Lecture 23 (Ch. 8)

We now have a method for hypothesis testing with p-values: (The rejection-region method is covered in a hw).
The method involves computing the prob. of getting a random sample mean more extreme than the obs. sample mean.
Then we check to see if that prob. is "small" compared to a pre-determined prob. which is denoted $\alpha$.

In practice:

1) Choose $\alpha$.

2) Compute the appropriate p-value, according to:

Because of the blue note, above, it is sufficient to test $\mu = \mu_0$

\[
\begin{align*}
\text{If } H_0: \mu &< \mu_0 \quad \text{(or } \mu = \mu_0) \\
H_1: \mu &> \mu_0 \\
p-value &= P(\bar{X} > \bar{X}_{\text{obs}} | \mu = \mu_0) = \text{right area} \\

\text{If } H_0: \mu &> \mu_0 \quad \text{(or } \mu = \mu_0) \\
H_1: \mu &< \mu_0 \\
p-value &= P(\bar{X} < \bar{X}_{\text{obs}} | \mu = \mu_0) = \text{left area} \\

\text{If } H_0: \mu = \mu_0 \\
H_1: \mu \neq \mu_0 \\
p-value &= \text{sum of tail areas, or twice one tail-area.}
\end{align*}
\]

3) If p-value < $\alpha$, then reject $H_0$ in favor of $H_1$.

Else, cannot reject $\cdots \cdots \cdots \cdots \cdots$
In prev. example, we had $n = 64, \bar{x} = 34.4, \ s = 1.1, $ and asked “Does data provide evidence to support $\mu > 34$?” Thus

$H_0: \mu \leq 34$

I always write these so that $H_0$ and $H_1$ have opposite directions, because it’s logical. The book does not. The “equality” in $H_0$ just reminds us that it’s sufficient to test $H_0: \mu = 34$. (The Blue note).

$H_1: \mu > 34$

$df = 64 - 1$

$p$-value = $P(X > X_{\text{obs}}) = P(t > t_{\text{obs}}) = P(t > 2.91) = 0.0025.$

$L = \frac{X_{\text{obs}} - \mu_0}{s/\sqrt{n}} = \frac{34.4 - 34}{1.1/\sqrt{64}} = 2.91$

Since $p$-value < $\alpha$, Thus There is evidence to support $\mu > 34$.

It is tempting to say the above “conclusion” (at $\alpha = 0.05$), That $\mu > 34$, is obvious and trivial. After all the sample gave $X_{\text{obs}} = 34.4$, which is greater than 34 already.

It’s NOT obvious! Suppose the sample/data gave $X_{\text{obs}} = 34.9$, i.e. still larger than 34 - Then

$t = \frac{34.9 - 34}{1.1/\sqrt{64}} = 1.73 \implies p$-value = $P(t > 1.73) = 0.24$

This $p$-value is larger than any reasonable $\alpha$. So, we cannot reject $H_0$ in favor of $H_1$, even though the $X_{\text{obs}}$ sample mean is bigger than 34. 34.9 is larger than 34, but just not enough (in units of standard error, $s/\sqrt{n}$) to justify rejecting $H_0 (\mu < 34)$ in favor of $H_1 (\mu > 34)$. 
There are many ways to rephrase the statement/question in a problem. Here are some of them:

Data Says: \( n = 64 \), \( \bar{x} = 34.4 \), \( s = 1.1 \)

\[ t_{ob} = \frac{34.4 - 34}{1.1 / \sqrt{64}} = 2.91 \]

- Does data support \( \mu > 34 \)?
  - Prior claim: \( H_0: \mu \leq 34 \)
  - \( H_1: \mu > 34 \)
  - \( p\)-value = \( \text{prob}(t > t_{ob}) = \text{prob}(t > 2.91) = 0.0025 < \alpha \)
  - \( \therefore \) Reject \( H_0 (\mu \leq 34) \) in favor of \( H_1 (\mu > 34) \).
  - \( \therefore \) Data does support \( \mu > 34 \).

- Does data support \( \mu < 34 \)?
  - \( H_0: \mu \geq 34 \)
  - \( H_1: \mu < 34 \)
  - \( p\)-value = \( \text{prob}(t < t_{ob}) = 1 - \text{prob}(t > 2.91) = 0.9978 > \alpha \)
  - \( \therefore \) Cannot Reject \( H_0 (\mu \geq 34) \) in favor of \( H_1 (\mu < 34) \).
  - \( \therefore \) Data does not support \( \mu < 34 \).

- Does data contradict \( \mu > 34 \)?
  - Prior claim: \( H_0: \mu \geq 34 \)
  - \( H_0: \mu > 34 \)
  - \( H_1: \mu < 34 \)
  - \( p\)-value = \( \text{prob}(t < t_{ob}) = 1 - \text{prob}(t > 2.91) = 0.9978 > \alpha \)
  - \( \therefore \) Cannot Reject \( H_0 (\mu > 34) \) in favor of \( H_1 (\mu < 34) \).
  - \( \therefore \) Data does not contradict \( \mu > 34 \).

- Does data contradict \( \mu < 34 \)?
  - \( H_0: \mu \leq 34 \)
  - \( H_1: \mu > 34 \)
  - \( p\)-value = \( \text{prob}(t > t_{ob}) = \text{prob}(t > 2.91) = 0.0025 < \alpha \)
  - \( \therefore \) Reject \( H_0 (\mu \leq 34) \) in favor of \( H_1 (\mu > 34) \).
  - \( \therefore \) Data does contradict \( \mu \leq 34 \).
We have learned that the p-value $p = \Pr(\bar{x} > \bar{x}_{\text{obs}} | \mu < \mu_0)$ measures evidence from data in favor of $H_1: \mu > \mu_0$. Then $(1 - p)$ measures evidence for

A) $H_1: \mu < \mu_0$  
B) $H_1: \mu > \mu_0$  
C) none of the above.

hint: $1 - p = \Pr(\bar{x} > \bar{x}_{\text{obs}} | \mu < \mu_0) = p = \Pr(\bar{x} < \bar{x}_{\text{obs}} | \mu > \mu_0)$

This switches  
This does not.

This prob is not any of the p-values listed in the above page.
Now, given the similarity between C.I. and the hypothesis testing approach (i.e., with p-value) guess what the hypotheses for a 2-sample test are:

\[ H_0 : \mu_2 \boxdot \mu_1 \quad \text{vs.} \quad H_1 : \mu_2 \boxdot \mu_1 \text{ (i.e., } \mu_2 - \mu_1 \boxdot 0 \text{)} \]

It turns out we can solve a more general problem:

\[ H_0 : \mu_2 - \mu_1 \boxdot \Delta \quad \text{vs.} \quad H_1 : \mu_2 - \mu_1 \boxdot \Delta \]

I.e., instead of zero, use \( \Delta \), the null parameter. You can always set it to zero, if desired.

\[ \text{YOU choose } \Delta \text{! Not Data.} \]

Then, if 2-samples are independent, then assuming \( H_0 = T_0 \),

\[ z = \frac{(\bar{x}_2 - \bar{x}_1) - \Delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1) \]

\[ t = \frac{(\bar{x}_2 - \bar{x}_1) - \Delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t \text{- dist. with } df = \text{Welch} \]

Then, p-values are computed just as before:

\[ p\text{-value} = \begin{cases} \text{prob}(t > t_{0.05}) & \text{if } H_1 : \mu_2 - \mu_1 > \Delta \\ \text{prob}(t < t_{0.05}) & \text{if } H_1 : \mu_2 - \mu_1 < \Delta \\ \text{twice "tail" if } H_1 : \mu_2 - \mu_1 \neq \Delta \end{cases} \]

(Table VI)

If the two samples are paired, make a new column:

| \( x_1 \) | \( x_2 \) | \( d = x_1 - x_2 \) | \( \bar{d} = \frac{\sum d}{n} \) |
|\hline|\hline|\hline|\hline|
| \text{\ldots} | \text{\ldots} | \text{\ldots} | \text{\ldots} |

\[ t = \frac{\bar{d} - \Delta}{\frac{S_d}{\sqrt{\bar{d}}} \sqrt{n}} \sim t \text{- dist. } df = n - 1 \]

p-value computed as before.
Reconsider this example from a past lecture:

**Example:** 82 students have picked up their test, but 30 have not, even 1 week after the test was returned.

Call these 2 groups “Attendees” and “Non-attendees”.

<table>
<thead>
<tr>
<th></th>
<th>Attendees</th>
<th></th>
<th>Non-attendees</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>82</td>
<td>n</td>
<td>30</td>
</tr>
<tr>
<td>X̄</td>
<td>13.25</td>
<td>X̄</td>
<td>11.8</td>
</tr>
<tr>
<td>s²</td>
<td>3.04</td>
<td>s²</td>
<td>3.32</td>
</tr>
</tbody>
</table>

Sample

\[ \mu_1 = \text{mean of test 1 for non-attend students who have ever taken 390.} \]

\[ \mu_2 = \text{mean of test 1 for attendees.} \]

Is there evidence from data that \( \mu_2 > \mu_1 \)?

We need to build the lower confidence bound for \( \mu_2 - \mu_1 \):

\[
(\bar{x}_2 - \bar{x}_1) - 1.645 \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}
\]

\[
(13.25 - 11.8) - 1.645 \sqrt{\frac{(3.04)^2}{82} + \frac{(3.32)^2}{30}} = 1.45 - 1.645(0.693)
\]

\[
1.45 - 1.14 = 0.31 \implies \frac{0.31}{0} \xrightarrow{\text{estimate}} \bar{x}_2 - \bar{x}_1
\]

**Corollary:** Zero is not included in that interval. So there is evidence that attending students have a higher pop. mean than non-attend.

Now, in Chapter 8’s way:

\[ H_0: \mu_2 - \mu_1 \leq 0 \]

\[ H_1: \mu_2 - \mu_1 > 0 \]

\[ t_{obs} = \frac{1.45 - 0}{0.693} = 2.1 \]

\[ \text{p-value} = \text{prob}(t > 2.1) \approx 0.0205 \implies \text{At } \alpha = 0.05, \text{ p-value} < \alpha. \]

\[ \text{Reject } H_0 \text{ in favor of } H_1 \]

\[ \mu_2 < \mu_1 \quad \mu_2 > \mu_1 \]

In English, there is evidence for \( \mu_2 > \mu_1 \).
We are done with 1-sample and 2-sample, z and t-tests for paired and unpaired data, but all of that has dealt with \( \mu \) or \( \mu_1 - \mu_2 \). What about \( p \) or \( p_1 - p_2 \)? Easy! Follow the pattern:

**C.I. for \( \mu \):**

\[
\bar{x} \pm z \frac{S_x}{\sqrt{n}} \quad | \quad \bar{x} \pm t \frac{S_x}{\sqrt{n}}
\]

\( df = n-1 \)

**Test for \( \mu \):**

\[
H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0
\]

\[
z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{S_x / \sqrt{n}}
\]

\[
p-value = \cdots \quad | \quad df = n-1
\]

**C.I. for \( \pi \):**

\[
p \pm z \sqrt{\frac{p(1-p)}{n}}
\]

**Test for \( \pi \):**

\[
H_0: \pi = \pi_0 \quad H_1: \pi \neq \pi_0
\]

\[
z_{obs} = \frac{\pi_{obs} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \quad | \quad \text{Because we assumed} \quad n \geq 10 \text{ and } 0.1 \leq p \leq 0.9
\]

\[
p-value = \cdots \quad | \quad H_0 = \pi_0
\]

**C.I. for \( \pi_2 - \pi_1 \):**

\[
\frac{\bar{x}_2 - \bar{x}_1 \pm z \sqrt{\frac{\pi_2(1-\pi_2)}{n_1} + \frac{\pi_1(1-\pi_1)}{n_2}}}{\sqrt{n_1} \sqrt{n_2}}
\]

**Test for \( \pi_2 - \pi_1 \):**

\[
p_2 - p_1 \pm z \sqrt{\frac{p_2(1-p_2)}{n_1} + \frac{p_1(1-p_1)}{n_2}}
\]

**C.I. for \( \mu_2 - \mu_1 \):**

\[
\bar{x}_2 - \bar{x}_1 \pm z \sqrt{\frac{\sigma_2^2}{n_1} + \frac{\sigma_1^2}{n_2}}
\]

**Test for \( \mu_2 - \mu_1 \):**

\[
H_0: \mu_2 = \mu_1 \quad H_1: \cdots
\]

\[
z_{obs} = \frac{(\bar{x}_2 - \bar{x}_1)_{obs} - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}
\]

**Test for \( \mu_2 - \mu_1 \):**

\[
H_0: \mu_2 = \mu_1 \quad H_1: \cdots
\]

\[
z_{obs} = \frac{(\bar{x}_2 - \bar{x}_1)_{obs} - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}
\]

Again:

\( df = \text{Welch} \)
We are supposed to transform our question into “Does data provide evidence for ...?” Usually the “...” is specified by you, the scientist. But just for practice, and to better understand the relationship between C.I.’s and p-values, let’s ask “Does data provide evidence for \( \mu_x < \) observed 95% upper confidence bound for \( \mu_x \)?”

a) Set-up \( H_0, H_1 \), b) compute the p-value.

Hint: Recall the defn of the 95% upper conf. bound, and note that the \( t^* \) that appears in that formula satisfies \( p(t > t^*) = 0.05 \)

---

hw-lect17-2 asked does it appear that \( \pi_x \) (the true proportion of defective screws) is at most 2.5%?

There, the appropriate interval is the upper conf. Bound for \( \pi_x \).

a) Set-up the appropriate hypotheses.

b) Compute the p-value (using the data in hw-lect17-2)

c) At \( \alpha = 0.05 \), is the conclusion consistent with the conclusion from the CI approach in hw-lect17-2