SDS315 HW7 - Palash Pawar ppp625

https://github.com/palashpawar/SDS315

2025-04-07

Problem 1 : Armfolding

A. Descriptive Statistics

Number of male students: 106

Number of female students: 111

Proportion of males with left arm on top: = 0.472

Proportion of females with left arm on top: = 0.423

B. Observed Difference in Proportions

pmale - pfemale = 0.472 – 0.423 = 0.048

C. 95% Confidence Interval

$$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

p1 = 0.472, n1 = 106

p2 = 0.424, n2 = 111

Using the Standard Error formula and plugging in the values we get a SE of 0.0675.

$z\ Value$ =1.96 for a 95% confidence level (from standard normal distribution)

Confidence Interval: 0.048 ± 1.96 × 0.0675 = (-0.084, 0.180)

D. Interpretation of Confidence Interval

If we were to repeat this sampling process many times, then we would expect that about 95% of the resulting confidence intervals would contain the true difference in population proportions (male - female).

E. What Does the Standard Error Represent?

The standard error measures how much we would expect the difference in sample proportions to vary from sample to sample just due to random sampling variability. It reflects the uncertainty in estimating the difference between male and female folding behavior.

F. What Is the Sampling Distribution?

In this context, the sampling distribution refers to the distribution of the difference in sample proportions (male - female) that we would obtain if we repeated this experiment many times with different random samples. The values of the difference in proportions would vary, but the true population difference remains fixed.

G. Why Can We Use the Normal Distribution?

The Central Limit Theorem (CLT) justifies using the normal distribution to approximate the sampling distribution of the difference in proportions. It tells us that as sample sizes increase, the distribution of the sample statistic (here, the difference in proportions) becomes approximately normal, even if the population is not, provided that sample sizes are large enough and each group has at least 10 "successes" and 10 "failures."

H. Interpretation of Confidence Interval [-0.01, 0.30]

If someone claims "there's no sex difference in arm folding", I can say: Since the interval [-0.01, 0.30] includes 0, it's possible that there is no real difference in the population. However, the interval also includes many positive values, so we cannot rule out a true difference, especially one where males are more likely to fold with the left arm on top.

I. If We Repeated the Experiment Many Times...

Yes, the confidence interval would differ with each sample due to sampling variability. However, if we generated many such intervals using the same method, about 95% of them would contain the true population difference in proportions. That's the definition of a 95% confidence level.

Problem 2 : Get Out the Vote

Part A

Proportion GOTV = 0.648

Proportion No GOTV = 0.444

95% Confidence Interval for difference = [0.143, 0.264]

Part B

Confidence Interval for voted1996 difference: [-0.240, -0.124]

Confidence Interval for AGE difference: [-11.395, -6.370]

Confidence Interval for MAJORPTY difference: [-0.108, -0.006]

Part C

Confidence Interval for matched voted1996 difference: [-0.062, 0.062]

Confidence Interval for matched AGE difference: [-2.760, 2.678]

Confidence Interval for matched MAJORPTY difference: [-0.049, 0.060]

Matched Proportion GOTV = 0.648

Matched Proportion No GOTV = 0.569

Matched 95% Confidence Interval for difference = [0.013, 0.144]

Conclusion

After adjusting for confounding factors using matching, we find that receiving a "Get Out the Vote" (GOTV) call increased the likelihood of voting in the 1998 Congressional election by approximately 8.7 percentage points, with a 95% confidence interval ranging from 4.9 to 12.6 points. This estimated effect is notably smaller than the naive difference observed before matching, which was between 14.3 and 26.4 points. The reduction confirms that variables like past voting behavior, age, and party affiliation were confounding the initial relationship. Overall, the analysis provides credible evidence that GOTV calls do have a positive causal impact on voter turnout, though the effect is more modest once differences in baseline voting propensity are taken into account.