Homework 7 - https://github.com/Henryblake2777/SDS315_homework7

Henry Blake - hrb987

2025-04-06

Contents

Problem 1 - Armfolding

Part A - Examine Data

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

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

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

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

number of Males - 106 number of Females - 111 % of Males with Left Hand Top - .4717% % of Females with Left Hand Top - .4234

Part B - Difference in Proportions

```
## [1] 0.0483
```

male% - female% = .0483

Part C - Confidence Interval

```
## [1] 0.1832122
```

```
## [1] -0.08661221
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  c out of c50 out of 10647 out of 111
## 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
```

I used The standard error value for difference in proportions (sqrt((p1)*(1 - p1)/n1 + (p2)*(1 - p2)/n2))

p1 = proportion of left top guys n1 = total guys p2 = proportion of left top girls n2 = total girls

I used a z of 2 to make a confidence interval of around 95%

95% Confint = (-.0839, .1805)

Part D - Interpretation

If we were to take many samples from the population and make a 95% confidence interval for each one, we could expect the population mean to be contained in 95 percent of the intervals

Part E - Standard Error

The standard error that I calculated is measuring the average fluctuation of the mean differences in proportions between samples.

Part F - Sampling Distribution

The sampling distribution represents the distribution of the sample differences in proportions. Between each sample, the sample average difference in proportion changes, but the population average stays constant

Part G - Normal Distribution

The Central Limit Theorem states that as you sample from a population many times, as long as the samples are large enough, then the distribution of the sample means will be normal. This is exactly what we're doing in this experiment.

Part H - No Difference

Although there is no significant statistical evidence to show that there is a difference between men and women because 0 is in the interval, the data shows that there could possibly be a difference as the confidence interval skews slightly positive. (-.01, .30)

Part I - Many Samples

Yes, the confidence interval would be different because each sample would have slightly different data. However, if a 95 percent confidence interval of the difference in proportions is created for each sample, then around 95 percent of those samples would contain the true population mean.

Problem 2 - Get out the Vote

Part A - Examine Data

```
## [1] 247
```

```
## [1] 10582
```

```
## [1] 0.6477733
```

```
## [1] 0.4442449
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  c out of c160 out of 2474701 out of 10582
## X-squared = 40.416, df = 1, p-value = 2.053e-10
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1432115 0.2638452
## sample estimates:
##    prop 1    prop 2
## 0.6477733 0.4442449
```

Prop of GOVT call that voted - 64.78% Prop of no GOVT that voted - 44.42% 95% confint - (.1433, .2638)

Part B - Confounders

```
##   GOTV_call voted1996
## 1         0 0.5308070
## 2         1 0.7125506
```

```
##   GOTV_call      AGE
## 1         0 49.42534
## 2         1 58.30769
```

```
##   GOTV_call  MAJORPTY
## 1         0 0.7447552
## 2         1 0.8016194
```

```
##   voted1998 voted1996
## 1         0 0.3496984
## 2         1 0.7623946
```

```
##   voted1998      AGE
## 1         0 44.91404
## 2         1 55.41535
```

```
##   voted1998  MAJORPTY
## 1         0 0.7005697
## 2         1 0.8018926
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  table1
## X-squared = 3.8248, df = 1, p-value = 0.0505
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.0004615944 0.0129180093
```

```
## sample estimates:
##    prop 1    prop 2
## 0.9821818 0.9754920

##
##  2-sample test for equality of proportions with continuity correction
##
## data:  table2
## X-squared = 144.63, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1111651 0.1534422
## sample estimates:
##    prop 1    prop 2
## 0.6498182 0.5175145

##
##  2-sample test for equality of proportions with continuity correction
##
## data:  table3
## X-squared = 31.32, df = 1, p-value = 2.188e-08
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01060767 0.02195834
## sample estimates:
##    prop 1    prop 2
## 0.9859015 0.9696185

##
##  2-sample test for equality of proportions with continuity correction
##
## data:  table4
## X-squared = 1832.4, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.3932429 0.4275349
## sample estimates:
##    prop 1    prop 2
## 0.7706513 0.3602624

##
##  Welch Two Sample t-test
##
## data:  vote$AGE by vote$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:  vote$AGE by vote$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
```

Since none of the confidence intervals, including both the correlation between the variable and 1998 voting turnout and the variable and receiving a GOTV_call, that means that all three of the variables are confounding variables. In addition, both confidence intervals per variables have the same sign, which means that all three confounding variables are boosting turnout rates for GOTV recipients, not decreasing.

Part C - Matching

```
##
## Call:
## matchit(formula = GOTV_call ~ AGE + voted1996 + MAJORPTY, data = vote,
##     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
## voted1996        0.7126        0.5308          0.4016          .    0.1817
## MAJORPTY         0.8016        0.7448          0.1426          .    0.0569
##           eCDF Max
## distance    0.2499
## AGE         0.2229
## voted1996   0.1817
## MAJORPTY    0.0569
##
## 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
## voted1996        0.7126        0.7126         -0.0000          .    0.0000
## MAJORPTY         0.8016        0.8073         -0.0142          .    0.0057
##           eCDF Max Std. Pair Dist.
## distance    0.0057          0.0001
## AGE         0.0057          0.0027
## voted1996   0.0000          0.0000
## MAJORPTY    0.0057          0.0183
##
## Sample Sizes:
##           Control Treated
## All         10582     247
## Matched      1235     247
## Unmatched    9347       0
## Discarded       0       0
```

The data is well matched because the Means of the treated and control under all three variables are equal. The standard mean diff is also close to 0 for all.

```
## [1] 247
```

```
## [1] 1235

## [1] 160

## [1] 703

##
##  2-sample test for equality of proportions without continuity correction
##
## data:  c out of c160 out of 247703 out of 1235
## X-squared = 5.2206, df = 1, p-value = 0.02232
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01288268 0.14420234
## sample estimates:
##    prop 1    prop 2
## 0.6477733 0.5692308
```

with a p-value of .02 (below .05) and a 95% confidence interval of .0129 - .1442 (0 not in interval), there is intriguing statistical evidence that the GOTV calls made people more likely to vote in the 1998 election.