Comparison of Two Independent Samples

→ Notation:

![Diagram showing notation for Population 1 to Sample 2 and Population 2 to 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 \((\bar{x}_1 - \bar{x}_2)\) is

mean = 
standard deviation = 

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)?
\( (\bar{x}_1 - \bar{x}_2) \) estimates

**Standard Error of \( (\bar{x}_1 - \bar{x}_2) \)**

\( SE_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{SE_1^2 + SE_2^2} \)

*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>( \bar{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.

\[ \text{group 1 = } \bar{y}_1, \quad \text{group 2 = } \bar{y}_2 \]

**CI for \( (\mu_1 - \mu_2) \)**

A 100\((1-\alpha)\)% confidence interval for \( (\mu_1 - \mu_2) \) p.468

\[
(\bar{y}_1 - \bar{y}_2) \pm t(SE_{\bar{y}_1 - \bar{y}_2})
\]

Where

\[
V_1 = \frac{s_1^2}{n_1}, \quad V_2 = \frac{s_2^2}{n_2}, \quad OR \text{ df } = \frac{(SE_1^2 + SE_2^2)^2}{SE_1^4/(n_1-1) + SE_2^4/(n_2-1)}
\]

→ Assumptions: the two samples are independently selected random samples; the sample sizes are large (both \( \geq 30 \)) or the population distributions are approximately normal

→ 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 than 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

→ What are the general forms of the hypotheses?

Ho: $H_0: \mu_1 - \mu_2 = \text{HypValue}$

Ha: $H_1: \mu_1 - \mu_2 \neq \text{HypValue}$

$t_s = \frac{(\bar{x}_1 - \bar{x}_2) - \text{HypValue}}{SE_{\bar{x}_1-\bar{x}_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - \text{HypValue}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

Where $V_1 = \frac{V_1^2}{(n_1 - 1)}$ and $V_2 = \frac{V_2^2}{(n_2 - 1)}$

$V_1 = \frac{s_1^2}{n_1}$ $V_2 = \frac{s_2^2}{n_2}$

OR $df = \frac{(SE_1^2 + SE_2^2)^2}{SE_1^2/(n_1 - 1) + SE_2^2/(n_2 - 1)}$

Assumptions: the two samples are independently selected random samples; the sample sizes are large (both $\geq 30$) or the population distributions are approximately normal.
When we compare two group means “nothing is going on” means:

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. Test using a significance level of 0.05.

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

Did we meet the assumptions? Explain.

Calculate the test statistic and p-value

What is the appropriate conclusion for this test?

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.

Example: cholesterol (cont’)

Suppose it is reasonable to assume that $\mu_1 < \mu_2$, in other words the researcher is hoping to show that this new medication lowers cholesterol.

If the researchers still had the same data, but specified the test as above, what would change?
If the researchers had specified an upper tailed test, what would change?

What is the difference between a correctly specified one-tailed p-value and a two-tailed p-value?

Example: A study was published in the Journal of Industrial Technology that examined the perceived importance of graphic communications curriculum as perceived by educators (n = 30) and those in industry (n = 56). The study was conducted via survey and significant topic differences included: characteristics of digital communications; develop marketing plans and research; use graphic communications terminology; explain digital photography process; evaluate desktop scanning technology. Each person was asked to evaluate a topic on a scale of 1 (not important) to 5 (important)

<table>
<thead>
<tr>
<th>Area</th>
<th>Educators (n = 30) mean</th>
<th>sd</th>
<th>Industry (n = 56) mean</th>
<th>sd</th>
<th>ts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify characteristics of digital</td>
<td>4.07</td>
<td>0.90</td>
<td>3.64</td>
<td>0.86</td>
<td>2.12**</td>
</tr>
<tr>
<td>communications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop marketing plans and research</td>
<td>2.86</td>
<td>0.93</td>
<td>3.45</td>
<td>1.01</td>
<td>2.59*</td>
</tr>
<tr>
<td>Use graphic communications terminology</td>
<td>4.46</td>
<td>0.64</td>
<td>3.88</td>
<td>1.05</td>
<td>2.73**</td>
</tr>
<tr>
<td>Explain digital photography process</td>
<td>3.86</td>
<td>1.03</td>
<td>3.11</td>
<td>1.23</td>
<td>2.83**</td>
</tr>
<tr>
<td>Evaluate desktop scanning technology</td>
<td>4.00</td>
<td>0.96</td>
<td>3.34</td>
<td>1.08</td>
<td>2.76**</td>
</tr>
<tr>
<td></td>
<td>* p&lt;0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>**p&lt;0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Which comparison had the most statistical significance?

Do you think these were carried out as one tailed or two tailed tests? Can you tell from this info?

Why didn't the researcher report the exact p-values?
Comparison of Paired Samples

→ What does it mean to say that data is paired?

Paired data

→ How do we calculate the differences?

\[
\bar{x}_d = \frac{\sum d}{n_d} = \bar{x}_1 - \bar{x}_2
\]

→ What does \( \bar{x}_d \) estimate?

→ What is \( s_d \)?

→ What is \( n_d \)?

Example: Suppose we measure the thickness of plaque (mm) in the carotid artery of 10 randomly selected patients with mild atherosclerotic disease. Two measurements are taken, thickness before treatment with Vitamin E (baseline) and after two years of taking Vitamin E daily.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Before</th>
<th>After</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.66</td>
<td>0.60</td>
<td>0.06</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
<td>0.65</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>0.85</td>
<td>0.79</td>
<td>0.06</td>
</tr>
<tr>
<td>4</td>
<td>0.62</td>
<td>0.63</td>
<td>-0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.59</td>
<td>0.54</td>
<td>0.05</td>
</tr>
<tr>
<td>6</td>
<td>0.63</td>
<td>0.55</td>
<td>0.08</td>
</tr>
<tr>
<td>7</td>
<td>0.64</td>
<td>0.62</td>
<td>0.02</td>
</tr>
<tr>
<td>8</td>
<td>0.70</td>
<td>0.67</td>
<td>0.03</td>
</tr>
<tr>
<td>9</td>
<td>0.73</td>
<td>0.68</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>0.68</td>
<td>0.64</td>
<td>0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>0.682</td>
<td>0.637</td>
<td>0.045</td>
</tr>
<tr>
<td>After</td>
<td>0.0742</td>
<td>0.0709</td>
<td>0.0264</td>
</tr>
</tbody>
</table>

→ Why would we consider this data paired?
→ Why would we want to use pairing in this example?

→ Calculate the mean of the differences and the standard error for that estimate.

→ Can we calculate $s_d$ as $s_1 - s_2$? Why or why not?

→ What effect does proper pairing have of the standard deviation (ie. $s_d$ rather than $s_1$ and $s_2$)? Explain.

**Paired CI for $\mu_d$**

A 100$(1-\alpha)$% confidence interval for $\mu_d$

$$\bar{d} \pm t \left(\frac{s_d}{\sqrt{n_d}}\right)$$

where $df = n_d - 1$

- Assumptions: The samples are paired; The differences are random; The number of pairs (nd) is large $\geq 30$ or the differences are approximately normal

*Example: Vitamin E (cont')*

→ Calculate a 90% confidence interval for the true mean difference in plaque thickness before and after treatment with Vitamin E.

→ Write a conclusion in the context of this example.
Should the researchers consider this confidence interval meaningful? Good or bad news?

Does the zero rule work with this type of CI? Explain.

Paired t test

What are the general forms of the null and alternative hypotheses?

How do you know when to use $H_a: <$ rather than $H_a: >$?

$$t_s = \frac{\bar{x}_d - 0}{s_d / \sqrt{n_d}}$$

Where df = nd – 1

- Assumptions: The samples are paired; The differences are random; The number of pairs (nd) is large ≥ 30 or the differences are approximately normal

Example: Vitamin E (cont’)

Do the data provide enough evidence to indicate that there is a difference in plaque before and after treatment with vitamin E for two years? Test using $\alpha = 0.10$

State the appropriate hypotheses for this test in symbols.

State the appropriate hypotheses for this test in words.

Calculate the test statistic.
→ What does the sign of this test statistic imply about the mean difference?

→ Calculate the p-value and state the meaningful conclusion for this test.

Corresponding Minitab Output:

**Paired T-Test and CI: Before, After**

Paired T for Before - After

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>10</td>
<td>0.682000</td>
<td>0.074207</td>
<td>0.023466</td>
</tr>
<tr>
<td>After</td>
<td>10</td>
<td>0.637000</td>
<td>0.070875</td>
<td>0.022413</td>
</tr>
<tr>
<td>Difference</td>
<td>10</td>
<td>0.045000</td>
<td>0.026352</td>
<td>0.008333</td>
</tr>
</tbody>
</table>

90% CI for mean difference: (0.029724, 0.060276)
T-Test of mean difference = 0 (vs not = 0): T-Value = 5.40  P-Value = 0.000

Results of ignoring pairing

**Example**: Vitamin E (cont')

→ Calculate the test statistic and p-value as if this were an independent t test

→ Why is this test statistic so much smaller than the test statistic for the paired test? Explain.
→ How does $SE_{\bar{y_1} - \bar{y_2}}$ compare to $SE_\bar{y}$? Explain any difference.

→ Calculate a 90% confidence interval for $\mu_1 - \mu_2$

→ How does the significance of this interval compare to the paired 90% CI calculated previously? Explain why this is happening.

→ Is there anything that is better in the independent CI when compared to the paired CI? Explain.

Comparison of MTB Output

**Paired T-Test and CI: Before, After**

Paired T for Before - After

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>10</td>
<td>0.682</td>
<td>0.074</td>
<td>0.023</td>
</tr>
<tr>
<td>After</td>
<td>10</td>
<td>0.637</td>
<td>0.071</td>
<td>0.022</td>
</tr>
<tr>
<td>Difference</td>
<td>10</td>
<td>0.045</td>
<td>0.026</td>
<td>0.008</td>
</tr>
</tbody>
</table>

90% CI for mean difference: (0.029724, 0.060276)
T-Test of mean difference = 0 (vs not = 0): T-Value = 5.40  P-Value = 0.000

**Two-Sample T-Test and CI: Before, After**

Two-sample T for Before vs After

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>10</td>
<td>0.682</td>
<td>0.074</td>
<td>0.023</td>
</tr>
<tr>
<td>After</td>
<td>10</td>
<td>0.637</td>
<td>0.070</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Difference = mu (Before) - mu (After)
Estimate for difference: 0.045000
90% CI for difference: (-0.011450, 0.101450)
T-Test of difference = 0 (vs not =): T-Value = 1.39  P-Value = 0.183  DF = 17
→ Why is the SE smaller for correctly paired data?

→ What is the price of pairing?

Purpose of Pairing
→ Pairing is used to reduce ___________ and ___________ precision.

→ (TRUE or FALSE) Pairing needs to be carried out before the data are observed.

Paired vs. Unpaired
→ If the variability between subjects is large then use...

→ If the experimental units are homogenous then use...