We continue to address the issue of comparing means between two groups, but now we’ll consider data collected with a different (and often better) study design.

**Example 8-1: Distracted Driving**
Researchers at the University of Utah (Strayer and Johnston, 2001) asked student volunteers to use a machine that simulated driving situations. At irregular intervals, a target would flash red or green. Participants were instructed to press a “brake button” as soon as possible when they detected a red light. The machine would calculate the mean reaction time to the red flashing targets for each student in milliseconds.

The students were given a warm-up period to familiarize themselves with the driving simulator. Then the researchers had each student use the driving simulation machine while talking on a cell phone about politics to someone in another room and then again with music or a book-on-tape playing in the background (control). The students were randomly assigned as to whether they used the cell phone or the control setting for the first trial. The data on reaction times (in milliseconds) for 16 students appears below:

<table>
<thead>
<tr>
<th>Subject</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone</td>
<td>636</td>
<td>623</td>
<td>615</td>
<td>672</td>
<td>601</td>
<td>600</td>
<td>542</td>
<td>554</td>
<td>543</td>
<td>520</td>
<td>609</td>
<td>559</td>
<td>595</td>
<td>565</td>
<td>554</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>604</td>
<td>556</td>
<td>540</td>
<td>522</td>
<td>459</td>
<td>544</td>
<td>513</td>
<td>547</td>
<td>556</td>
<td>531</td>
<td>599</td>
<td>537</td>
<td>619</td>
<td>536</td>
<td>554</td>
<td></td>
</tr>
</tbody>
</table>

**a)** Is this an observational study or a randomized experiment? Explain.

**b)** What’s the primary difference in how this study was designed and conducted, as compared to the sleep deprivation study?

The study presented here is an example of a **matched pairs** design. Each student in the study experienced both treatments (driving while conversing on the cell phone and driving with background music/book.) Randomization was used in this study, but only to randomly determine which treatment each student experienced first. This type of design is preferable to a completely randomized design here because the pairing helps to control for natural differences in reaction times across subjects. If a subject performs differently on the two treatments, we feel much more comfortable attributing that difference to the treatment than if we compared two different people.

**c)** What are the observational units of this study? What makes sense to use as the response variable?
d) We will consider the response variable for this study to be the difference in reaction time (cell phone – control) for each student. If it is actually the case that cell phone use delays reaction time, what should we see in these differences?

A theory-based approach to analyzing paired data is to apply a one-sample $t$-test/interval (now called a paired $t$-test/interval) to the sample differences.

- Null hypothesis $H_0$: $\mu_{\text{diff}} = 0$, where $\mu_{\text{diff}}$ represents the population mean of the differences.
- Test statistic: $t = \frac{\bar{x}_{\text{diff}}}{s_{\text{diff}} / \sqrt{n_{\text{diff}}}}$, where
  - $\bar{x}_{\text{diff}}$ is the sample mean of the differences
  - $s_{\text{diff}}$ is the sample standard deviation of the differences,
  - $n_{\text{diff}}$ is the sample size of the differences
- The $p$-value is based on the $t$-distribution with $(n-1)$ degrees of freedom
- Confidence interval for $\mu_{\text{diff}}$: $\bar{x}_{\text{diff}} \pm t^* s_{\text{diff}} / \sqrt{n_{\text{diff}}}$
- Technical conditions
  - Random sample (of differences) from the population
  - Large sample size ($n_{\text{diff}} \geq 30$) or normal distribution (of differences)

e) Express the null and alternative hypotheses to be tested here (in symbols and in words).

f) Enter the sample differences (DrivingDifferences.txt) into the Descriptive Statistics applet. Examine a dotplot and boxplot of the differences. Comment on what these reveal, particularly with regard to the research question that motivated this study.

g) Calculate the relevant statistics for conducting a paired $t$-test from these data. Report these, along with appropriate symbols.
h) Calculate (by hand) the value of the \( t \)-test statistic. Also interpret this value. Then determine the p-value of the \( t \)-test as accurately as possible.

i) Summarize your conclusion from the paired \( t \)-test.

j) Produce (by hand) a 95% confidence interval for the population mean difference in reaction times.

k) Interpret what this interval reveals.

l) Verify your results with the Theory-Based Inference applet.

m) How many of the 16 subjects reacted more quickly in the control condition rather than while talking on a cell phone?

n) Let the random variable \( X \) represent the number who would respond more quickly in the control condition. What probability distribution would \( X \) have, if in fact there was no difference between the two treatment conditions?

o) Use your answer to n) to determine the probability that \( X \) (the number who would respond more quickly in the control condition) would be at least as large as your answer to m), if in fact there was no difference between the two treatment conditions.
p) Summarize your conclusion from the probability you calculated in o).

- The **sign test**, based on a binomial distribution, provides an alternative to a paired $t$-test.
  - The sign test requires fewer assumptions but ignores much information.

**Example 8-2: Marriage Ages**
A student of mine wanted to test whether husbands tend to be older than their wives on average. He went to the county courthouse and took a sample of 24 couples who had applied for marriage licenses, recording the ages of the man and woman in each case (*MarriageAges.txt*).

Some summary statistics are:

<table>
<thead>
<tr>
<th></th>
<th>Sample size</th>
<th>Sample mean</th>
<th>Sample std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Husbands</td>
<td>24</td>
<td>35.71</td>
<td>14.56</td>
</tr>
<tr>
<td>Wives</td>
<td>24</td>
<td>33.83</td>
<td>13.56</td>
</tr>
</tbody>
</table>

a) Explain why this information is not sufficient to conduct the appropriate test.

Some more summary statistics and a dotplot, with differences calculated as husband’s age minus wife’s age:

<table>
<thead>
<tr>
<th></th>
<th>Sample size</th>
<th>Sample mean</th>
<th>Sample std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences</td>
<td>24</td>
<td>1.875</td>
<td>4.812</td>
</tr>
</tbody>
</table>

b) Determine and interpret a 90% confidence interval for the population *mean difference* in ages between husbands and wives. Also check and comment on the technical conditions for applying this procedure. Finally, comment on whether 90% of the age differences in the sample fall within this interval.
c) Conduct a paired $t$-test of whether the sample data provide strong evidence that husbands tend to be older than their wives in the population. Include all components of a significance test. Indicate the test decision at the .10 significance level, and summarize your conclusion.

d) What would change in your results above if the differences had been calculated/subtracted in the opposite order (wife – husband rather than husband – wife)?

e) Was the student wise for gathering paired data rather than independent-samples data to investigate his research question? Explain. (Hint: What varies a lot, and what varies not so much in this context?)

The pairing is effective here because there is a lot of variation in the ages of people who apply for marriage licenses, but there is a strong correlation between the ages of the husband and wife within a couple. Thus, there is much less variation in the differences in ages than there is in the ages themselves.

Notice that a paired analysis follows from the data having been collected according to a matched-paired design in the first place.

f) Would it be appropriate to compare the 65 men’s and 65 women’s body temperatures with a paired $t$-test? Explain.

Example 8-3: Melting Morsels
Suppose that you want to investigate whether there is a difference in the melting times of semi-sweet milk chocolate chips and peanut butter chips. You could take a group of volunteers, randomly assign half to take a chocolate chip and the other half to take a peanut butter chip, and then time how long it takes before the chip melts in their mouth.
Suppose that you want to investigate whether there is a difference in the melting times of semi-sweet milk chocolate chips and peanut butter chips. You could take a group of volunteers, randomly assign half to take a chocolate chip and the other half to take a peanut butter chip, and then time how long it takes before the chip melts in their mouth.

a) Would this be an independent-samples or a matched-pairs design? Explain.

b) Describe how you could alter the data collection process to make this a paired design.

c) Explain why it’s a good idea to conduct this study with a matched-paired design.

d) Describe how (and why) a paired design to investigate this issue should make use of randomness.

Example 8-4: Waterproofing Boots
Suppose that you want to compare two methods of water-proofing boots and that you have recruited 100 subjects to participate in an experiment.

a) Suggest a better experimental design than randomly assigning 50 subjects to wear boots with each type of water-proofing.

b) What is the primary advantage of the matched-pairs design in this study?

c) What would be the key step in analyzing the resulting data?