Today we continue to examine how to compare two groups, but we turn our attention to studies involving a \textit{quantitative} response variable. You will again see that the reasoning process of statistical significance, as summarized by a p-value, is the same as we have studied previously, as is the concept of a confidence interval for a population parameter.

\textbf{Example 20-21: Got a Tip?}\n
Can waitresses increase their tips simply by introducing themselves by name when they greet customers? Garrity and Degelman (1990) report on a study in which a waitress collected data on two-person parties that she waited on during Sunday brunch (with a fixed price of $23.21) at a Charley Brown’s restaurant in southern California. For each party the waitress used a random mechanism to determine whether to give her name as part of her greeting or not. Then she kept track of how much the party gave for a tip at the end of their meal.

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

b) Identify and classify the explanatory and response variables.

c) State the null and alternative hypotheses, in symbols and in words, for testing the waitress’ conjecture.

We can use a two-sample \textit{t}-test for comparing two \textit{means}:

- The null hypothesis is $H_0: \mu_1 = \mu_2$
- The alternative hypothesis is $H_a: \mu_1 > \mu_2$ or $H_a: \mu_1 < \mu_2$ or $H_a: \mu_1 \neq \mu_2$.
- The test statistic is:  
  \[ t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]
- The \textit{p}-value is the area under the \textit{t}-distribution (with the smaller of $n_1-1$ and $n_2-1$ degrees of freedom) more extreme than the test statistic, in the direction indicated by the alternative hypothesis.
- The test decision is to reject $H_0$ whenever \textit{p}-value $< \alpha$ (significance level).
- A confidence interval for estimating $\mu_1 - \mu_2$ is given by: \( (\bar{x}_1 - \bar{x}_2) \pm t * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \).
- The technical conditions required for this procedure to be valid are that:
  - the data come from independent random samples or random assignment to groups
  - the sample sizes are large ($\geq 30$ in each group) or the populations follow normal distributions

\textcopyright Fall 2015
The sample mean tip amount for the 20 parties to which the waitress gave her name was $5.44, with a standard deviation of $1.75. These statistics were $3.49 and $1.13, respectively, for the 20 parties to which the waitress did not give her name.

e) Use this information to calculate the test statistic and p-value (by hand).

f) Interpret the p-value.

g) What test decision would you reach at the $\alpha = .05$ level? What does this decision mean in the context of this study?

h) Do you have enough information to check whether the technical conditions of the two-sample $t$-test are satisfied here? If so, check them. If not, explain what additional information you would request from the waitress.

i) Calculate a 95% confidence interval for the difference in population mean tip amounts between the two experimental treatments (giving name or not). Also interpret what the interval means in this context.

j) Is the confidence interval consistent with the test decision? Explain.

k) Use Minitab (Stat> Basic Statistics> 2-Sample t...) or the Theory-Based Inference applet to confirm your calculations.

l) Regardless of whether or not the technical conditions are met, summarize your conclusions from this test. Be sure to comment on issues of causation and generalizability.
Example 20-2: Body Temperatures
Recall the data (BodyTemps.mtw) on body temperatures for a sample of healthy adult males and females.

a) Examine and comment on comparative dotplots and boxplots of the body temperatures between the two sexes.

b) Calculate summary statistics from the sample data. Use these to conduct a $t$-test of whether the data provide strong evidence that the two sexes differ with regard to average body temperature. Include all components of the test, and also check the technical conditions. Summarize your conclusions.

c) Use summary statistics from the sample data to construct a 95% confidence interval for the difference in population means. Interpret the interval, paying particular attention to whether it includes the value zero.

d) Verify your calculations with Minitab (Stat> Basic statistics> 2-sample t) or the Theory-Based Inference applet.

e) Summarize your conclusions from this analysis.
Example 20-3: SAT coaching (hypothetical)
Suppose that 5000 students are randomly assigned to either take an SAT coaching course or not, with the following results in their improvements in SAT scores:

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sample mean</th>
<th>Sample SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coaching group:</td>
<td>2500</td>
<td>46.2</td>
</tr>
<tr>
<td>Control group:</td>
<td>2500</td>
<td>44.4</td>
</tr>
</tbody>
</table>

a) Use Minitab (Stat> Basic statistics> 2-sample t, summarized data) or the Theory-Based Inference applet to conduct a test of whether the sample data provide evidence that SAT coaching is helpful. State the hypotheses, and report the \( p \)-value. Draw a conclusion in the context of this study.

b) Use software to produce a 99% CI for the difference in population mean improvements between the two groups. Interpret this interval.

c) Are the test conclusion and CI consistent with each other? Explain.

d) Do the sample data provide very strong evidence that SAT coaching is helpful? Explain whether the \( p \)-value or the CI helps you to decide.

e) Do the sample data provide strong evidence that SAT coaching is very helpful? Explain whether the \( p \)-value or the CI helps you to decide.

f) Explain how this (hypothetical) example relates to the difference between statistical significance and practical importance.