Chapter 7, Section 7.1

Inference for population means, $\sigma$ unknown

When the population standard deviation $\sigma$ is unknown we have to estimate it first based on the collected data using the sample standard deviation $s$.

$$
s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}
$$

Using $s$ instead of $\sigma$ adds more variability to the distribution of $\frac{\bar{x} - \mu}{s}$, so we need a distribution with heavier tails.

The $t$-distribution accounts for the additional variation by having heavier tails (see graph next page).

Recall that the normal distribution is characterized by two parameters:

- $\mu$ (the mean) and
- $\sigma$ (the standard deviation)

The $t$-distribution is characterized by a single parameter, the so-called

- “degrees of freedom” (short: df)

As the degrees of freedom increase, the $t$-distribution approaches the standard normal distribution $N(0,1)$.

Why? As the sample size increases, $s$ estimates $\sigma$ more accurately because we have more information about the population standard deviation in our sample.
Section 7.1 — Inference for population means, $\sigma$ unknown

**Reading the $t$-table (Table D)**

- $t$-distribution is symmetric
- It is characterized by the degrees of freedom

**Finding critical values $t^*$ for a $t$-distribution (Table D)**

**Example:**
- 95% percentile of a $t$-distribution with df=5:
  
  $t^*$ is the critical value such that the area to the right (upper tail probability) of $t^*$ is equal to 0.05 (or 5%)

- Look down the “df” column (first column on left) to 5
- At the top of the table, find the right tail (upper tail) probability of 0.05
- The critical value $t^*$ corresponds to where row and column intersect, which is $t^* = 2.015$

Note: $t$-table works with the area above (right) while $z$-table works with area below (left)
Section 7.1 — Inference for population means, $\sigma$ unknown

CONFIDENCE INTERVALS FOR $\mu$ WHEN $\sigma$ IS UNKNOWN

- A $(1 - \alpha) \cdot 100\%$ confidence interval for $\mu$ is given by
  \[
  \left[ \bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}} \right]
  \]
- just change $\sigma$ to $s$ and $z^*$ to $t^*$
- look up $t^*$ corresponding to a $t$-distribution with $df = n - 1$

Example: A random sample of 30 pills yielded a mean level of 20.5 mg of aspirin and a standard deviation of 1.5 mg. Find a 95% confidence interval for the mean level $\mu$ of aspirin in a pill.

What about assumptions for the $t$-test?

- simple random sample (ensuring independence of observations)
- data following a normal distribution or sufficiently large sample size for the CLT to apply

Handout ($t$-test examples)