Comparison of Two Independent Samples

→ Notation:

Population 1  →  Sample 2  →  Population 2  →  Sample 2

→ Two Approaches for Comparison

1.
2.

→ What seems like a reasonable way to compare two groups?

→ What parameter are we trying to estimate?

→ Sampling distribution \( \left( \bar{y}_1 - \bar{y}_2 \right) \) is

\[
\text{mean} = \frac{\sigma^2}{n}
\]
\[
\text{standard deviation} = \frac{\sigma}{\sqrt{n}}
\]

→ In the formula for the standard deviation why are we using something of the form \( \frac{\sigma^2}{n} \) instead of \( \frac{\sigma}{\sqrt{n}} \)?

→ In the formula for the standard deviation why are we adding \( \frac{\sigma^2}{n} \) for the two samples instead of subtracting \( (\mu_1 - \mu_2) \) is the mean after all)?
(\overline{y}_1 - \overline{y}_2) estimates

**Standard Error of** (\overline{y}_1 - \overline{y}_2)

→ To attach precision to our estimate of (\mu_1 - \mu_2) we need to calculate

\[
SE_{(\overline{y}_1 - \overline{y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{SE_1^2 + SE_2^2}
\]

**Example:** A study is conducted to quantify the benefits of a new cholesterol lowering medication. Two groups of subjects are compared, those who took the medication twice a day for 3 years, and those who took a placebo. Assume subjects were randomly assigned to either group and that both groups data are normally distributed. Results from the study are shown below:

<table>
<thead>
<tr>
<th></th>
<th>Medication</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \overline{y} )</td>
<td>209.8</td>
<td>224.3</td>
</tr>
<tr>
<td>n</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>s</td>
<td>44.3</td>
<td>46.2</td>
</tr>
<tr>
<td>SE</td>
<td>14.0</td>
<td>14.6</td>
</tr>
</tbody>
</table>

→ Calculate an estimate of the true mean difference between treatment groups and this estimate’s precision.

\[ \text{group 1} = \quad \text{group 2} = \]

**Pooled vs. Unpooled**

→ \( \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \) is also know as

→ Pooled variance is a weighted average of
NOTE: If \((\sigma_1 = \sigma_2)\) or if \((s_1 = s_2)\) the pooled and unpooled will give the same answer for \(SE_{(\bar{x}_1 - \bar{x}_2)}\).

It is when \(n_1 \neq n_2\) that we need to decide whether to use pooled or unpooled:

- if \(s_1 = s_2\) then use pooled (unpooled will give similar answer)
- if \(s_1 \neq s_2\) then use unpooled (pooled will NOT give similar answer)

→ Why do we choose to always use the unpooled version?

\[ SE_{pooled} = \sqrt{s_{pooled}^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \]

\[ \text{CI for } (\mu_1 - \mu_2) \]

\[
(\bar{x}_1 - \bar{x}_2) \pm t(df)_{\alpha/2} \left( SE_{\bar{x}_1 - \bar{x}_2} \right)
\]

where \(df = \frac{(SE_{1}^2 + SE_{2}^2)^2}{SE_{1}^4/(n_1 - 1) + SE_{2}^4/(n_2 - 1)}\)

→ What is the df rule of thumb that you can use to check your work? *NOTE:* you should use the df formula above, the rule of thumb will just verify that you are in the ballpark.

*Example:* Cholesterol medication (cont’)

→ Calculate a 95% confidence interval for \((\mu_1 - \mu_2)\)

→ What does this mean conclusion-wise?

→ Could we say that the data shows for certain that one group has higher cholesterol that the other? Explain.
→ Explain the zero rule.

→ Suppose the CI came out to be (5.2, 28.1), would this indicate a true mean difference? Explain.

Hypothesis Testing: The independent t-test

→ Explain the idea behind a hypothesis test.

→ Identify the main parts of a hypothesis test:
   1.
   2.
   3.
   4.

Hypothesis Testing: #1 The hypotheses

→ (#1) There are two types of hypotheses:
   1. notation:
   2. notation:

→ When we compare two group means “nothing is going on” means:

→ For the independent t test the null and alternative hypotheses are:
   Ho:
   Ha:
Example: Cholesterol medication (cont’)

Suppose we want to carry out a hypothesis test to see if the data show that there is enough evidence to support a difference in treatment means.

→ Find the appropriate null and alternative hypotheses. What do these mean in terms of the example?

Hypothesis Testing: #2 Test Statistic

→ (#2) A test statistic measures:

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{\bar{y}_1 - \bar{y}_2}}$$

→ If there were absolutely NO difference between the sample means what would the value of $t_s$ be?

→ If there were absolutely NO difference between the sample means what would we probably conclude about the null hypothesis?

→ If the test statistic is close to 0, this shows

→ If the test statistic is far from 0, this shows

→ With respect to the formula for $t_s$, why does this make sense? Explain.
**Example:** Cholesterol medication (cont’)

→ Calculate the test statistic.

→ What does the value of \( t_s \) mean? Why?

→ From the value of \( t_s \) above does it seem like this supports or goes against Ho? Explain.

**Hypothesis Testing: #3 P-value**

→ The p-value is the area…

→ Mark the area on the t distribution that corresponds to a p-value

*Definition (p.238):* The p-value for a hypothesis test is the probability, computed under the condition that the null hypothesis is true, of the test statistic being at least as extreme or more extreme as the value of the test statistic that was actually obtained [from the data].

→ A large p-value implies…? What is large?

→ A small p-value implies…? What is small?

→ The significance level of a hypothesis test is denoted by what?

→ Why is it important to select your significance level before you collect and analyze your data?
What are the rules for making a decision?

Why do refer to our decision as fail to reject Ho and not accept Ho?

*Example: Cholesterol medication (cont’)

Find the p-value that corresponds to the results of the cholesterol lowering medication experiment.

Hypothesis Testing: #4 Conclusion

Suppose the researchers had set \( \alpha = 0.05 \), what would our quick conclusion be?

Give a conclusion in the context of this example.

If we calculated the difference in mean cholesterol to be \(-14.5\), why are we concluding that there is no true difference? \(-14.5\) isn’t 0, explain.
Hypothesis Testing and your HW

→ What are the four parts of a hypothesis test?

→ Do you need to include all 4 parts on your HW even if the problem doesn’t specifically ask for them?

→ How can you identify a hypothesis test problem?

→ What are the important parts of a hypothesis test conclusion?

1.

2.

3.

4.

5.