Homework07

Tucker Bullock

2025-04-07

github link: "https://github.com/tuckerbullock/Homework07"

```
## [1] 106
```

```
## [1] 111
```

```
## Proportion of males (left on top): NaN
```

```
## Proportion of females (left on top): NaN
```

```
## Observed difference (male - female): NaN
```

```
## [1] NaN
```

1A) the number of students in the dataset is 217; 106 are male, and 111 are female. The proportion of males that place their left arm over their right arm is 0.472, and for females that proportion is 0.423.

1B) The observed difference in proportions between these groups is (0.472-0.423) 0.048.

1C) R's built in confidence interval function provides an interval of (-0.093, 0.190). To do this by hand, you take the difference in proportions from part B, then add/subtract the standard error times a Z* (using Z* of 1.96 because confidence interval is set at 95%.) to get the bounds of the CI. The formula for SE of a difference in proportions is SE = (sqrt(p1(1-p1)/n1) - sqrt(p2(1-p2)/n2)), where p1 is the proportion of left-over-right males, and p2 is the proportion of left-over-right females. This ends up being (sqrt(0.472(1-0.472)/106) - sqrt(0.423(1-0.423)/111)) = about 0.067. So taking the 0.048 difference from 1B and +- 0.067, we get an interval of (-0.084, 0.182), very similar to the R calculated value.

1D) If we were to repeat this sampling and repeat the calculations many times, we can expect that roughly 95% of the confidence intervals we get would contain the actual difference in proportions between males who fold with their left arm on top and females who fold with their left arm on top.

1E) The standard error attempts to quantify how much sampling variability is in the estimate; it represents the average amount we can expect the difference in sample proportions to vary from the actual difference, solely from the random chance in sampling

1F) We know that the true population proportion stays fixed no matter what, but the sample proportions and the differences in proportions vary from sample to sample. The sample distribution in this context is essentially just referring to the distribution of test statistics. It just describes if you were repeatedly sampling and calculating a difference in sample proportions over and over to create a distribution.

1G) The Central Limit Theorem justifies using a normal distribution to approximate the sampling distribution of a difference in sample proportions. If the sample size is large enough, the CLT states, then the distribution of the sample statistic becomes approximately a normal distribution, no matter what kind of distribution the population distribution is.

1H) We cannot conclude that there is no difference between how males and females hold their arms because 0 is included in the confidence interval. It suggests a positive difference but that is not strong enough evidence to conclude that there is no difference.

1I) Across different samples the confidnce intervals would likely differ, because of random variation in the sampling data. But if we repeat this process many times and calculate an interval every time, about 95% of those intervals would include the actual difference in proportions.

Problem 2

```
## $Proportion_Treated
## [1] 0.6477733
##
## $Proportion_Control
## [1] 0.4442449
##
## $Difference
## [1] 0.2035283
##
## $CI
## [1] 0.1432104 0.2638463
```

Interpretation: The observed difference in turnout between those who received a GOTV call and those who did not is about 20.35% with a 95% confidence interval of approximately [0.143, 0.264].

Part B: Evidence of Confounding

```
##        0        1
## 49.42534 58.30769
```

```
##        0        1
## 44.91404 55.41535
```

```
##          0          1
## 0.01781818 0.02450798
```

```
##         0         1
## 0.3501818 0.4824855
```

```
##          0          1
## 0.01409849 0.03038149
```

```
##         0         1
## 0.2293487 0.6397376
```

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

```
## [1] 0.006107401 0.107620976
## attr(,"conf.level")
## [1] 0.95
```

```
## [1] 0.1241269 0.2393602
## attr(,"conf.level")
## [1] 0.95
```

Interpretation: - People who received a GOTV call tend to be older, more likely to be registered with a major party, and more likely to have voted in 1996. - These same characteristics are also associated with a higher probability of voting in 1998, indicating confounding.

Part C: Matching Analysis

```
##
## Call:
## matchit(formula = GOTV_call ~ AGE + MAJORPTY + voted1996, data = turnout,
##     ratio = 5)
##
## Summary of Balance for All Data:
##           Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance         0.0297        0.0226          0.5130     1.3026    0.1572
## AGE             58.3077       49.4253          0.4475     1.1228    0.1114
## MAJORPTY         0.8016        0.7448          0.1426          .    0.0569
## voted1996        0.7126        0.5308          0.4016          .    0.1817
##           eCDF Max
## distance    0.2499
## AGE         0.2229
## MAJORPTY    0.0569
## voted1996   0.1817
##
## Summary of Balance for Matched Data:
##           Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance         0.0297        0.0297          0.0001      1.004    0.0000
## AGE             58.3077       58.2664          0.0021      1.008    0.0006
## MAJORPTY         0.8016        0.8073         -0.0142          .    0.0057
## voted1996        0.7126        0.7126         -0.0000          .    0.0000
##           eCDF Max Std. Pair Dist.
## distance    0.0057          0.0001
## AGE         0.0057          0.0027
## MAJORPTY    0.0057          0.0183
## voted1996   0.0000          0.0000
##
## Sample Sizes:
##           Control Treated
## All         10582     247
## Matched      1235     247
## Unmatched    9347       0
## Discarded       0       0
```

```
## $Proportion_Treated
## [1] 0.6477733
##
## $Proportion_Control
## [1] 0.5692308
##
## $Difference
## [1] 0.07854251
##
## $CI
## [1] 0.01288148 0.14420354
```

Conclusion: After matching on confounding variables, the estimated causal effect of aGOTV call is about **7.7 percentage points**, with a **95% confidence interval of [2.9%, 12.5%]**. This interval does not include zero, suggesting a positive effect of the calls on turnout. We can conclude that recieving a GOTV call does have an effect on a person's likelihood to vote in 1998.

‘‘‘