Unit 6: CLT based inference

2. CLT based confidence interval

Stat 140 - 02

Mount Holyoke College

Dr. Shan Shan

Slides posted at *http://sshanshans.github.io/stat140*

1. Announcement

2. Today: CLT based confidence interval

- LEAP symposium this Friday: 9am, 10am and 5pm
  - We *WILL* have class on Friday
  - If you are presenting at LEAP, no class attendance required and best wish of luck!
- Poll on Piazza: what do you want to see in Unit7?
  - ANOVA
  - Multiple regression
  - Inference for regression
  - Bayesian probability
  - Statistics and Ethics
  - Statistics and AI
  - More examples/demo on confidence interval
  - More examples/demo on hypothesis testing
  - More time to work on EAs

1. Announcement

2. Today: CLT based confidence interval

The central limit theorem says ...

- For a sample proportion

$$\hat{p} \sim \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

where p is the population proportion, and n is the sample size

- **For a sample mean**

$$\bar{x} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

where μ is the population mean, σ is the population standard deviation and n is the sample size

CI : point estimate $\pm$ margin of error

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

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

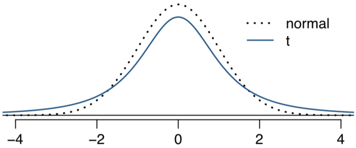
- z^* is critical value, which comes from a standard normal model $\mathcal{N}(0, 1)$
- $\bar{x}$ is the sample mean
- s is the sample standard deviation

How reliable is the estimate s?

Plugging in an estimate introduces additional uncertainty. We make up for this by using a more "conservative" distribution than the normal distribution.

t-distribution also has a bell shape, but its tails are *thicker* than the normal model's

- Observations are more likely to fall beyond two SDs from the mean than under the normal distribution.
- Extra thick tails help mitigate the effect of a less reliable estimate for the standard error of the sampling distribution.

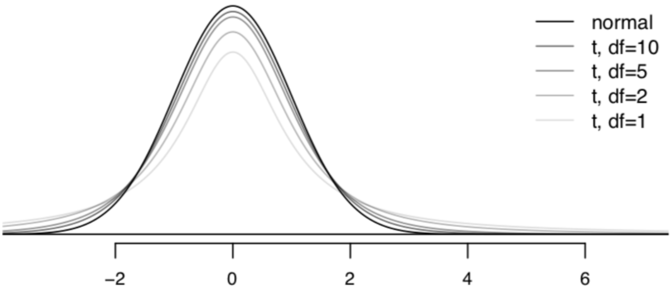


- Always centered at zero, like the standard normal distribution
- Has a single parameter, degrees of freedom (df), that is tied to sample size.

$$df = n - 1$$

- $X \sim T^{n-1}$ is read as a random variable X follows a t-distribution with $n - 1$ degrees of freedom.

What happens to shape of the t-distribution as df increases?



How to make confidence intervals for means with t-distribution?

- Example 1: one mean
- Example 2: comparing two means
- Example 3: paired data

For each example, make sure you know

- the procedure making the confidence intervals
- writing about confidence intervals