Hypothesis Testing and CI using
\textit{t}-distribution
Is population SD known or unknown?

Recall (Section 6.2):

- Tests for population mean
- Confidence Intervals
- Connection between the 2 procedures (382–390)

Today (Section 7.1):

- Tests for population mean when the population SD is unknown (418–428)
Example to keep in mind: SAT

- In a discussion of SAT scores, someone comments:
  - “Because only some students take the test, the scores overestimate the ability of typical seniors. The mean SAT–M score is about 475, but I think if all seniors took the test, the mean would be 450.”
- You gave the test to an SRS of 500 seniors from California. They had an average score of 461. (The SAT–M scores follow a normal distribution with a standard deviation of 100.)
- Is there sufficient evidence against the claim that the mean for all California seniors is 450 with a significance level of 0.05?
- Give a 95% CI for the mean score $\mu$ of all seniors.
Equivalence between CI & 2–Sided Tests

- A level $\alpha$ 2–sided test
  - Accepts $H_0: \mu = \mu_0$ exactly when the value $\mu_0$ falls inside a level $1 - \alpha$ confidence interval for $\mu$.
  - rejects $H_0$ when the value $\mu_0$ falls outside the CI.

- CI can be used to test hypotheses:
  - Calculate the $1 - \alpha$ level confidence interval, then
    - if $\mu_0$ falls within the interval, accept the null hypothesis,
    - Otherwise, reject the null hypothesis.
SAT Example

• The hypotheses are $H_0 : \mu = 450$. $H_a : \mu \neq 450$.

• A 95% confidence interval for $\mu$ is

$$\bar{x} \pm z*\sigma / \sqrt{n} = 461 \pm 1.96 \times 100 / \sqrt{500} = [452.2, 469.8].$$

• The value 450 does not lie in the above interval

• At 5% significance level we reject the null hypothesis.
Review

- Problem of interest:
  - Population mean $\mu$ of a normal distribution
  - known $\sigma$
- $Z$-confidence interval:
  $$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$
- $Z$-test:
  $$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

<table>
<thead>
<tr>
<th>$H_a$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu \neq \mu_0$</td>
<td>$2 \ P(Z \geq</td>
</tr>
<tr>
<td>$\mu &gt; \mu_0$</td>
<td>$P(Z \geq z)$</td>
</tr>
<tr>
<td>$\mu &lt; \mu_0$</td>
<td>$P(Z \leq z)$</td>
</tr>
</tbody>
</table>

What if $\sigma$ is unknown?
An investor is estimating the return on investment in companies that won quality awards last year.

A random sample of 50 companies is selected, and the return on investment is calculated had he invested in them. The data is summarized as follows: \( \bar{x} = 14.75, \ s = 8.18. \)

Construct a 95% CI for the mean return.
For CI and hypothesis testing about a normal mean $\mu$, when $\sigma$ is not known, the sample standard deviation $s$ is used to estimate $\sigma$.

Recall: $$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$ where $\sigma$ is assumed to be known.

When $\bar{X}$ is the mean of a random sample of size $n$ from a normal distribution with mean $\mu$, then

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

has a Student $t$ distribution with $n - 1$ degrees of freedom (df).
Student $t$ distribution

- The $t$ distribution is also bell-shaped, and symmetric about zero.

- The “degrees of freedom” determines how spread the distribution is (Note: all $t$-curves are more spread out compared to the standard normal distribution which is equivalent to having $d.f. = \infty$).

$$n_1 < n_2$$
## Table D: t distribution critical values

<table>
<thead>
<tr>
<th>Degrees of Freedom</th>
<th>$t_{.10}$</th>
<th>$t_{.05}$</th>
<th>$t_{.025}$</th>
<th>$t_{.01}$</th>
<th>$t_{.005}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td>63.657</td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td>2.92</td>
<td>4.303</td>
<td>6.965</td>
<td>9.925</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>50</td>
<td>1.299</td>
<td>1.676</td>
<td>2.009</td>
<td>2.403</td>
<td>2.678</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>200</td>
<td>1.286</td>
<td>1.653</td>
<td>1.972</td>
<td>2.345</td>
<td>2.601</td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.282</td>
<td>1.645</td>
<td>1.96</td>
<td>2.326</td>
<td>2.576</td>
</tr>
</tbody>
</table>
t Confidence Interval

- A 100(1 − α)% confidence interval for μ is
  \[ [\bar{x} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}, \bar{x} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}] \]
- or, more compactly,
  \[ \bar{x} \pm t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \]
Investment Return

- An investor is estimating the return on investment in companies that won quality awards last year.
- A random sample of 50 companies is selected, and the return on investment is calculated had he invested in them. The data is summarized as follows: \( \bar{x} = 14.75, \ s = 8.18, \ n=50. \)
- Construct a 95% CI for the mean return:

\[
\bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} = 14.75 \pm 2.009 \frac{8.18}{\sqrt{50}} = [12.43, 17.07]
\]
Take home message

- t-distribution, Table D, d.f.
- Comparison to N(0,1) curve: similarity, difference, connection
- Application: when to use N(0,1), when to use t-distribution?
- SD known or unknown?
- Lecture 22: more about t-tests ...
- Exercises: (don’t turn in) 7.15, 7.21, 7.24 (b) (c)