315HW7

Elan Satchit

2025-04-06

GITHUB REPO:

Problem 1: Armfolding

Part A

```
## Rows: 217 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Sex, W.Hnd
## dbl (1): LonR_fold
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

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

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

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

Part B

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

Part C

We use the following formula for the standard error of the difference in sample proportions:

SE = sqrt((p1(1 - p1)/n1) + (p2(1 - p2)/n2))

Where: - p1 = 0.472 (Male proportion) - n1 = number of males = 106 - p2 = 0.423 (Female proportion) - n2 = number of females = 111 - z* = 1.96 for 95% confidence

```
## Standard Error: 0.06745634
```

```
## 95% CI Lower Bound: -0.08393973

## 95% CI Upper Bound: 0.1804891

## Manual 95% CI Lower Bound: -0.08393973

## Manual 95% CI Upper Bound: 0.1804891
```

Part D

If we were to repeat this experiment many times with different random samples of university students, then we would expect that about 95% of the confidence intervals we calculate will contain the true difference in proportions.

Part E

The standard error tells us how much we expect the difference in sample proportions to vary from sample to sample. It measures variability in the sampling process.

Part F

The sampling distribution refers to the distribution of the difference in sample proportions (male minus female) that we would get if we repeatedly sampled new groups of males and females from the same population. What varies from sample to sample are the sample proportions (p1 and p2), and therefore their difference. The true population proportions (p1 and p2) stay fixed.

Part G

We can use a normal distribution because of the Central Limit Theorem. It says the sampling distribution of the sample statistic (like a proportion) is approximately normal when the sample size is large enough.

Part H

If the confidence interval was [-0.01, 0.30], we would say we can't rule out no difference, but there's also a chance of a real difference. Since 0 is inside the interval, we don't have strong evidence for a difference.

Part I

If we repeated this experiment many times with random samples, the interval might change, but about 95% of them would capture the true population difference.

Problem 2: Get Out The Vote

Part A

```
## Rows: 10829 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
```

```
## dbl (6): voted1998, GOTV_call, voted1996, PERSONS, AGE, MAJORPTY
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## 95% CI for GOTV difference (Lower): 0.1432104
```

```
## 95% CI for GOTV difference (Upper): 0.2638463
```

Part B

```
## # A tibble: 2 x 5
##   GOTV_call voted1996   AGE MAJORPTY voted1998
##       <dbl>     <dbl> <dbl>    <dbl>     <dbl>
## 1         0     0.531  49.4    0.745     0.444
## 2         1     0.713  58.3    0.802     0.648
```

```
##                 Diff       Lower      Upper
## voted1996 0.18174358 0.124507040  0.2389801
## MAJORPTY  0.05686419 0.006442534  0.1072858
## AGE       8.88234738 6.381461909 11.3832329
```

People who got a call were more likely to have voted before, be older, and belong to a major party.

Part C

```
## # A tibble: 2 x 4
##   GOTV_call voted1996   AGE MAJORPTY
##       <dbl>     <dbl> <dbl>    <dbl>
## 1         0     0.713  58.3    0.807
## 2         1     0.713  58.3    0.802
```

```
##             Difference 95% CI Lower 95% CI Upper
## voted1996  0.000000000  -0.06182822   0.06182822
## MAJORPTY  -0.005668016  -0.06004875   0.04871271
## AGE        0.041295547  -2.66844540   2.75103650
```

Matched GOTV Impact

```
## Proportion voted1998, GOTV group: 0.6477733
```

```
## Proportion voted1998, Control group: 0.5692308
```

```
## Difference: 0.07854251
```

```
## 95% Confidence Interval: [ 0.01288148 , 0.1442035 ]
```

After matching, the three confounding variables (`voted1996`, `AGE`, `MAJORPTY`) are balanced — their confidence intervals all include 0. This means any observed difference in 1998 voting between groups is more likely to reflect a causal effect of the GOTV call.