SDS315_HW7: https://github.com/Panda-nny/SDS315_HW7

Danny Pan

2025-04-05

Problem 1 - Armfolding

Sex	Count	LeftTopProp
Female	111	0.4234
Male	106	0.4717

The table above describes several statistics relating to a sample of students in an experiment. LeftTopProp refers to the proportion of students for each sex whose left arm was on top when folding their arms. The difference in proportion between the two sexes (Male - Female) was approximately 0.049. A test for difference proportions was conducted using the above values to result in a 95% confidence interval, demonstrated by the formula: difference in sample proportions ± z* times the standard error of the sample difference in proportions. The specific numbers come out to be about 0.049 ± 1.96 * 0.067 resulting in the interval:

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  tally(LonR_fold ~ factor(Sex, levels = c("Male", "Female")))
## 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
```

The sampling distribution in this scenario refers to the distribution of differences in proportions of left-arm-top arm folding between males and females undergraduates at this university under repeated trials (of different sample populations). By the Central Limit Theorem, we believe that under repeated sampling, the sample distribution looks more and more normal. Consequently, a z* value of 1.96 was used, due the Empirical Rule which states that for normal distributions, the true difference in proportions is captured 95% of the time within 1.96 * the standard error of our sample proportion.

The standard error details the variability of the sampling distribution primarily as a function of the sample size. The standard error calculation for a difference of proportions is demonstrated below, where p is the value of LeftPropTop, N is the sample size, and the subscript of 1 refers to the male sex and 2 for females. Crunching in the numbers we get a standard error of about 0.0674 (refer to the table for the exact values used).

$$se(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{N_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{N_2}}$$

Using the confidence interval previously computed, We are 95% confident that the true difference in proportions is reflected within the interval [-0.083, 18.05]. Although upon repeated experimentation with different samples, the confidence interval will likely be different across samples due to natural variation within samples. However, if we were to recalculate the interval with differing samples, while all the intervals would be calculated in the same manner, we would expect that the true difference in proportions would be reflected 95% of the time.

For our interval and if the interval was [-0.01, 0.30], we'd agree that there is no conclusive evidence for a sex difference in arm folding, since the interval contains 0.

Problem 2 - Get Out the Vote

A)

According to the sample, 64.8% of recipients of a GOTV callup voted and 44.4% of people who didn't recieve the callup voted. This results in a difference of around 20.35%, with a 95% confidence interval of the true difference in proportions of [14.32, 26.38] percent.

B)

Voting in 1996

Table 2: GOTV Call Proportions by 1996 Voting Status

voted1996	GOTV_call_prop
0	0.0141
1	0.0304

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  tally(GOTV_call ~ factor(voted1996, levels = c(1, 0)))
## X-squared = 32.047, df = 1, p-value = 1.505e-08
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01079327 0.02177274
## sample estimates:
##     prop 1     prop 2
## 0.03038149 0.01409849
```

Voting status in 1996 (1 = Voted in the tables) is shown to result in around a 1.63% increase in an individual's likelihood to recieve a GOTV callup, with a 95% confidence interval of the true difference of [1.08, 2.18] percent.

Table 3: 1998 Voting Proportions by 1996 Voting Status

voted1996	Voted1998Prop
0	0.2293
1	0.6397

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  tally(voted1998 ~ factor(voted1996, levels = c(1, 0)))
## X-squared = 1834.1, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.3934285 0.4273493
## sample estimates:
##    prop 1    prop 2
## 0.6397376 0.2293487
```

Additionally, voting status in 1996 is shown to result in around a 41.03% increase in an individual's likelihood to vote in 1998, with a 95% confidence interval of the true difference of [39.34, 42.73] percent.

Participation in a Major Party

Table 4: GOTV Call Proportions by Major Party Participiation

MAJORPTY	GOTV_call_prop
0	0.0178
1	0.0245

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  tally(GOTV_call ~ factor(MAJORPTY, levels = c(1, 0)))
## X-squared = 4.1195, df = 1, p-value = 0.04239
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.0007053014 0.0126743023
## sample estimates:
##     prop 1     prop 2
## 0.02450798 0.01781818
```

Participation in a major party (1 = Yes, in the tables) is shown to result in around a 0.67% increase in an individual's likelihood to recieve a GOTV callup, with a 95% confidence interval of the true difference of [0.07, 1.27] percent.

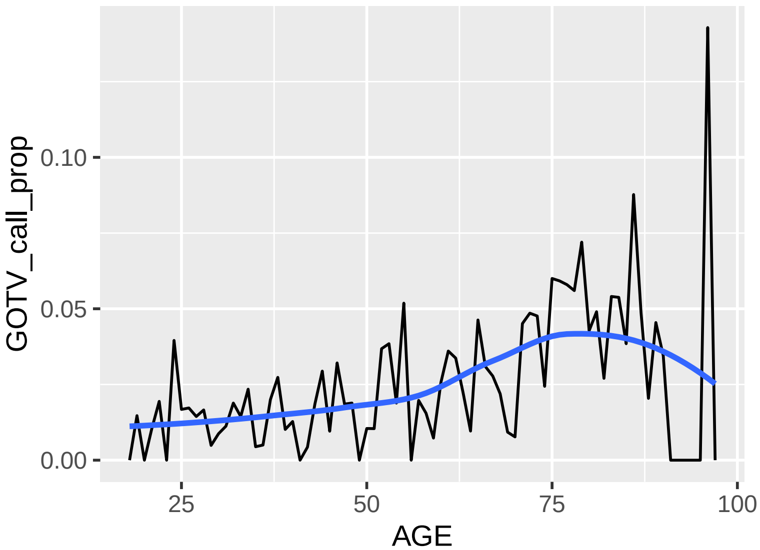
Table 5: 1998 Voting Proportions by Major Party Participation

MAJORPTY	Voted1998Prop
0	0.3502
1	0.4825

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  tally(voted1998 ~ factor(MAJORPTY, levels = c(1, 0)))
## X-squared = 145.17, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1114088 0.1531985
## sample estimates:
##    prop 1    prop 2
## 0.4824855 0.3501818
```

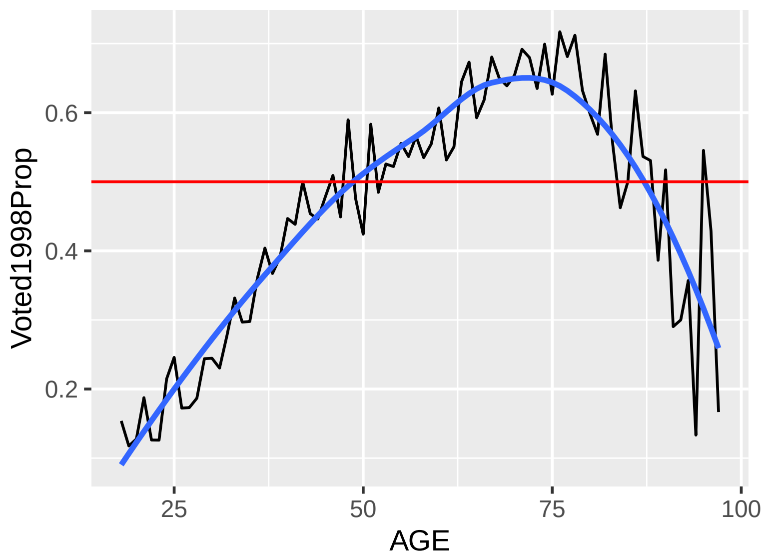
Additionally, participation in a major party is shown to result in around a 13.23% increase in an individual's likelihood to vote in 1998, with a 95% confidence interval of the true difference of proportions of [11.14, 15.32] percent.

Age




```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  tally(GOTV_call ~ factor(age_half, levels = c("UpperHalf", "LowerHalf")))
## X-squared = 30.143, df = 1, p-value = 4.013e-08
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01013515 0.02137160
## sample estimates:
##     prop 1     prop 2
## 0.03069527 0.01494189
```

The line graph of GOTV Call Proportions vs. Age shows a slight, nonlinear trend in likelihood of receiving a call about the senior ages. Comparing the elder vs younger halves of the database sample shows that more senior people are slightly more likely to be called up, with a 95% confidence interval of a [1.01, 2.14] percent difference in proportions.




```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  tally(voted1998 ~ factor(age_half, levels = c("UpperHalf", "LowerHalf")))
## X-squared = 717.93, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.2380287 0.2742393
## sample estimates:
##   prop 1   prop 2
## 0.577108 0.320974
```

Similarly, the line graph for 1998 Voting Proportions vs. Age reflects a nonlinear trend where ages around 70 were more likely to vote than younger and older ages. Comparing the elder vs younger halves of the database sample shows that more senior people are more likely to vote, with a 95% confidence interval of a [23.80, 27.42] percent difference in proportions.

Conclusion For each of the three variables listed (voting status in 1996, participation in a major party, and age), we believe to be confounders for establishing the relationship between receiving a GOTV call and likelihood of voting in 1998 as each variable led to a similar noticeable change in both the likelihood of receiving a GOTV call and of voting in 1998. This is supported by the fact that none of the confidence intervals of the difference of proportions for binary outcomes contained 0.

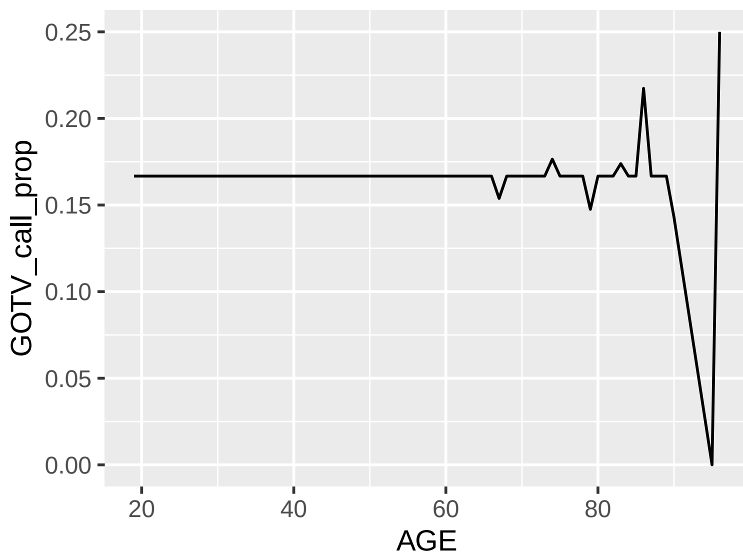
C)

Table 6: GOTV Call Proportions by 1996 Voting Status (matched)

voted1996	GOTV_call_prop
0	0.1667
1	0.1667

Table 7: GOTV Call Proportions by Participation in Major Party (matched)

MAJORPTY	GOTV_call_prop
0	0.1707
1	0.1657



A new matched dataset was created. Now, the proportion of people who received a GOTV call is approximately equal regardless of whether or not they voted, and likewise for the factor of whether the recipient was in a major party. The factor of age's effect on GOTV call rates was also attempted to be matched, with an outlier at the age of 88, where no individual within the original sample at that age got a GOTV call, reflected by the matched dataset.

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  tally(voted1998 ~ factor(GOTV_call, levels = c(1, 0)))
## 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
```

According to the matched data, 64.77% of recipients of a GOTV call voted in 1998 while 56.92% of those who didn't receive a GOTV voted, with a 95% confidence of the true difference in proportions of [1.29,13.42] percent. This suggests that while attempting to control for the confounders of age, 1996 voting status, and participation in a major party, regardless, GOTV calls had a positive effect on the likeliness of voting in the 1998 election.