Unit 3: Foundations for inference
2. Confidence intervals

STA 104 - Summer 2017

Duke University, Department of Statistical Science

Prof. van den Boom

Slides posted at
http://www2.stat.duke.edu/courses/Summer17/sta104.001-1/

Announcements

Midterm 1: June 1, Thursday

- Preparation
  - Come to class with questions on Wednesday
  - Sample MT posted on course website
- Rules
  - Bring a calculator + cheat sheet (one sheet, both sides, handwritten, must be prepared by you) + (optional) scratch paper and writing utensil
  - Webcam and microphone need to be working

1. Statistical inference methods based on the CLT depend on the same conditions as the CLT

Always check these in context of the data and the research question!

1. *Independence:* Sampled observations must be independent.
   * This is difficult to verify, but is more likely if
   - random sampling/assignment is used, and,
   - if sampling without replacement, $n < 10\%$ of the population.
2. *Sample size/skew:* Either the population distribution is normal or $n > 30$ and the population distribution is not extremely skewed (the more skewed the distribution, the higher n necessary for the CLT to apply).
   * This is also difficult to verify for the population, but we can check it using the sample data, and assume that the sample mirrors the population.

2. Use confidence intervals to estimate population parameters

$$CI : point\ estimate \pm margin\ of\ error$$

If the parameter of interest is the population mean, and the point estimate is the sample mean,

$$\bar{x} \pm Z^* \frac{s}{\sqrt{n}}$$

3. Critical value depends on the confidence level

Clicker question

What is the critical value ($Z^\star$) for a confidence interval at the 91% confidence level?

(a) $Z^\star = 1.34$
(b) $Z^\star = 1.65$
(c) $Z^\star = 1.70$
(d) $Z^\star = 1.96$
(e) $Z^\star = 2.33$

Common misconceptions about confidence intervals

1. The confidence level of a confidence interval is the probability that the true population parameter is in the confidence interval you construct for a single sample.
   The confidence level is equal to the proportion of random samples that result in confidence intervals that contain the true pop. parameter.
2. A narrower confidence interval is always better.
   This is incorrect since the width is a function of both the confidence level and the standard error.
3. A wider interval means less confidence.
   This is incorrect since it is possible to make very precise statements with very little confidence.

4. Calculate the sample size a priori to achieve desired margin of error

$$ME = z^\star \frac{s}{\sqrt{n}}$$

So if we know the desired *ME*, and confidence level (and hence $z^\star$), and the sample standard deviation, we can solve for n.

Application exercise: 3.1 Confidence interval for a single mean

See course website for details.

1. Statistical inference methods based on the CLT depend on the same conditions as the CLT
2. Use confidence intervals to estimate population parameters
3. Critical value depends on the confidence level
4. Calculate the sample size a priori to achieve desired margin of error