EPSY 5261 : Introductory Statistical Methods

Day 18

Confidence Intervals for a Single Proportion

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 sample size we haves 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 Proportion

- $n\pi \geq 10$
- $n(1 - \pi) \geq 10$
- Independence

Since we don't know the population parameter π, we substitute $\hat{p}$ (sample proportion) in for π when checking the first two assumptions.

CI for a Single Proportion

- **Research Question:** What is the proportion of something for some population (e.g., what is the proportion of Americans who subscribe to Netflix?)
- We need to determine:
  - **Sample Statistic**
  - **Standard Error**
  - **Multiplier**

CI for a Single Proportion (cntd.)

- **Sample Statistic:** Since we are trying to estimate the population proportion (π), we will use the sample proportion ($\hat{p}$) as our sample statistic

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

CI for a Single Proportion (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	$\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

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

$$CI = \hat{p} \pm \text{Multiplier}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

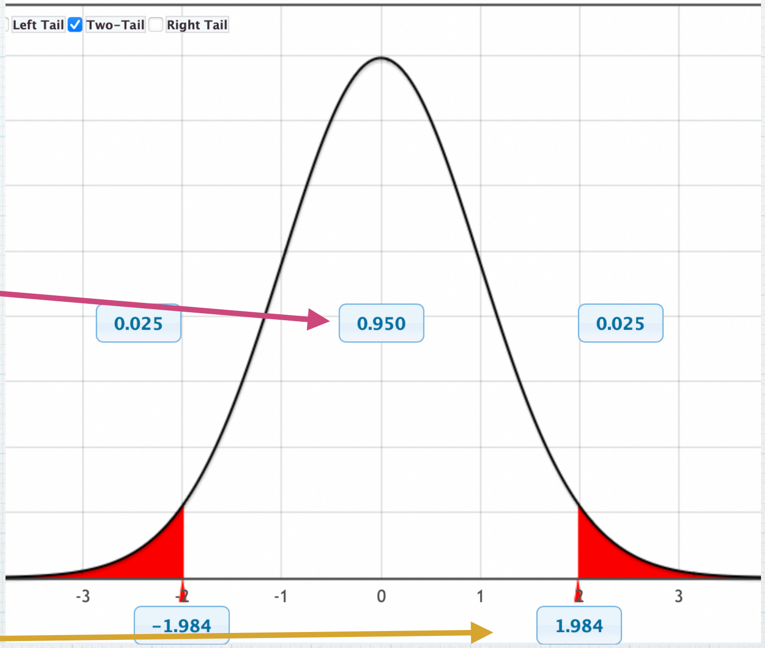
CI for a Single Proportion (cntd.)

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

$$CI = \hat{p} \pm 2\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

TMI on z*

- Recall the z-distribution is a normal distribution with a mean of 0 and a standard deviation of 1.
- To get z^* 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 z-distribution is 95%.
- z* = 1.96



T-distribution with sample size 100

z* in Practice

- R will compute the appropriate value for z* when you use the `confint()` function.
- If you ever need to get z*, you can use algebra to find it by substituting in values of your upper CI limit, sample proportion, and n.

$$z^* = \frac{(\text{Upper Limit} - \hat{p})}{\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)}$$

Worry Activity

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

What was the relationship between the sample size 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 larger sample size we expect a smaller confidence interval.