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

- Sampling Variability = Samples vary
- We need something to quantify the uncertainty in our estimates

Confidence Intervals

Terminology

- 95% confidence interval:
  - Sample statistic +/- (2 x SE)
- Margin of error:
  - A specified number of standard errors that we add and subtract from the sample statistic to get a confidence interval.
  - Margin of error quantifies the amount of sampling error due to variation from sample to sample.

Assumptions needed to use z-distribution for single proportion

- Assumptions
  - $n_1\hat{p}_1 \geq 10$
  - $n_1(1-\hat{p}_1) \geq 10$

AND

  - $n_2\hat{p}_2 \geq 10$
  - $n_2(1-\hat{p}_2) \geq 10$

- We use $\hat{p}$ as a stand-in for p, which is unknown.

Formula

$$CI = (\hat{p}_1 - \hat{p}_2) \pm z * SE$$

Table 17.1 in text

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

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

Formula

$$CI = (\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{\sqrt{n_1}} + \frac{\hat{p}_2(1-\hat{p}_2)}{\sqrt{n_2}}}$$

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