

Confidence Intervals Pt. 5

Conditions

Grinnell College

Review

We have seen how to make CI's for pop. means or the diff. in pop. means, proportions, diff. in proportions

Confidence intervals will always have the following form:

$$\text{statistic} \pm C \times SE$$

Tough part is finding out formula for SE and deciding what to use for C

- ▶ C choice previously determined by t- or Normal distribution

Outline

Today we will talk more about the conditions necessary to create a CI, and consequences for violating them.

Conditions for Making CI for a mean

The following conditions are necessary for making a CI for a population mean μ with the Normal or t-distribution

- ▶ representative/random sample
- ▶ Normal population **OR** large sample size*

How do we check condition 2? If there is previous evidence of a Normal population \rightarrow OK.

Otherwise look at the sample distribution (histogram of data).

- ▶ Normal: OK
- ▶ symmetric: 15+
- ▶ moderately skewed: 30+
- ▶ extremely skewed 40+

Note: Not always ideal as we are seeing if sample size is OK after already collecting data

Conditions for Making CI for a mean

Looking at the sample distribution is a way to get a *rough* idea of what the population distribution could look like.

We are going to (again) explore what the CLT says. Key idea: CLT is a result that tells us about the sampling distribution.

- ▶ as sample size increases, sampling distribution looks more and more Normal
- ▶ says nothing about Population or Sample distributions
- ▶ Population shape does play into what sample size is required for Normal methods to work well

Note: CLT only works for sampling distributions of means and proportions

Sample Size Demonstration

For the baseball example, roughly what percent of 95% CIs correctly estimated the mean with sample size $n=10$?

For the baseball example (with $n=10$), what "bad" thing do we notice happens relatively frequently?

For the baseball example, roughly what percent of 99% CIs correctly estimated the mean with sample size $n=10$? Should we really call these 99% CIs?

For the baseball example, roughly what percent of 99% CIs correctly estimated the mean with sample size $n=40$?

Sample Size Condition Takeaways

Some of the big takeaways from this:

Sample size can affect the validity of our intervals

- ▶ too small sample sizes can result in inaccurate confident percentages with non-Normal data
- ▶ too small sample sizes can give us really wide intervals with non-Normal data
 - ▶ Negative values often do not make sense for many variables
- ▶ problems with small sample sizes are more noticeable at high confidence levels because intervals depend more on tails (extreme values)

Conditions for CIs for Proportions

Conditions to make a CI for pop. proportion

- ▶ representative sample
- ▶ $n \times \hat{p} \geq 10$ (Successes Condition)
- ▶ $n \times (1 - \hat{p}) \geq 10$ (Failures Condition)

Conditions for CIs for Proportions

What happens when S/F conditions are violated?

Space for Notes:

Conditions for Differences

One of conditions we added when looking at differences in means or proportions is **Independence**: We require the 2 groups to be independent of each other.

- ▶ Knowing something about one group does not tell you something about another
- ▶ observations from one group do not tell us about observations from the other

Examples of Independence:

- ▶ on-campus vs. off-campus students
- ▶ students from two different colleges
- ▶ students from different sections of 209

Examples of Association:

- ▶ same class split into two groups
- ▶ dominant vs. non-dominant hand
- ▶ students from different sections of a class taught by same Prof.