EPSY 5261 : Introductory Statistical Methods

Day 17

Confidence Intervals for a Single Mean

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 mean
  - Interpret a confidence interval
  - Explain how the confidence level we choose affects our interval

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.

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

Assumptions needed to use CI for Single Mean

- Data comes from a population with a normal distribution **OR** Sample size is large enough (>30)
- Independence

CI for a Single Mean

- **Research Question:** What is the mean value of something for some population (e.g., what is the mean income for students who graduate from UMN?)
- We need to determine:
  - **Sample Statistic**
  - **Standard Error**
  - **Multiplier**

CI for a Single Mean (cntd.)

- **Sample Statistic:** Since we are trying to estimate the population mean (μ), we will use the sample mean ($\bar{x}$) as our sample statistic

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

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

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

$$CI = \bar{x} \pm \text{Multiplier}(\frac{SD_x}{\sqrt{n}})$$

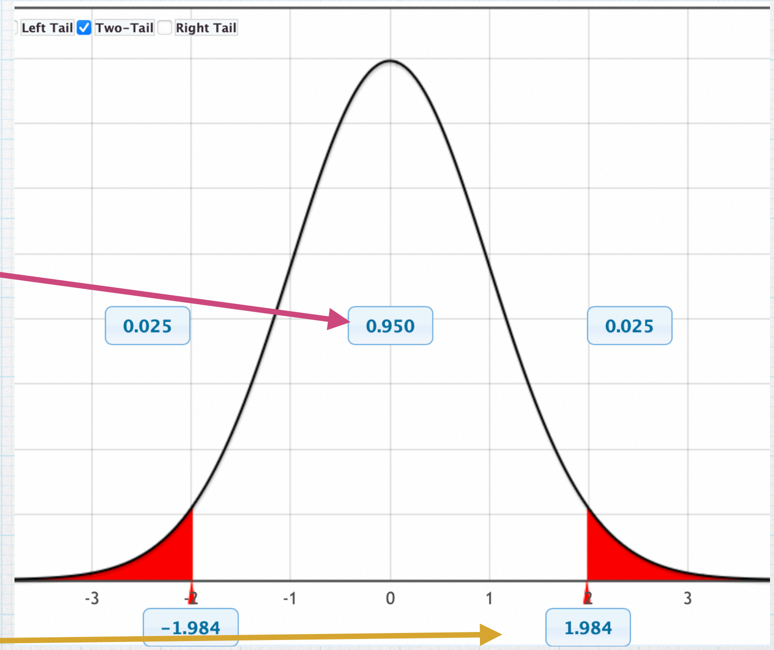
CI for a Single Mean (cntd.)

- **Multiplier:** The multiplier we use is based on the confidence level we want for our interval. In a CI for the mean, 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.)

$$CI = \bar{x} \pm 2(\frac{SD_x}{\sqrt{n}})$$

TMI on t*

- Recall that the shape of the *t*-distribution is based on degrees of freedom (basically our sample size, n-1)
- To get t^* for a confidence level of 95% we need to find at how many standard errors away from the mean we need to be so that the middle proportion of the *t*-distribution is 95%.
- t* = 1.984



T-distribution with sample size 100

t* in Practice

- R will compute the appropriate value for t* when you use the `confint()` function.
- If you ever need to get t*, you can use algebra to find it by substituting in values of your upper CI limit, SD, n, and $\bar{x}$.

$$t^* = \frac{(\text{Upper Limit} - \bar{x}) \times \sqrt{n}}{SD_x}$$

Body Temperature Activity

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

What was the relationship between the confidence level and the 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
- With a higher confidence level we expect a larger confidence interval (more uncertainty in the estimate).