Homework 7

Advik Ravichandran

April 7th, 2025

UTEID: asr3897

GitHub Link: https://github.com/advikr/SDS-HW-7

Problem 1: Armfolding

Part A

```
## # A tibble: 2 x 2
##   Sex    count
##   <chr>  <int>
## 1 Female   111
## 2 Male     106
```

```
## [1] 0.4716981
```

```
## [1] 0.4234234
```

The number of males and females in the dataset are 106 and 111, respectively. The sample proportions of males and females who folded their left arm on top is 0.472 and 0.423, respectively.

Part B

The observed difference in proportions is 0.048 between the two groups, calculated by the proportions of males folding their left arm on top minus the proportions of females folding their left arm on top.

Part C

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  recurrences out of sample_size
## X-squared = 0.51118, df = 1, p-value = 0.4746
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.08393731  0.18048668
## sample estimates:
##    prop 1    prop 2
## 0.4716981 0.4234234
```

```
## [1] 0.06745634
```

```
## [1] 0.1804891
```

```
## [1] -0.08393974
```

Standard Error for Difference in Proportions Equations:

$$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

The following values were used to calculate the standard error:

p1 = 0.4716981

p2 = 0.4234234

n1 = 106

n2 = 111

I used the Z* value of 1.96 because that is derived from the 95% large-sample confidence interval of a normal distribution.

Part D

If we were to repeat this study multiple times, 95% of the time, the true difference in proportions between males who fold their left hand on top and females who fold their left hand on top would be between the confidence interval (-0.08393731, 0.18048668).

Part E

The standard error measures the variability between the two sample proportions, which in this case are males and females that fold their hands with their left hand on top, due to chance.

Part F

The sampling distribution refers to the distribution of the difference in sample proportions if were to repeatedly sample the same population. In this case, it is referring to the difference in proportion of males and females that fold their hands with their left hand on top.

Part G

The Central Limit Theorem supports this theory because it states that if a sample size is large enough, the distribution of the sample statistic will be approximately normal, regardless of the population distribution.

Part H

Since the confidence interval contains 0, I think it is probable that there is no sex difference in arm folding.

Part I

The confidence intervals would be different across all samples because you are randomly sampling from a population, thus the data will not be the same due to chance. However, the collection of all these intervals are that 95% of them will contain the sample statistic.

Problem 2

```
## [1] 160
```

```
## [1] 0.01477514
```

```
## [1] 4701
```

```
## [1] 0.4341121
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  recurrences1 out of sample_size1
## 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
## 0.6477733 0.4442449
```

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

The proportion of those not receiving a GOTV call who voted in 1998 was 0.444.

The large-sample 95% confidence interval for the difference in these two proportions was (0.141, 0.266).

Part B

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  call_voted1996_table
## 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
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  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
```

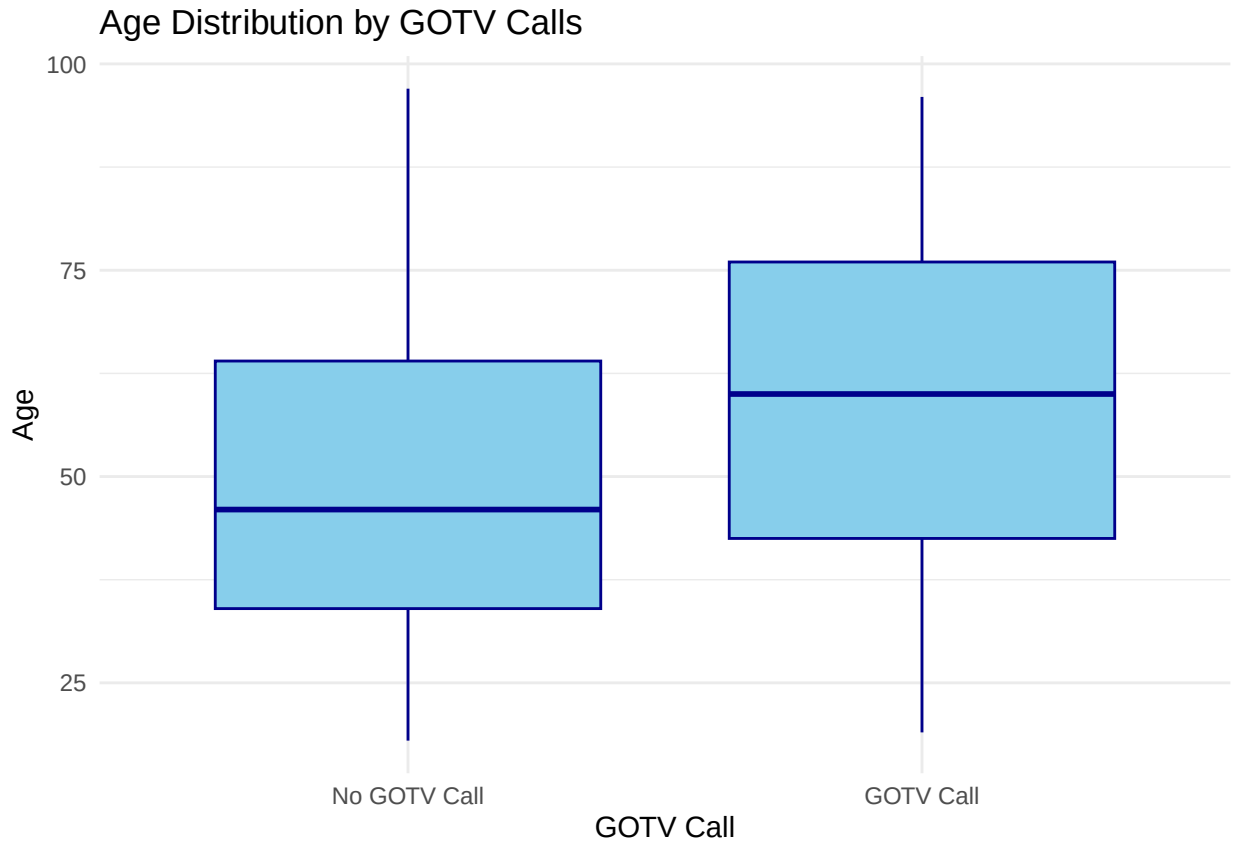
The voted1996 variable is a confounding variable because when calculating the 95% confidence intervals, they both do not include 0. This proves that voted1996 is a confounding variable because it increases the likelihood of both GOTV_call and voted1998.

```
##
```

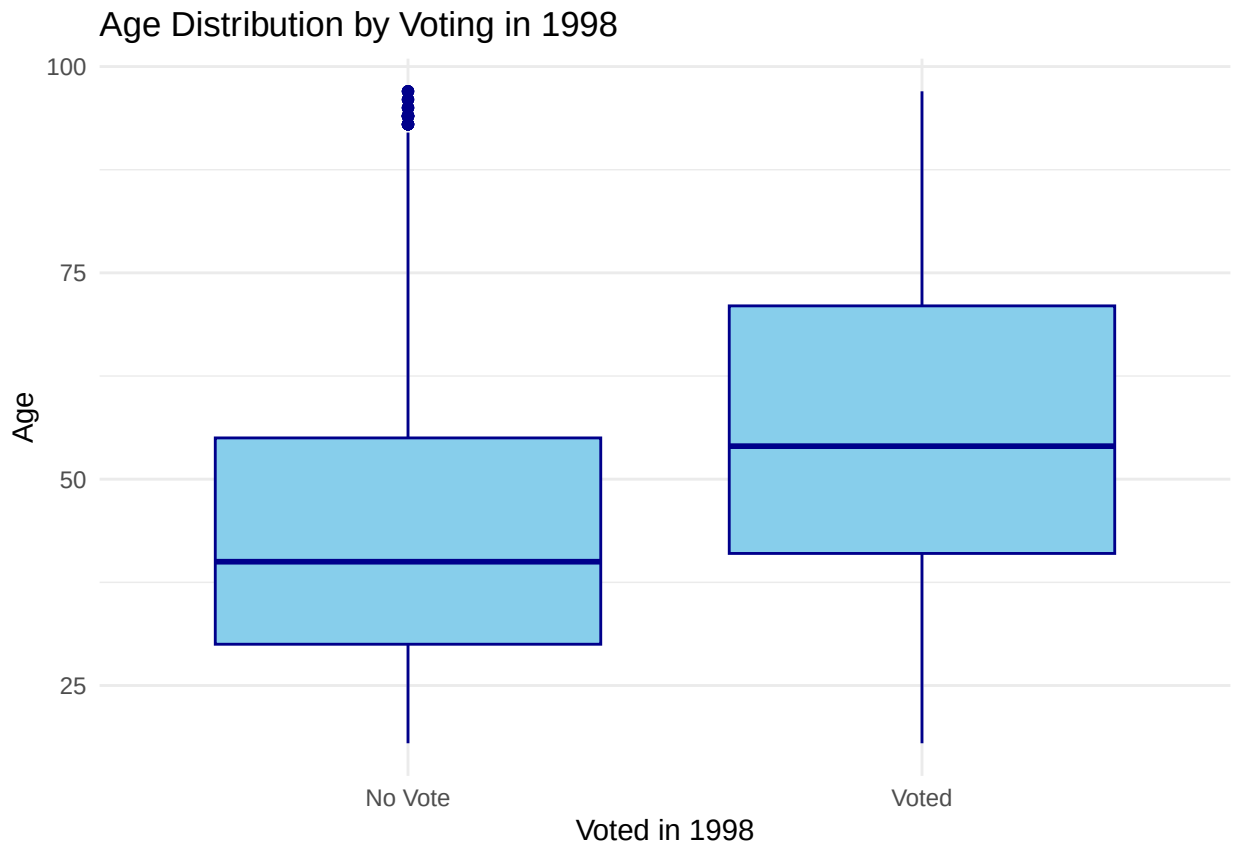
```
##  2-sample test for equality of proportions with continuity correction
##
## data:  gotv_call_majorpty_table
## X-squared = 3.8248, df = 1, p-value = 0.0505
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.004371919 0.109356458
## sample estimates:
##    prop 1    prop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


```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  gotv_call_majorpty_table
## 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
```


The MAJORPTY variable is a confounding variable on the casual effect of GOTV calls impacting the likelihood of a person voting in 1998. This is proven by the prop test 95% confidence intervals as they both do not contain 0, which means that being registered to one of two major political parties increases your chance of getting a GOTV call, and voting in 1998.



```
##
##  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
```




```
##
##  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 AGE variable is a confounder on the casual effect of GOTV calls affecting the likelihood if a person voted in 1998 because it increased the likelihood of both variables. This is proven through the box plots, where the median age is significantly higher in both plots and also with the T test confidence intervals not including 0.

Part C

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  call_96_matched
## 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
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  call_party_matched
## 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
```

```
##
##  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
```

All three confidence intervals include 0, which proves that this matched data set is balanced with respect to voted1996, AGE, and MAJORPTY.

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  votes_count_matched out of total_matched
## 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
## 0.6477733 0.5692308
```

The proportion of those receiving a GOTV call who voted in 1998 was 0.648. The sample proportion of those not receiving a GOTV call who voted in 1998 0.569. The 95% confidence interval for the difference in these two proportions was (0.010, 0.147).