Comparison of Paired Samples

- In chapter 7 we discussed how to compare two independent samples.
- In chapter 9 we discuss how to compare two samples that are paired.
  - In other words, the two samples are not independent, Y_1 and Y_2 are linked in some way, usually by a direct relationship.
  - For example, measure the weight of subjects before and after a six-month diet.

Paired data

- To study paired data, we would like to examine the differences between each pair.
  - \( d = Y_1 - Y_2 \)
  - Each \( Y_1, Y_2 \) pair will have a difference calculated.
- With the paired t-test, we would like to concentrate our efforts on this difference data.
  - We will be calculating the mean of the differences and the standard error of the differences.

Paired data (cont')

- The mean of the differences is calculated just like the one-sample mean we calculated in chapter 2:
  \[ \bar{d} = \frac{\sum d}{n_d} \]
  - It also happens to be equal to the difference in the sample means – this is similar to the t-test.
- This sample mean differences is an estimate of the population mean difference:
  \[ \mu_d = \mu_1 - \mu_2 \]

Paired data (cont')

- Because we are focusing on the differences, we can use the same reasoning as we did for a single sample in chapter 6 to calculate the standard error:
  - Aka. the standard deviation of the sampling distribution of \( \bar{d} \)
- Recall:
  \[ SE = \frac{s_d}{\sqrt{n_d}} \]
- Using similar logic:
  \[ SE_d = \frac{s_d}{\sqrt{n_d}} \]
  - Where \( s_d \) is the standard deviation of the differences and \( n_d \) is the sample size of the differences.

Paired data (cont')

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>0.68</td>
<td>0.60</td>
<td>0.08</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
<td>0.65</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>0.65</td>
<td>0.65</td>
<td>0.00</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>

What makes this paired data rather than independent data?

<table>
<thead>
<tr>
<th>Subject</th>
<th>Before</th>
<th>After</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.68</td>
<td>0.60</td>
<td>0.08</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
<td>0.65</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>0.65</td>
<td>0.65</td>
<td>0.00</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>

Why would we want to use pairing in this example?
Paired data (cont’)

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

\[ \bar{d} = 0.045 \]
\[ s_d = 0.0264 \]
\[ SE_d = \frac{s_d}{\sqrt{n}} = \frac{0.0264}{\sqrt{10}} = 0.00833 \]

Paired CI for \( \mu_d \)

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

\[ \bar{d} \pm t_{df} \left( \frac{SE_d}{\sqrt{n}} \right) \]
where \( df = n_d - 1 \)

– Very similar to the one sample confidence interval we learned in section 6.3, but this time we are concentrating on a difference column rather than a single sample

Paired CI for \( \mu_d \) (cont’)

Example: Vitamin E (cont’)
Calculate a 90% confidence interval for the true mean difference in plaque thickness before and after treatment with Vitamin E

\[ \bar{d} \pm t_{df} \left( \frac{SE_d}{\sqrt{n}} \right) \]
\[ = 0.045 \pm t_{9}(0.00833) \]
\[ = 0.045 \pm 1.383(0.00833) \]
\[ = (0.0297, 0.0603) \]

Paired t test

• Of course there is also a hypothesis test for paired data
  • #1 Hypotheses:
    \[ H_0: \mu_d = 0 \]
    \[ H_a: \mu_d < 0 \quad \text{or} \quad H_a: \mu_d > 0 \]
  • #2 test statistic
    \[ t = \frac{\bar{d} - 0}{SE_d} \]
    Where \( df = n_d - 1 \)
  • #3 p-value and #4 conclusion similar idea to that of the independent t test

Paired t test (cont’)

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? Testing using \( \alpha = 0.10 \)

\[ H_0: \mu_d = 0 \quad \text{(mean thickness in plaque is the same before and after treatment with Vitamin E)} \]
\[ H_a: \mu_d \neq 0 \quad \text{(mean thickness in plaque after treatment is different than before treatment with Vitamin E)} \]

\[ t = \frac{0.045 - 0}{0.00833} = 5.402 \]
\[ df = 10 - 1 = 9 \]
\[ p < 2(0.0005) = 0.001, \text{ so we reject } H_0. \]
Paired t test (cont')

CONCLUSION: These data show that the true mean thickness of plaque after two years of treatment with Vitamin E is statistically significantly different than before the treatment ($p < 0.001$).

In other words, vitamin E appears to be effective in changing carotid artery plaque after treatment

- May have been better to conduct this as an upper-tailed test because we would hope that vitamin E will reduce clogging
- However, researchers need to make this decision before analyzing data

Paired T-Test and CI: Before, After

Paired T for Before - After

N  Mean  StDev  SE Mean
Before 10  0.6820  0.0742  0.0235
After 10  0.6370  0.0709  0.0224

Difference 10  0.0450  0.0264  0.0083

90% CI for mean difference: (0.029, 0.060)
T-Test of mean difference = 0 (vs not = 0): T-Value = 5.40  P-Value = 0.000

Results of Ignoring Pairing

- Suppose we accidentally analyzed the groups independently (like an independent t-test) rather than a paired test?
  - Keep in mind this would be an incorrect way of analyzing the data
- How would this change our results?

Paired T-Test and CI: Before, After

Paired T for Before - After

N  Mean  StDev  SE Mean
Before 10  0.6820  0.0742  0.0235
After 10  0.6370  0.0709  0.0224

Difference 10  0.0450  0.0264  0.0083

90% CI for mean difference: (0.029, 0.060)
T-Test of mean difference = 0 (vs not = 0): T-Value = 5.40  P-Value = 0.000

Results of Ignoring Pairing (cont')

- What happens to a CI?
  - Calculate a 90% confidence interval for $\mu_1 - \mu_2$
    
    $\bar{y}_1 - \bar{y}_2 \pm t_{\alpha/2} s_d / \sqrt{n}$
    
    $= (0.682 - 0.637) \pm 1.753(0.0325)$
    
    $= 0.045 \pm (0.1016)$
    
    How does the significance of this interval compare to the paired 90% CI (0.03 mm and 0.06 mm)?
  - Why is this happening?
  - Is there anything better about the independent CI? Is it worth it in this situation?
Results of Ignoring Pairing (cont’)

• Why would the SE be smaller for correctly paired data?
  – If we look at the within each sample at the data we notice variation from one subject to the next
  – This information gets incorporated into the SE for the independent t-test via $s_1$ and $s_2$
  – The original reason we paired was to try to control for some of this inter-subject variation
  – This inter-subject variation has no influence on the SE for the paired test because only the differences were used in the calculation.
• The price of pairing is smaller df.
  – However, this can be compensated with a smaller SE if we had paired correctly.

Conditions for the validity of the paired t test

• Conditions we must meet for the paired t test to be valid:
  – It must be reasonable to regard the differences as a random sample from some large population
  – The population distribution of the differences must be normally distributed.
  – The methods are approximately valid if the population is approximately normal or the sample size $n_0$ is large.
  – These conditions are the same as the conditions we discussed in chapter 6.

Conditions for the validity of the paired t test (cont’)

• How can we check:
  – check the study design to assure that the differences are independent (ie no hierarchical structure within the d's)
  – create normal probability plots to check normality of the differences
  – NOTE: p.355 summary of formulas

The Paired Design

• Ideally in the paired design the members of a pair are relatively similar to each other
• Common Paired Designs
  – Randomized block experiments with two units per block
  – Observational studies with individually matched controls
  – Repeated measurements on the same individual
  – Blocking by time – formed implicitly when replicate measurements are made at different times.
• IDEA of pairing: members of a pair are similar to each other with respect to extraneous variables

The Paired Design (cont’)

Example: Vitamin E (cont’)
  – Same individual measurements made at different times before and after treatment (controls for within patient variation).
Example: Growing two types of bacteria cells in a petri dish replicated on 20 different days.
  – These are measurements on 2 different bacteria at the same time (controls for time variation).

Purpose of Pairing

• Pairing is used to reduce bias and increase precision
  – By matching/blocking we can control variation due to extraneous variables.
• For example, if two groups are matched on age, then a comparison between the groups is free of any bias due to a difference in age distribution
• Pairing is a strategy of design, not analysis
  – Pairing needs to be carried out before the data are observed
  – It is not correct to use the observations to make pairs after the data has been collected
Paired vs. Unpaired

• If the observed variable Y is not related to factors used in pairing, the paired analysis may not be effective
  – For example, suppose we wanted to match subjects on race/ethnicity and then we compare how much ice cream men vs. women can consume in an hour
• The choice of pairing depends on practical considerations (feasibility, cost, etc…) and on precision considerations
  – If the variability between subjects is large, then pairing is preferable
  – If the experimental units are homogenous then use the independent t

The Sign Test

• The sign test is a non-parametric version of the paired t test
• We use the sign test when pairing is appropriate, but we can’t meet the normality assumption required for the t test
• The sign test is not very sophisticated and therefore quite easy to understand
• Sign test is also based on differences
  \[ d = Y_1 - Y_2 \]
  The information used by the sign test from this difference is the sign of d (+ or -)

The Sign Test (cont’)

• #1 Hypotheses:
  Ho: the distributions of the two groups is the same
  Ha: the distribution of group 1 is less than group 2
  or Ha: the distribution of group 1 is greater than group 2
• #2 Test Statistic B_s

  Method:
  1. Find the sign of the differences
  2. Calculate \( N_+ \) and \( N_- \)
  3a. If Ha is non-directional, \( B_s \) is the larger of \( N_+ \) and \( N_- \)
  3b. If Ha is directional, \( B_s \) is the \( N \) that jives with the direction of Ha (if Ha: \( Y_1 < Y_2 \) then we expect a larger \( N_- \), if Ha: \( Y_1 > Y_2 \) then we expect a larger \( N_+ \))

  NOTE: If we have a difference of zero it is not included in \( N_+ \) or \( N_- \), therefore \( n_d \) needs to be adjusted

The Sign Test (cont’)

• #3 p-value:
  Table 7 p.684
  Similar to the WMW
  Use the number of pairs with "quality information"
• #4 Conclusion:
  Similar to the WMW
  Do NOT mention any parameters!

Example: Twelve sets of identical twins are given psychological tests to determine whether the first born of the set tends to be more aggressive than the second born. Each twin is scored according to aggressiveness, a higher score indicates greater aggressiveness.
• Because of the natural pairing in a set of twins these data can be considered paired.
The Sign Test (cont’)

Do the data provide sufficient evidence to indicate that the first born of a set of twins is more aggressive than the second? Test using \( \alpha = 0.05 \).

\( H_0: \) The aggressiveness is the same for 1st born and 2nd born twins

\( H_a: \) The aggressiveness of the 1st born twin tends to be more than 2nd born.

NOTE: Directional Ha (we’re expecting higher scores for the 1st born twin), this means we predict that most of the differences will be positive

\( N_+ = \) number of positive = 7
\( N_- = \) number of negative = 4
\( n_d = \) number of pairs with useful info = 11

\( B_s = N_+ = 7 \) (because of directional alternative)

\( P > 0.10, \) Fail to reject \( H_0 \)

CONCLUSION: These data show that the aggressiveness of 1st born twins is not significantly greater than the 2nd born twins (\( P > 0.10 \)).

Applicability of the Sign Test

• Valid in any situation where d’s are independent of each other
• Distribution-free, doesn’t depend on population distribution of the d’s
  – although if d’s are normal the t-test is more powerful
• Can be used quickly and can be applied on data that do not permit a t-test

Practice

• Suppose Ha: one-tailed, \( n_d = 11 \) and \( B_s = 10 \), find the appropriate p-value

\[ 0.005 < p < 0.01 \]

pick the smallest p-value for \( B_s = 10 \) and bracket

– NOTE: Distribution for the sign test is discrete, so probabilities are somewhat smaller (similar to WMW)
Applicability of the Sign Test (cont')

Ho: The escape times (sec.) of rats are the same before and after training.
Ha: The escape times (sec.) of rats are different before and after training.

\[ N_+ = 9 \quad N_- = 1 \quad n_d = 10 \]

\[ B_s = \text{larger of } N_+ \text{ or } N_- = 9 \]

\[ 0.02 < p < 0.05, \text{ reject } Ho \]

CONCLUSION: These data show that the escape times (sec.) of rats before training are different from the escape times after training \((0.02 < p < 0.05)\).