EPSY 5261 : Introductory Statistical Methods

Day 20

Confidence Intervals for Comparing Two Proportions

Learning Goals

- At the end of this lesson, you should be able to…
  - Identify when to answer a research question with a confidence interval
  - Explain the need for creating a confidence interval to do statistical inference
  - Know how to calculate a confidence interval by hand and using R Studio for a difference in proportions
  - Interpret a confidence interval
  - Explain the connection between the confidence interval estimate and the likely outcome of a hypothesis test

Confidence Intervals

- Confidence intervals give a <u>range of plausible values</u> for the **population parameter** by including the uncertainty due to sampling variability in the estimate.

$$CI = \text{Sample statistic} \pm \underbrace{\boxed{\text{Multiplier x (SE)}}}_{\text{Margin of error}}$$

Assumptions needed to use CI for Comparing Proportions

- $n_1(\pi_1) \geq 10$ **and** $n_1(1 - \pi_1) \geq 10$
- $n_2(\pi_2) \geq 10$ **and** $n_2(1 - \pi_2) \geq 10$
- Independence
- Since we don't know the population parameters (π_1, π_2), we substitute $\hat{p}_1$ and $\hat{p}_2$ (sample proportion) for these values, respecticvely, when checking the first two assumptions.

CI for a Comparing Proportions

- **Research Question:** What is the difference in proportion of something between two populations (e.g., what is the difference in the proportion of Americans who subscribe to Netflix and the proportion of Americans who subscribe to Hulu?)
- We need to determine:
  - **Sample Statistic**
  - **Standard Error**
  - **Multiplier**

CI for a Comparing Proportions (cntd.)

- **Sample Statistic:** Since we are trying to estimate the difference between population proportions $(\pi_1 - \pi_2)$, we will use the difference in sample proportions $(\hat{p}_1 - \hat{p}_2)$ as our sample statistic

$$CI = (\hat{p}_1 - \hat{p}_2) \pm \text{Multiplier}(\text{Standard Error})$$

CI for a Comparing Proportions (cntd.)

Table 19.1:
Formulas to compute the standard error (SE) for the different situations we have studied in EPsy 5261.

Situation	SE
Single Mean	$\dfrac{\text{SD}}{\sqrt{n}}$
Single Proportion	$\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
Difference in Means	$\sqrt{\dfrac{\text{SD}_1^2}{n_1}+\dfrac{\text{SD}_2^2}{n_2}}$
Difference in Proportions	$\boxed{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$

Standard Error: We consult Table 19.1 (in textbook) to find the formula for the appropriate SE)

$$CI = (\hat{p}_1 - \hat{p}_2) \pm \text{Multiplier}\left(\sqrt{\frac{\hat{p}_1(1-\hat{p})_1}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2}{n_2}}\right)$$

CI for a Comparing Proportions (cntd.)

- **Multiplier:** The multiplier we use is based on the <u>confidence level</u> we want for our interval. In a CI for comparing proportions, this multiplier is again referred to as z^*. For a 95% CI the value of z^* is approximately 2. (When we compute a CI by hand, we use $z^*=2$. When R computes t^*, it will use a more exact value.)

$$CI = (\hat{p}_1 - \hat{p}_2) \pm 2\left(\sqrt{\frac{\hat{p}_1(1-\hat{p})_1}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2}{n_2}}\right)$$

Phone Survey Incentive Part 2

Write your final confidence interval interpretation on the white board for your group.

Did your interval contain 0? What does this suggest?

Summary

- For a research question asking for an estimate, the best way to answer is with a confidence interval
- The confidence interval allows us to take into sampling account variability