Lecture 9:
Confidence Interval for One Mean
STAT 310, Spring 2021

Introduction

- Similar to how we can model the sample proportion $\hat{p}$ using a normal distribution, the sample mean $\bar{x}$ can also be modeled using a normal distribution when certain conditions are met.
- However, we'll learn that a new distribution, called the t-distribution, tends to be more useful when working with the sample mean.
- In this lecture we'll first learn about this new distribution, and then use it to construct confidence intervals for the mean.

Central Limit Theorem (CLT) for the Mean

When the sample size n is sufficiently large, the sampling distribution for $\bar{x}$ follows an approximate normal distribution centered around the population mean μ, and with standard error $SE = \sigma/\sqrt{n}$

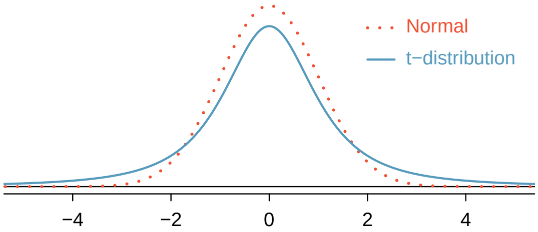
- Based on the CLT, we can construct a 95% confidence interval for the population mean μ as

$$\bar{x} \pm z^* SE \implies \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

- However, one issue with this confidence interval formula is that the standard error is in terms of the population standard deviation σ, which is unknown.
- We can resolve this by plugging in the sample standard deviation s as the estimate of σ. That is, use $SE \approx s/\sqrt{n}$.
- This is a sensible approach when the sample size is large. But when the sample size is small, the confidence interval needs to be adjusted to account for additional uncertainty in estimating σ with s.
- It turns out that we can get a more accurate confidence interval, which accounts for this additional uncertainty, by using the t-distribution to calculate the critical value instead of the z-distrubiton.

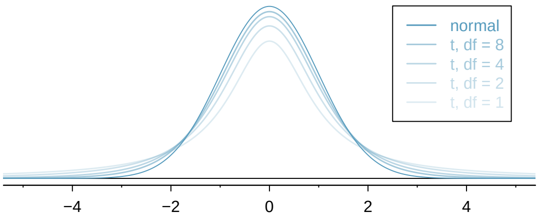
t-disribution

- A t-distribution is bell-curve shaped distribution that is centered around zero.
- It looks similar to a standard normal distribution, but it has wider tails.



t-disribution

- The shape of the t-distribution depends on the degrees of freedom, which is defined as $df = n - 1$
- When the sample size is small the t-distrubiton has noticeably wider tails than a normal distribution.
- When the sample size is large (about 30 or more), the t-distribution is nearly identical to a normal distribution.



Confidence Interval for μ

A confidence interval for the the population mean μ is given by

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

The critical value t^* is found using t-distribution. It depends on the confidence level and degrees of freedom, and can be computed using the R function `qt()`.

Example: Calculate the critical value t^* when the confidence level is 95% and $n = 15$.

Confidence Interval for μ

The confidence interval formula is valid if the following conditions are satisfied:

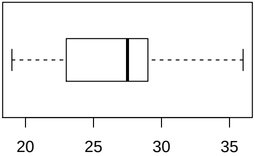
- The data come from a random sample. (This is called the **independence condition** in the textbook.)
- The sample size n is large ($n \geq 30$). Otherwise, if the sample size is small ($n < 30$), the data should have an approximate normal distribution. (This is called the **normality condition** in the textbook.)
- Additionally, the data should not contain any extreme outliers.

Graphical methods (box plots, histograms) can be used to check if the data have an approximate normal distribution when the sample size n is small.

Example

Below are some summary statistics and a box plot for the ages of a random sample of $n = 26$ female athletes who participated in the 2012 Olympic Games in London. Calculate and interpret a 95% confidence interval for the population mean age. Also comment on whether the conditions for the confidence interval appear satisfied.

n	$\bar{x}$	s	min	max
26	26.9	4.5	19	36



Example

Calculate and interpret a 99% confidence interval for the population mean age of female athletes who participated in the 2012 Olympic Games.