HW 7

Maggie Ma

2025-04-03

Maggie Ma, mm227339, Github Link: https://github.com/maggiema9966/Homework-7

Problem 1

Part A

There are 111 female students and 106 male students in this dataset. 47.17% of the males fold their left arm on top, while 42.34% of the females fold their left arm on top.

Part B

The observed difference in proportions between the genders is 0.04827.

Part C

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(LonR_fold ~ Sex)
## X-squared = 0.33454, df = 1, p-value = 0.563
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.09315879  0.18970817
## sample estimates:
##    prop 1    prop 2
## 0.5765766 0.5283019
```

I used the following formula:

$$SE = \sqrt{((pMales * (1 - pMales))/nMales) + ((pFemales * (1 - pFemales))/nFemales)}$$

where pMales = 0.4716981 and pFemales = 0.4234234. Then, I calculated the left and right bounds of the confidence interval by adding/subtracting z*, which was 1.96 since we are calculating a 95% confidence level. The 95% confidence interval for the difference in proportions is -0.0932 to 0.1897.

Part D

If we were to find the true difference in proportions of males versus females who fold their left arms on top, then we would expect that there is no significant difference between them, since 0 is in the confidence interval, which tells us that it is possible that there is no significant difference in the proportions between the genders.

Part E

The standard error calculated, 0.06745634, measures the variability of the sample difference in proportions. It represents how much we would expect the sample proportion difference to vary from the true population difference.

Part F

In this context, the sampling distribution is the distribution of the difference in proportions of students that fold their left arm on top between males and females that you'd get by repeated random sampling of students from the university population. The arm that is on top is what is varied from sample to sample, and what stays fixed is the true population mean, the sample size, and the sampling method.

Part G

The Central Limit Theorem justifies using a normal distribution to approximate the sampling distribution of the difference in sample proportions. If the sample size is sufficiently large, then the statistical fluctuations in the difference in proportions can be well approximated by a normal distribution.

Part H

Although we cannot say definitively, I would agree with their claim. Because 0 is in the confidence interval, that tells us it is possible that there is no significant difference in the proportions between the genders. Therefore, we do not have evidence to reject the null hypothesis that there is no significant difference in the arm on top when folding arms between the genders.

Part I

Due to random chance, the confidence intervals would be different across different samples. However, at a 95% confidence level, about 95% of all of these confidence intervals should contain the true population difference in proportions.

Problem 2

Part A

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(voted1998 ~ GOTV_call)
## X-squared = 39.597, df = 1, p-value = 3.122e-10
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1411399 0.2659167
## sample estimates:
##    prop 1    prop 2
## 0.5557551 0.3522267
```

The proportion of those who recieved a GOTV call and voted in 1998 is 0.6478, and the proportion of those who didn't receive the GOTV call and voted in 1998 is 0.4442 The 95% confidence interval for the difference in proportions is 0.1411399 to 0.2659167. The proportion of people who voted in 1998 was 0.2035 higher among those who received a GOTV call compared to those who did not.

Part B

voted1996

Table 1: Proportion That Voted in 1996 by GOTV Call Status

	No Call	Call
1996 Proportion Voted	0.530807	0.7125506

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(voted1996 ~ GOTV_call)
## X-squared = 31.32, df = 1, p-value = 2.188e-08
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1224366 0.2410506
## sample estimates:
##    prop 1    prop 2
## 0.4691930 0.2874494
```

Table 2: Proportion That Voted in 1996 by Voting Status in 1998

	No Vote in 1998	Vote in 1998
1996 Proportion Voted	0.3496984	0.7623946

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(voted1996 ~ voted1998)
## X-squared = 1832.4, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.3954939 0.4298985
## sample estimates:
##    prop 1    prop 2
## 0.6503016 0.2376054
```

The first table shows us that the proportion of people who voted in 1996 was 0.1817 higher among those who received a GOTV call compared to those who did not. This means those who voted in 1996 were more likely to receive a GOTV call. This tells us that voting status in 1996 was associated with getting a GOTV call (the independent variable). The second table shows us that the proportion of people who voted in 1996 was 0.4127 higher among those who voted in the 1998 election compared to those who did not. This means that those who voted in the 1996 election were more likely to vote again in the 1998 election. This tells us that voting in 1996 was associated with voting in 1998 (the dependent variable). Additionally, the confidence

intervals for both the independent and dependent variables do not include 0, which tells us that voting status in 1996 is significantly different and is a confounder for both variables. Thus, we have concluded that voting status in 1996 is associated with both the independent (GOTV calls) and dependent (voting status in 1998) variables, which means that it is a confounding variable.

AGE

Table 3: Mean Age by GOTV Call Status

	No Call	Call
Mean Age	49.42534	58.30769

```
##
##  Welch Two Sample t-test
##
## data:  AGE by GOTV_call
## t = -6.9613, df = 256.33, p-value = 2.817e-11
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.395051  -6.369644
## sample estimates:
## mean in group 0 mean in group 1
##        49.42534        58.30769
```

Table 4: Mean Age by Voting Status in 1998

	No Vote	Vote
Mean Age	44.91404	55.41535

```
##
##  Welch Two Sample t-test
##
## data:  AGE by voted1998
## t = -30.24, df = 10568, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.182008  -9.820602
## sample estimates:
## mean in group 0 mean in group 1
##        44.91404        55.41535
```

The first table shows us that the mean age was 8.8824 higher among those who received a GOTV call compared to those who did not. This means older people were more likely to receive a GOTV call. This tells us that age was associated with getting a GOTV call (the independent variable). The second table shows us that the mean age was 10.5013 higher among those who voted in the 1998 election compared to those who did not. This means that older people were more likely to vote in the 1998 election. This tells us that age was associated with voting in 1998 (the dependent variable). Additionally, the confidence intervals for both the independent and dependent variables do not include 0, which tells us that age is significantly different and is a confounder for both variables. Thus, we have concluded that age is associated with both the independent (GOTV calls) and dependent (voting status in 1998) variables, which means that it is a confounding variable.

MAJORPTY

Table 5: Proportion in a Major Party by GOTV Call Status

	No Call	Call
Proportion in Party	0.7447552	0.8016194

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(MAJORPTY ~ GOTV_call)
## X-squared = 3.8248, df = 1, p-value = 0.0505
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.004371919 0.109356458
## sample estimates:
##    prop 1    prop 2
## 0.2552448 0.1983806
```

Table 6: Proportion in a Major Party by Voting Status in 1998

	No Vote	Vote
Proportion in Party	0.7005697	0.8018926

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(MAJORPTY ~ voted1998)
## X-squared = 144.63, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.08499419 0.11765163
## sample estimates:
##    prop 1    prop 2
## 0.2994303 0.1981074
```

The first table shows us that the proportion of people associated with a major party was 0.0569 higher among those who received a GOTV call compared to those who did not. This means those who were associated with a major party were more likely to receive a GOTV call. This tells us that association with a major party was associated with getting a GOTV call (the independent variable). The second table shows us that the proportion of people associated with a major party was 0.1013 higher among those who were associated with a major party compared to those who did not. This means that those who voted in the 1996 election were more likely to vote again in the 1998 election. This tells us that association with a major party was associated with voting in 1998 (the dependent variable). Additionally, the confidence intervals for both the independent and dependent variables do not include 0, which tells us that association with a major party is significantly different and is a confounder for both variables. Thus, we have concluded that association with a major party is associated with both the independent (GOTV calls) and dependent (voting status in 1998) variables, which means that it is a confounding variable.

Part C

Table 7: Matched Mean Age by GOTV Call Status

	No Call	Call
Mean Age	58.2664	58.30769

```
##
##  Welch Two Sample t-test
##
## data:  AGE by GOTV_call
## t = -0.02987, df = 350.55, p-value = 0.9762
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -2.760374  2.677783
## sample estimates:
## mean in group 0 mean in group 1
##        58.26640        58.30769
```

Table 8: Matched Proportion That Voted in 1996 by GOTV Call Status

	No Call	Call
1996 Proportion Voted	0.7125506	0.7125506

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  tally(voted1996 ~ GOTV_call)
## X-squared = 1.9277e-29, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.06182709  0.06182709
## sample estimates:
##    prop 1    prop 2
## 0.2874494 0.2874494
```

Table 9: Matched Proportion in a Major Party by GOTV Call Status

	No Call	Call
Proportion in Party	0.8072874	0.8016194

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(MAJORPTY ~ GOTV_call)
## X-squared = 0.013828, df = 1, p-value = 0.9064
## alternative hypothesis: two.sided
```

```
## 95 percent confidence interval:
##  -0.06247690  0.05114086
## sample estimates:
##    prop 1    prop 2
## 0.1927126 0.1983806
```

In these tables, we can see that the matched confounding variables are much closer to being balanced. For the AGE confounding variable, the unmatched mean ages for no calls versus calls do not match up (the mean age for those called was ~9 years higher), but the matched mean ages were much more balanced, differing by only 0.041 years. For the voted1996 confounding variable, the unmatched proportion of those who voted in the 1996 election was not balanced (the proportion for those who received a call was 0.1817 higher), but the matched proportions were much more balanced, differing by 0. Finally, for the MAJORPTY confounding variable, the unmatched proportion of those who were registered under a major party was not balanced (the proportion for those who received a call was 0.0569 higher), but the matched proportions were much more balanced, differing by 0.0057. Additionally, all of the confidence intervals contain 0, which tells us that there is no difference in the proportions (for MAJORPTY and voted1996 variables) or means (AGE). Thus, we have balanced these confounding variables, as they are now balanced.

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(voted1998 ~ GOTV_call)
## X-squared = 4.9027, df = 1, p-value = 0.02682
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01045353 0.14663149
## sample estimates:
##    prop 1    prop 2
## 0.4307692 0.3522267
```

After matching the dataset, the proportion of those who recieved a GOTV call and voted in 1998 is 0.1854, and the proportion of those who didn't receive the GOTV call and voted in 1998 is 0.8146. The 95% confidence interval for the difference in proportions is 0.01045353 to 0.14663149. The proportion of people who voted in 1998 was 0.0785 higher among those who received a GOTV call compared to those who did not. Because 0 is not in the confidence interval, that tells us it is possible that there is a significant difference in the proportion differences between the whether people got a GOTV call. Therefore, we have evidence to reject the null hypothesis that there is no significant difference in voting rates between the whether people got a GOTV call.