Homework 2

Trent Latz

2025-04-07

Name: Trent Latz **UT EID:** tjl2597 **GitHub Link:** https://github.com/trentjlatz/SDS315-HW7

Problem 1: Armfolding

A.

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

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

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

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

There are 111 female students and 106 male students in the dataset. The sample proportion of males who folded their left arm on top is .47, while it is .42 for females.

B.

```
## [1] 0.04827469
```

The observed difference in proportions (males minus females) is .048.

C.

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(LonR_fold ~ Sex)
## X-squared = 0.33454, df = 1, p-value = 0.563
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.09315879  0.18970817
## sample estimates:
##    prop 1    prop 2
## 0.5765766 0.5283019
```

```
## [1] -0.08393973  0.18048911
```

With 95% confidence, the built-in R function gives an interval for the difference in proportions (male - female) of -.084 to .180. The formula for the standard error of the difference in proportions is: SE = sqrt(p1(1-p1)/n1 + p2(1-p2)/n2) Where p1 = .4717(male_prop) and n1 = 106(total_males) and p2 = .4234(female_prop)) and n2 = 111(total_females) This gave us a standard error of .067. I calculated my hand-calculated interval using the formula: (p1 - p2) ± z * SE Where p1 and p2 are the same as before, and z is 1.96, which corresponds to 95% of values falling within ± 1.96 standard errors under the normal curve. This also gave me an interval of -.084 to .180, confirming that the two methods agree.

D.

If this study were to be repeated many times with new random samples from the same population, then we would expect that 95% of the constructed confidence intervals would contain the true difference in the proportion of males and females that fold their left arm over their right.

E.

The standard error tells us how much differences in proportions would typically vary across repeated random samples.

F.

In this context, the sampling distribution refers to the distribution of the difference in sample proportions that we would get if we took many random samples from the same population. The observed difference is varying sample to sample due to random chance. The true difference in proportion is unknown, but it stays fixed. Additionally, the sample sizes will stay fixed.

G.

The Central Limit Theorem justifies the use of a normal distribution. This is because the sampling distribution of a statistic like difference in proportions will be normally distributed since both groups have large enough sample sizes.

H.

I would say that the data does not provide strong evidence of a difference in arm folding between sexes. The interval includes 0, which means it is possible there is no real difference in the population. We can't rule out a difference, we just cannot confidently conclude anything.

I.

Yes, the confidence interval would vary from sample to sample because each random sample would give slightly different proportions. However, if this experiment was repeated, about 95% of confidence intervals should contain the true difference in proportion.

Problem 2: Get out the vote

A.

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

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

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(voted1998 ~ GOTV_call)
## 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
## 0.5557551 0.3522267
```

The proportion of those receiving a GOTV call who voted in 1998 was .65, while the sample proportion of those not receiving a GOTV call who voted in 1998 was .44. A large-sample 95% confidence interval for the difference in these two proportions was .14 to .26.

B.

```
## # A tibble: 2 x 4
##   GOTV_call mean_voted1996 mean_age mean_majorpty
##       <int>          <dbl>    <dbl>         <dbl>
## 1         0          0.531     49.4         0.745
## 2         1          0.713     58.3         0.802
```

```
## # A tibble: 2 x 4
##   voted1998 mean_voted1996 mean_age mean_majorpty
##       <int>          <dbl>    <dbl>         <dbl>
## 1         0          0.350     44.9         0.701
## 2         1          0.762     55.4         0.802
```

```
## 
##  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
```

```
## 
##  2-sample test for equality of proportions with continuity correction
## 
## data:  tally(voted1996 ~ GOTV_call)
## 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:  tally(MAJORPTY ~ GOTV_call)
## 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
```

```
## 
##  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
```

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

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  tally(MAJORPTY ~ voted1998)
## X-squared = 144.63, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.08499419 0.11765163
## sample estimates:
##    prop 1    prop 2
## 0.2994303 0.1981074
```

People who received a GOTV call were significantly older (mean age 58 vs. 49, 95% CI: [-11.40, -6.37]), more likely to have voted in 1996 (47% vs. 29%, CI: [0.125, 0.239]), and more likely to be affiliated with a major party (26% vs. 20%, CI: [0.006, 0.107]). These same variables were also significantly associated with voting in 1998. For example, 1998 voters were older (mean age 55 vs. 45), more likely to have voted in 1996 (65% vs. 24%, CI: [0.396, 0.430]), and more likely to be major party members (30% vs. 20%, CI: [0.085, 0.117]). Since all three variables relate to the GOTV call and voting in 1998, they are confounders that influence the difference observed in part A.

C.

```
## # A tibble: 2 x 4
##   GOTV_call mean_voted1996 mean_age mean_majorpty
##       <int>          <dbl>    <dbl>         <dbl>
## 1         0          0.713     58.3         0.807
## 2         1          0.713     58.3         0.802
```

```
## # A tibble: 2 x 2
##   GOTV_call voted1998_rate
##       <int>          <dbl>
## 1         0          0.569
## 2         1          0.648
```

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

```
##
## data:  tally(voted1998 ~ GOTV_call)
## 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
## 0.4307692 0.3522267
```

After matching, the means for the treatment and control groups are extremely similar, with nearly identical averages for prior voting, age, and party affiliation. This suggests that the matched data is no longer confounded by these variables. In the matched sample, 64.8% of individuals who received a GOTV call voted in 1998, compared to 57.0% of those who did not. The 95% confidence interval for the difference in proportions was [0.013, 0.144], and the difference was statistically significant ($p = 0.022$). After adjusting for confounding variables, the GOTV had a positive causal effect on the likelihood of voting in 1998.