.

Table 10: Proportion Table of those Receiving a GOTV call and
Whether They Voted in 1998

Voted?	Proportion
0	0.352
1	0.648

Table 11: Proportion Table of those Not Receiving a GOTV call
and Whether They Voted in 1998

Voted?	Proportion
0	0.426
1	0.574

In the matched data, the proportion of those who voted in the group that received GOTV calls is still higher than the group who didn't receive the calls, but it is noticeably smaller than in part A(64.8% vs. 57.4%).

Table 12: Confidence Interval of the Difference of Proportions

	method	mean.of.x	lower	upper	level
diff_prop	t.test	0.0739046	0.0732505	0.0745587	0.95

The confidence interval tells us that we are 95% confident that the true difference of proportion of voters in the group that received the GOTV call and those that did not receive the call is between .0733 and .0746. This result is statistical significant as the confidence interval does not contain 0. Even though this difference is smaller than before adjusting for confounders, I would argue that a 7% difference is still practically significant as 7% would amount to a large group of people in a context such as the 1998 Congressional midterms. Hence, I would say that GOTV calls creates an increase in voter turnout.

Problem 2 - Manufacturing Flaws in Circuit Boards

Part A: Evidence that Opening Size and Alloy Thickness is Related to Skips

Opening Size of Solder Gun and Number of Skips
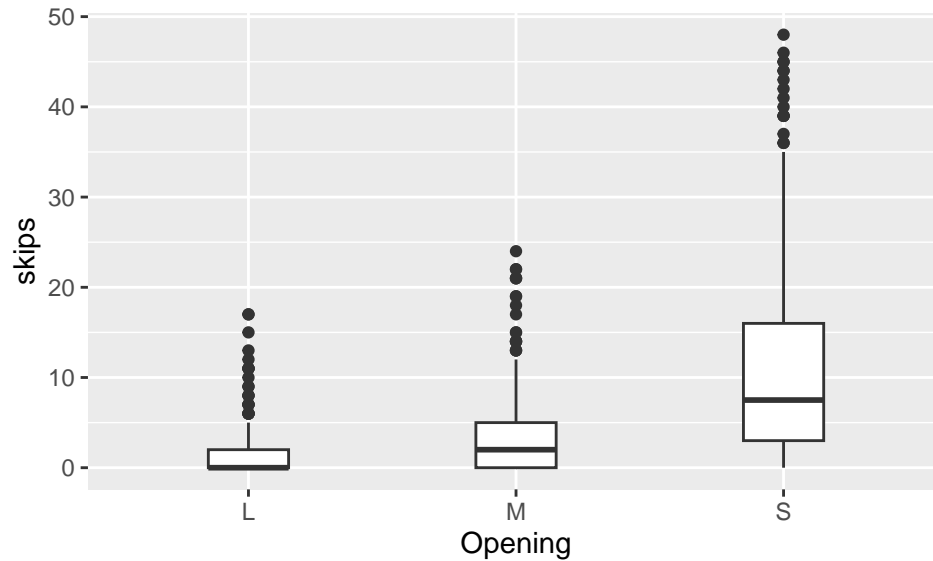
This box-plot shows the number skips related to the opening of the solder gun used on the circuit board. Notice that the median values for size S is the largest followed by M, and then L. Thus, it seems as though that there is a relation between smaller solder gun openings and a larger number of skips.

Thickness of the Alloy Used for Soldering vs. Number of Skips
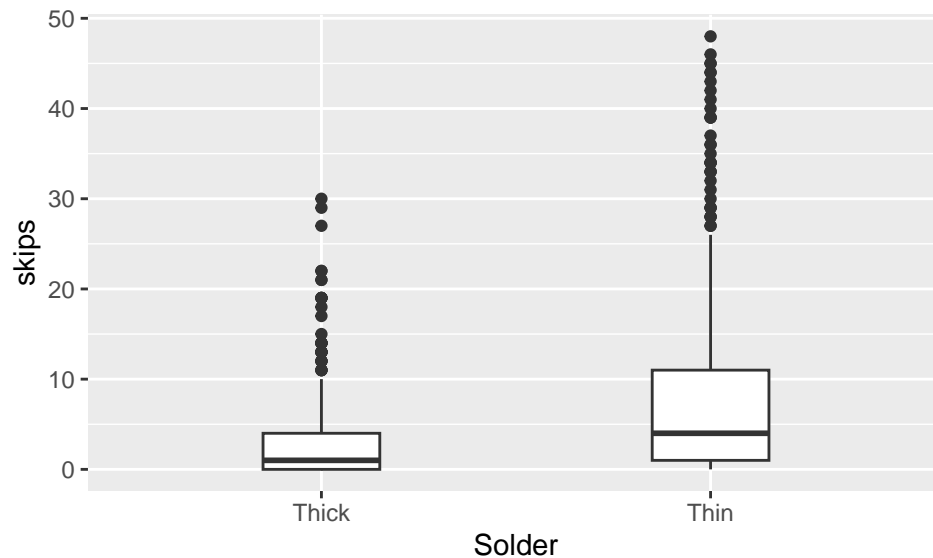
Here is a box-plot showing thickness of the solder alloy and the number of skips. Because the median of the thin alloy has a higher median than the thick alloy, the plot suggests that a thinner alloy is related to a higher number of skips.

Part B: Regression Model Predicting Skips With Multiple Predictors

Table 13: Regression Model For Predicting Number of Skips Using
Opening Size and Solder Type and their Interaction as Predictors

(Intercept)	0.393
OpeningM	2.407
OpeningS	5.127
SolderThin	2.280
OpeningM:SolderThin	-0.740
OpeningS:SolderThin	9.653

Table 14: Confidence Intervals For Each Coefficient in the Regression
Model

	2.5 %	97.5 %
(Intercept)	-0.628	1.415
OpeningM	0.962	3.851
OpeningS	3.682	6.571
SolderThin	0.836	3.724
OpeningM:SolderThin	-2.782	1.302
OpeningS:SolderThin	7.611	11.696

Part C: Interpretation of Each Coefficient in The Regression Model

- The baseline number of skips is .393. The baseline is when we use a large opening size in the solder gun and a thick solder alloy.

- The main effect of using a medium opening size in the solder gun is 2.407 skips.

- The main effect of using a small opening size in the solder gun is 5.127 skips.

- The interaction effect between a medium opening size and a thin solder alloy is -.740 skips.

- The interaction effect between a small opening size and a thin solder alloy is 9.653 skips.

Part D: Recommendation to AT&T

If I were to recommend a combination of opening size and solder thickness to AT&T based on my analysis, I would recommend a large opening size and a thick solder thickness. If we observe the confidence intervals, we can be 95% confident that the effects of using smaller opening sizes (medium and small) and a thin solder alloy would result in an increased number of skips since both ends of the intervals are positive. This undesired effect is especially pronounced in the interaction term between a small opening size and a thin solder alloy. Observing it, we can see that the confidence interval tells us that we can be 95% confident that using this combination would give us around 7.611 to 11.696 skips more than the baseline. This is a large amount of added skips. Therefore, to minimize the number of skips, sticking with a large opening size and a thick alloy is likely the best choice here.