EPSY 5261 : Introductory Statistical Methods

Day 19

Confidence Intervals for Comparing Two Means

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 means
    - Interpret a confidence interval

Confidence Intervals

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

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

Assumptions needed to use CI for for Comparing Means

- Both populations are normally distributed **or** both samples have a sample size greater than 30 ($n_1 \geq 30$ and $n_2 \geq 30$)
- Both populations have equal variances (in practice we check that the larger sample variance is no more than four times as great as the smaller sample variance)
- Independence
  -

CI for Comparing Means

- **Research Question:** What is the difference in the average value of something for between two populations (e.g., what is the difference between the mean income for students who graduate from UMN and those that graduate from Anoka Ramsey Community College?)
- We need to determine:
  - **Sample Statistic**
  - **Standard Error**
  - **Multiplier**

CI for Comparing Means (cntd.)

- **Sample Statistic:** Since we are trying to estimate the difference between two population means ($\mu_1 - \mu_2$), we will use the difference between sample means ($\bar{x}_1 - \bar{x}_2$) as our sample statistic

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

CI for Comparing Means (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	$\frac{\mathrm{SD}}{\sqrt{n}}$
Single Proportion	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Difference in Means	$\sqrt{\frac{\mathrm{SD}_1^2}{n_1}+\frac{\mathrm{SD}_2^2}{n_2}}$
Difference in Proportions	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

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

$$CI = (\bar{x}_1 - \bar{x}_2) \pm \text{Multiplier}\left(\sqrt{\frac{SD_1^2}{n_1}+\frac{SD_2^2}{n_2}}\right)$$

CI for Comparing Means (cntd.)

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

$$CI = (\bar{x}_1 - \bar{x}_2) \pm 2\left(\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}\right)$$

College Debt Activity

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

In what cases would we want a single confidence interval vs. a difference in means confidence interval?

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