Matched Pairs (Dependent Samples)

*t-distibutions: confidence intervals and hypothesis tests*

Matched Pairs are when two samples are taken from two populations such that the samples are *dependent* on each other and can be matched up. This is not a two-sample distribution, rather, we treat the differences of the two samples as one sample. We therefore apply t-procedures to the one population of differences.

1. Imagine that you want to compare two sets of golf club brands: Titleist and Calloway. Suppose that we want to show that Titleist hits farther than Calloway. We do this by comparing 10 random pairs of golf clubs where each pair consists of one Titleist club and one Calloway club of the same type. Each pair is hit by a different randomly chosen professional golfer (this makes the pairs *dependent* since one pair can be matched up through the one golfer hitting those two clubs). Measurements in the number of yards hit are recorded below:

<table>
<thead>
<tr>
<th>Golfer</th>
<th>Titleist</th>
<th>Calloway</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>188</td>
<td>194</td>
<td>-6</td>
</tr>
<tr>
<td>2</td>
<td>201</td>
<td>197</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>195</td>
<td>195</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>195</td>
<td>190</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>202</td>
<td>195</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>175</td>
<td>178</td>
<td>-3</td>
</tr>
<tr>
<td>7</td>
<td>182</td>
<td>172</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>188</td>
<td>184</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>193</td>
<td>201</td>
<td>-8</td>
</tr>
<tr>
<td>10</td>
<td>194</td>
<td>192</td>
<td>2</td>
</tr>
</tbody>
</table>

Is this sample of differences sufficient evidence to claim that Titleist hits farther than Calloway? Choose $\alpha = 0.05$.

**ANSWER:**

First note that the first population consists of the number of yards a Titleist club is hit, and the second population consists of the number of yards a Calloway club is hit. Therefore, denote

$\mu_T =$ the **average** number of yards a Titleist club hits

$\mu_C =$ the **average** number of yards a Calloway club hits

We will assume $H_0 : \mu_T = \mu_C$

and we want to show $H_a : \mu_T > \mu_C$

It is important to note that we always subtracted Calloway from Titleist in our sample. If we subtract off $\mu_C$ then it follows that

We will assume $H_0 : \mu_T - \mu_C = 0$

and we want to show $H_a : \mu_T - \mu_C > 0$

Define the average of the population of differences as $\mu_d = \mu_T - \mu_C$. Then

We will assume $H_0 : \mu_d = 0$

and we want to show $H_a : \mu_d > 0$
Steps conducted for the golf club Hypothesis Test:

1. Let $\mu_d = \mu_T - \mu_C$.
   
   $H_0 : \mu_d = 0$
   
   $H_a : \mu_d > 0$

2. $\alpha = 0.05$

3. Statistics taken from the difference column ONLY!
   
   $\bar{x} = 1.5$
   
   $S_x \approx 5.74$
   
   $n = 10$

4. $t = \frac{\bar{x} - \mu_0}{\frac{S_x}{\sqrt{n}}} = \frac{1.5 - 0}{\frac{5.74}{\sqrt{10}}} \approx 0.8264$

   P-value = $P(t \geq 0.8264) = tcdf(0.8264, 1E99, 9) \approx 0.215$

   Also, STAT → TESTS 2 (T-Test) yields same results.

5. P-value > $\alpha$
   
   $0.215 > 0.05$

   do NOT Reject $H_0$

6. Our sample is NOT significant evidence that the average number of yards hit by a Titleist club is greater than the average number of yards hit by a Calloway club.

2. Some researchers studying vitamin C in a corn soy blend (CSB) were also interested in a similar commodity called wheat soy blend (WSB). Both of these commodities are mixed with other ingredients and cooked. Loss of vitamin C as a result of this process was another concern of the researchers. One preparation used in Haiti called gruel (or “bouillie” in Creole) can be made from WSB, salt, sugar, milk, banana, and other optional items to improve the taste. Samples of gruel prepared in Haitian households were collected. The vitamin C content (in milligrams per 100 grams of blend, dry basis) was measured before and after cooking. Here are the results:

<table>
<thead>
<tr>
<th>Sample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>73</td>
<td>79</td>
<td>86</td>
<td>88</td>
<td>78</td>
</tr>
<tr>
<td>After</td>
<td>20</td>
<td>27</td>
<td>29</td>
<td>36</td>
<td>17</td>
</tr>
</tbody>
</table>

Carry out a significance test for these data at 5% to see if vitamin C is lost after cooking.
Steps conducted for the vitamin C and cooking Hypothesis Test:

1. Let $\mu_d = \mu_B - \mu_A$.
   
   $H_0 : \mu_d = 0$
   
   $H_a : \mu_d > 0$

2. $\alpha = 0.05$

3. Statistics taken for differences only (BEFORE—AFTER).
   
   $\bar{x} = 55$
   
   $S_x \approx 3.937$
   
   $n = 5$

4. $t = \frac{\bar{x} - \mu_0}{\frac{S_x}{\sqrt{n}}} = \frac{55 - 0}{\frac{3.937}{\sqrt{5}}} \approx 31.24$

   P-value = $P(t \geq 31.24) = \text{tcdf}(31.24, 1E99, 4) \approx 3.13E-6$

   Also, STAT TESTS 2 (T-Test) yields same results.

5. P-value $\leq \alpha$
   
   $3.13E-6 \leq 0.05$

   Reject $H_0$

6. There is sufficient evidence that vitamin C is lost after cooking (on average).