Homework5

Kaleb Luna | kl36927

2025-04-07

GitHub Link:

Problem 1: Armfolding

A. There are **111 female students** and **106 male students** in the given dataset.

The sample proportion of **males who folded their left arm on top** is **0.472**

The sample proportion of **females who folded their left arm on top** is **0.423**

B. **The observed difference** between the two groups (males and females) is approximately **0.0483.**

C. The **95% confidence interval** for the difference of proportions (males minus females) is **(-0.0839, 0.180)**

D. If we were to repeat the study a multitude of times with different random samples of university students, then we'd expect that about 95% of the confidence intervals we obtain would include the true difference between males and females in who folds their left arm on top.

E. The standard error represents how much the difference in sample proportions would vary from sample to sample by chance. It essentially measures the uncertainty in the estimate of the difference between males and females.

F. In this context, the term "sampling distribution" refers to the distribution of the difference in sample proportions between males and females who fold their arms on top, if multiple samples were to be obtained from the same population of university students. The true population proportions remain fixed while the sample proportions vary from one sample to another.

G. The Central Limit Theorem justifies using a normal distribution to approximate the sampling distribution of the difference in sample proportions. According to the Central Limit Theorem, as long as there is a large enough sample size, the distribution of the difference in sample proportions will be approximately normal.

H. If the 95% confidence interval for the difference in proportions was [-0.01, 0.30], there is a possibility that there would be no difference between the two groups in arm folding as the confidence interval contains zero. Therefore, the claim that "there's no sex difference in arm folding" can be true to an extent.

I. The confidence interval would likely be different across samples due to the variability that arises from continuous random sampling. It is important to note, as previously mentioned before, if the study were to take place multiple times, about 95% of the confidence intervals would be expected to contain the true population difference in arm folding between the two groups (male and female).

Problem 2: Get out the vote

Part A

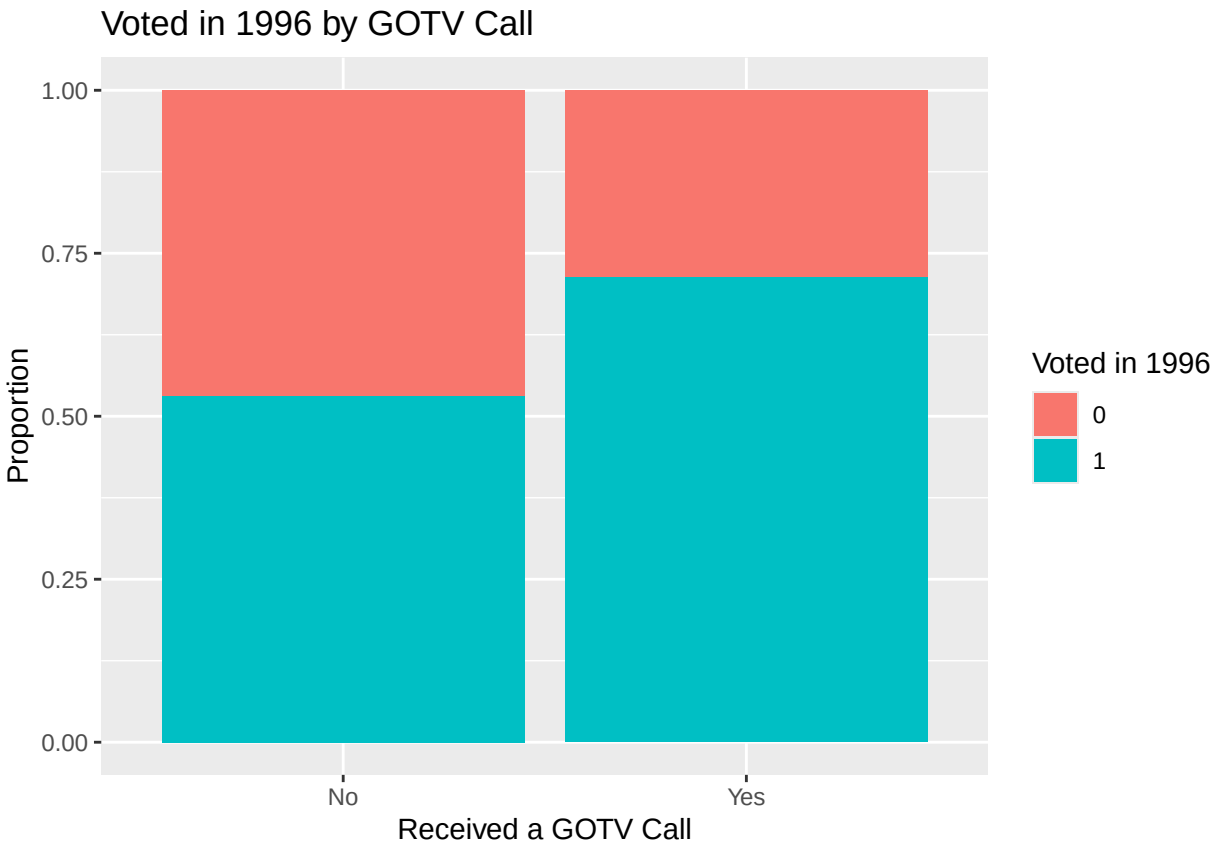
The proportion of those receiving a GOVT call who voted in 1998 is **0.648.**

The sample proportion of those not receiving a GOTV call who voted in 1998 is **0.444.**

The large-sample 95% for the difference in the two proportions (the proportions of voting 1998 for those who received a GOTV call versues those who didn't) is **(0.143, 0.264)**.

Part B

Analyzing the Confounders with Respect to GOTV_call

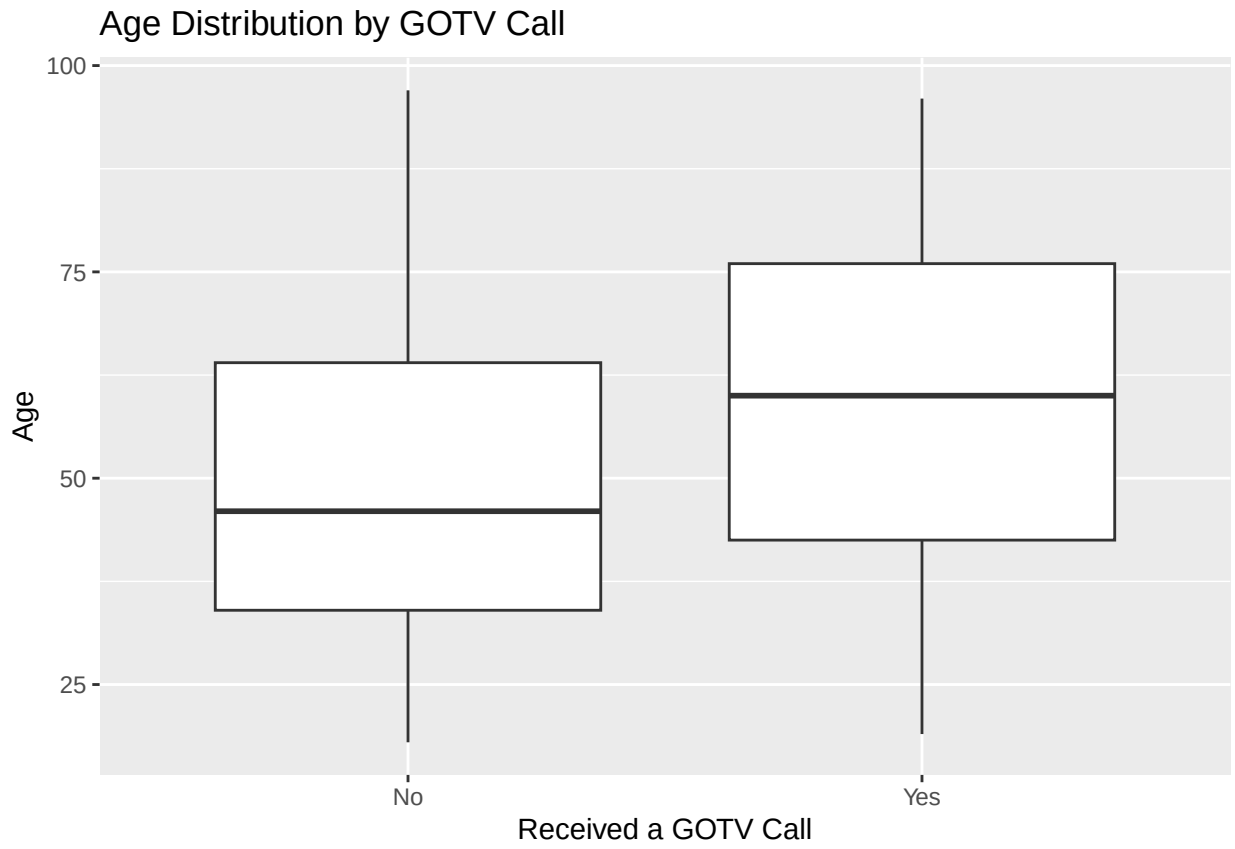



```
## [1] 0.1245081 0.2389791
## attr(,"conf.level")
## [1] 0.95
```

Whether or not an individual voted in 1996 is in fact a confounder. People who received a GOTV call were more likely to have voted in 1996. Furthermore, the 95% confidence interval for the difference in proportions (0.125, 0.239) shows a significant difference in voting history between those who received and didn't receive a GOTV call.


```
## [1] 0.006443461 0.107284916
## attr(,"conf.level")
## [1] 0.95
```

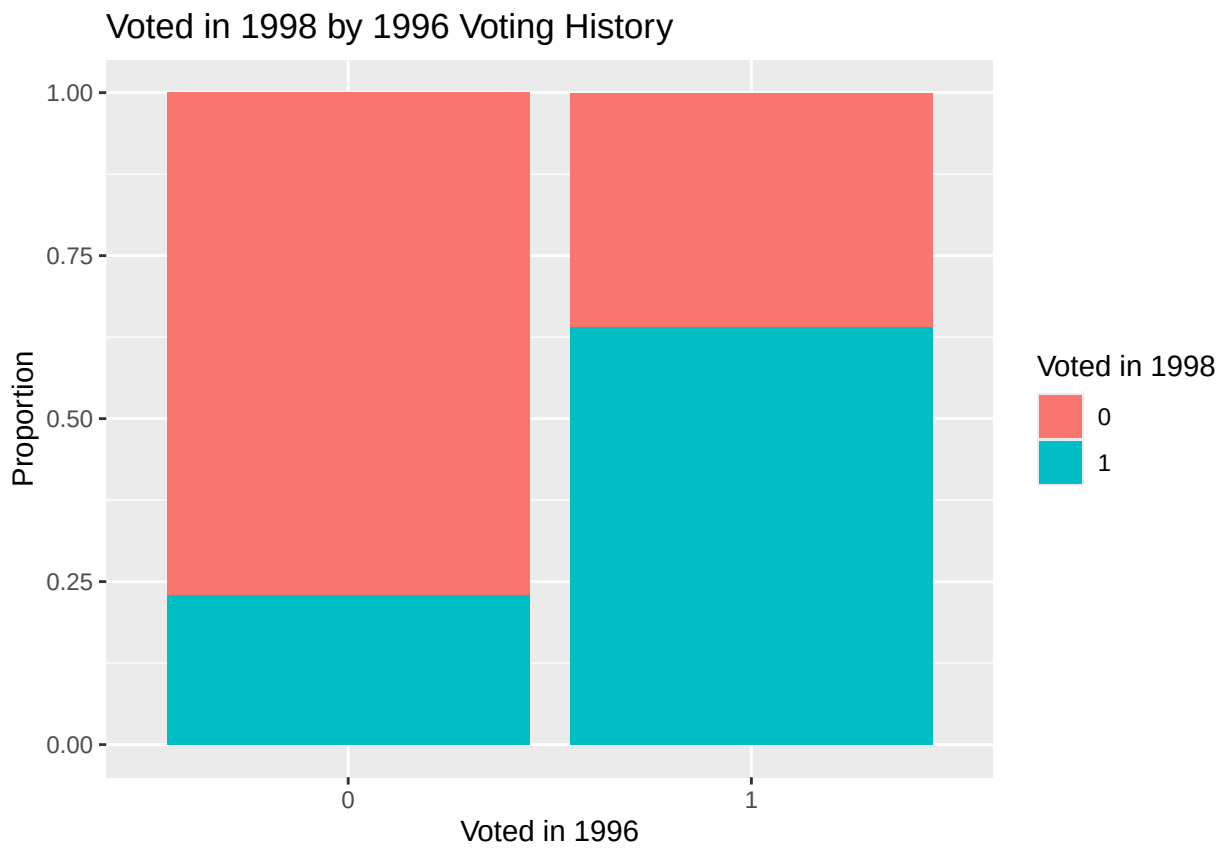
MAJORPTY is in a fact a confounding variable. Those who received a GOTV call were more likely to be registered with a major political party. According to the 95% confidence interval for the difference in proportions (0.00644, 0.1070), there is a significant difference between major party registration between those who received a GOTV call and those who didn't.



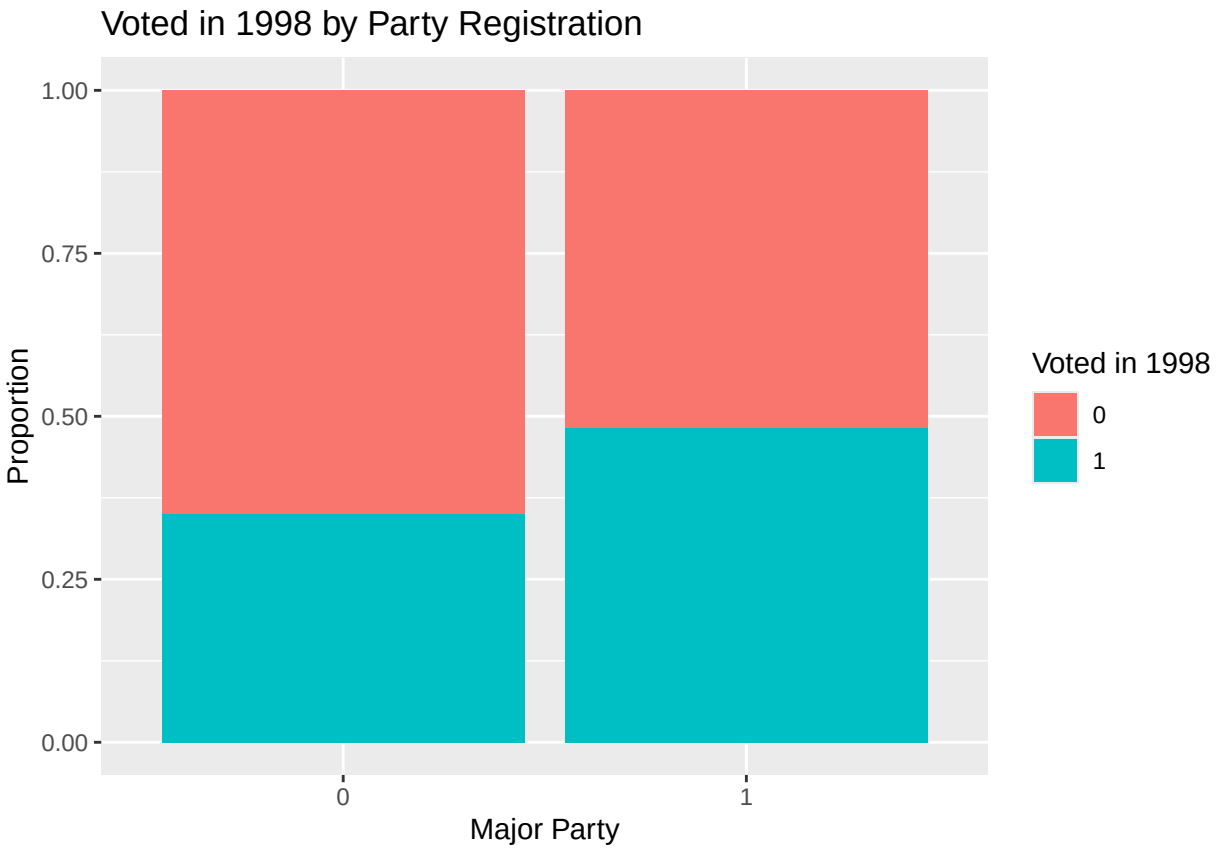

```
## [1] -11.395051  -6.369644
## attr(,"conf.level")
## [1] 0.95
```

According to the plot and t-test above, age is in fact a confounder as the plots indicate that age is associated with voting behavior. The 95% confidence interval, (-1140, -6.37), indicates that the age difference is between 6.37 and 11.40 years, which means that those receiving a GOTV call are usually older.

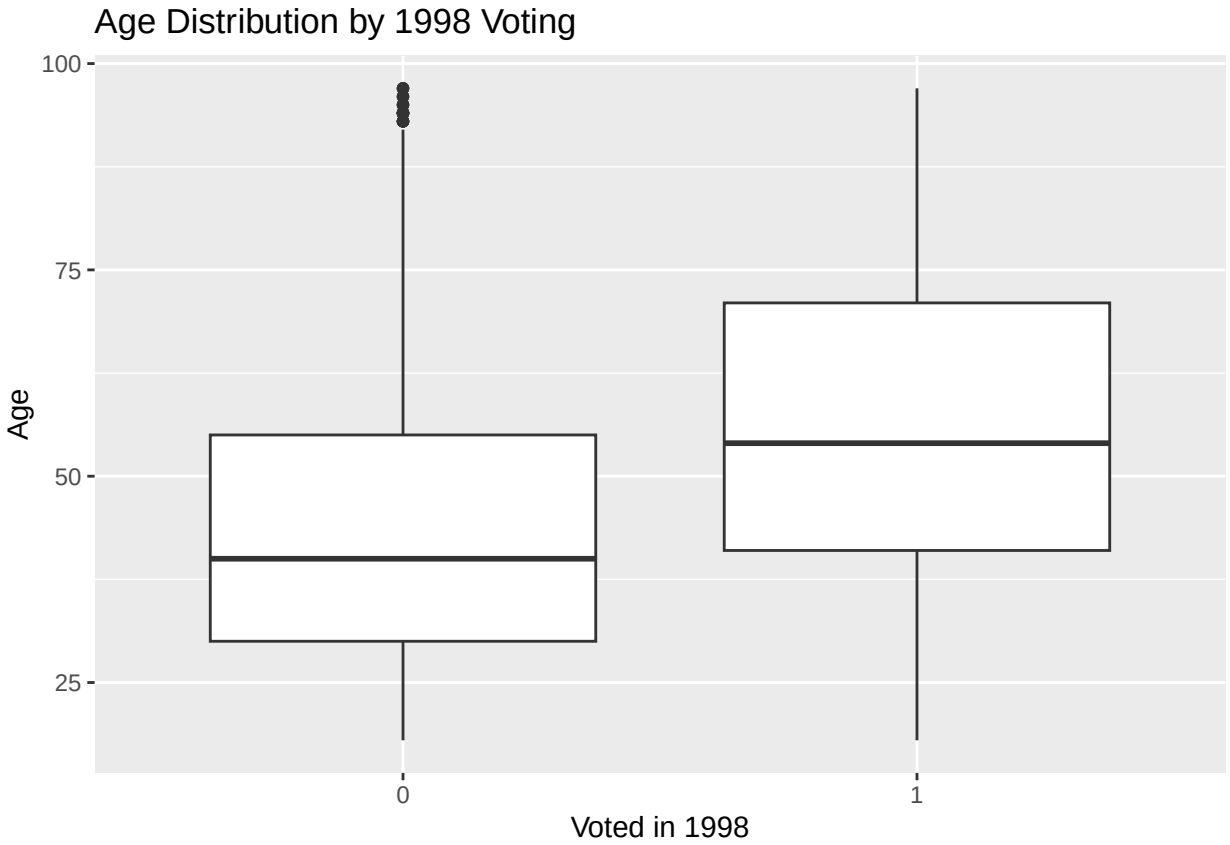
Determining whether the confounding variables are related to whether a person voted in 1998:



Voters in 1996 were more likely to vote again in 1998.



Those registered under a major party were more likely to vote in 1998 than those who were not.



People who voted in 1998 are usually more older than those who did not vote.

Part C

The proportion of those receiving a GOTV call who voted in 1998 is **0.648.**

The sample proportion of those not receiving a GOTV call who voted in 1998 is **0.569.**

A large-sample 95% confidence interval for the difference in these two proportions (the proportions of voting in 1998 for those who received a GOTV call versus those who didn't) is **(0.0129, 0.144)**

Overall, after taking the confounding variables (voting in 1996, age, and party registration) into account, the impact of receiving a GOTV call in the 1998 election was smaller than what it was before taking the confounding variables into account.