Chapter 18 Part One: Confidence Intervals about Means

Recall: Remember from Chapter 15 Part II, that there are three conditions necessary for inference about sample means.

- **Randomization Condition:** Is the sample taken randomly
- **10% Condition:** is the sample less than 10% of the population.
- **Nearly Normal Condition:**
  - our population must be normally distributed
  - sample size must be sufficienty large
    - symmetric : n= 20 or larger
    - skewed population distribution: n = 50 or larger
    - very skewed population distribution: n= 100 or larger

Once these conditions are met, then :

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

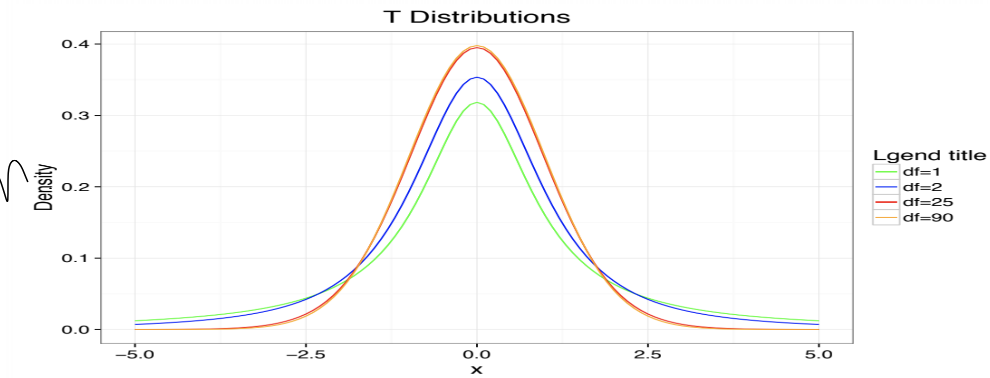
However, there's an issue. If we don't know the population standard deviation, σ, ahead of time (we usually don't) we need to substitute in with the sample mean s. But if we were to substitute in the sample mean, when we standardize scores using our normal z-score fomula, the resulting test statistic doesn't follow a normal distribution.

$$t = \frac{\bar{y} - \mu}{\underbrace{\frac{s}{\sqrt{n}}}_{\text{standard error}}}$$

The t-distribution

The t-distribution is characterized by a parameter called the degrees of freedom.

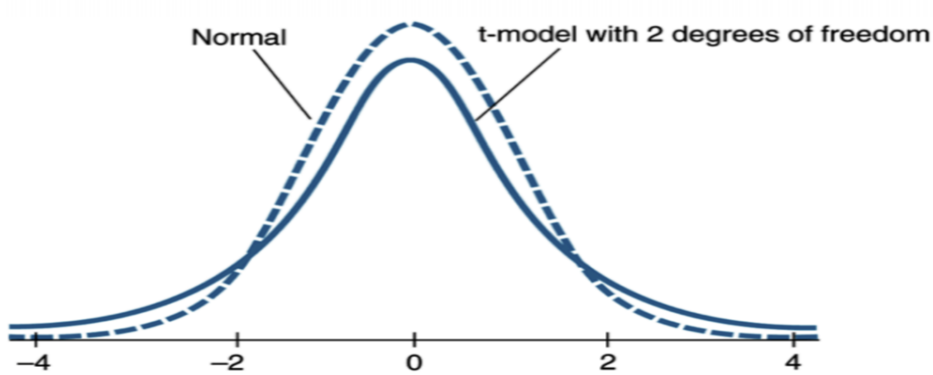
longer tails than normal distribution



Fun Fact: The t-distribution was developed by William S. Gosset who was a statistician and the head brewer at the Guinnes brewery in Dublin, Ireland. He needed a way to analyze data even if the population standard deviation, σ is not known.

Properties of the t-distribution:

- Symmetric around 0
- Bell-shaped
- More probability in the tails of the distribution compared to the Normal curve
- Gets closer and closer to the Normal model as sample size gets larger



If the following conditions are met:

- Randomization
- 10% condition
- Nearly Normal condition

then

$$t = \frac{\bar{y} - \mu}{\frac{s}{\sqrt{n}}}$$

follows a t-distribution with n-1 degrees of freedom (df)

Confidence Intervals

The confidence interval for a population mean can be computed using the following formula:

$$\bar{y} \pm t^* \left(\frac{s}{\sqrt{n}} \right)$$

Where the t^* comes from a t-distribution with n-1 degrees of freedom. This is chosen based on the desired confidence level (using a t-table).

row: df

col: confidence level

interior: t^*

	0.20 0.10	0.10 0.05	0.05 0.025	0.02 0.01	0.01 0.005
df					
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
140	1.288	1.656	1.977	2.353	2.611
180	1.286	1.653	1.973	2.347	2.603
250	1.285	1.651	1.969	2.341	2.596
400	1.284	1.649	1.966	2.336	2.588
1000	1.282	1.646	1.962	2.330	2.581
∞	1.282	1.645	1.960	2.326	2.576
Confidence levels	80%	90%	95%	98%	99%

t-table posted under ch. 18 lecture materials.

1. Find the t* for a 95% CI for a sample of size n = 10.

   $t^* = 2.262$ $\quad df = 10-1$

2. Find the t* for a 90% CI for a sample of size n = 15.

   $df = 15-1 = 14$, $\quad t^* = 1.761$

3. Find the t* for a 95% CI for a sample of size n = ∞

   1.96 same as Normal z-score

Interpretation:

We are ______ (confidence level) confident that the mean number of ______ (event of interest) for all ______ (population) is between ______ (lower bounds) and ______ (upper bound) ______ (units).

Example *A survey of 755 randomly selected US cell phone users age 18 or older was taken in May 2011. The average number of text messages sent or received per day in the sample was 41.5 messagess with a standard deviation of 167.6. Construct a 95% confidence interval for the population mean.*

Check conditions:

1). Survey given randomly

2) More than 755 × 10 US cell phone users

3) regardless of population distribution sample size of n = 755 is sufficiently large

$t^* \approx 1.964$ $\quad SE = \frac{167.6}{\sqrt{755}} \approx 6.1$ $\quad df = 755-1$

CI: $(41.5 - 1.964 \cdot 6.1,\ 41.5 + 1.964 \cdot 6.1)$

$= (29.5,\ 53.5)$

Interpretation:

We are 95% confident that the mean number of text messages sent or received per day for all U.S cell phone users is between 29.5 and 53.5 texts per day.

Example *In a random sample of 130 adults, the sample mean body temperature was 98.25 with a sample standard deviation of 0.73. Compute and interpret a 99% confidence interval for the population mean body temperature.*

$t^* = 2.617$ SE $= .064$

CI: $(98.25 - 2.617(.064),$
$98.25 + 2.617(.064))$

$= (98.08, 98.42)$

Check Conditions

1) Random sample was taken
2) More than $130 * 10$ adults in the world.
3) Regardless of population distribution sample size large enough

interpretation: we are 99% confident that the average body temperature of adults is between (98.08, 98.42) degrees fahrenheit.

Effects of Confidence Level and Sample Size:

$t^* \frac{s}{\sqrt{n}}$

- An increase in confidence level ⇒ an increase in CI width
- a decrease in confidence level ⇒ a decrease in CI width.
- So as $n \uparrow$ the width of CI actually decreases
- as $n \downarrow$ the width of the CI actually get larger.

Meaning of Confidence

Confidence Interval: The actual interval that is attempting to estimate the population parameter.

↳ An estimate of plausible values of the population parameter.

Confidence Level:

Amount of confidence we have in the process.

↳ same idea as the C-L for proportions

↳ In C% of all samples taken, the population mean would be contained in our C.I. calculated using sample data.