Last, we revisited correlation coefficients with respect to significance testing. We learned how to determine whether an observed correlation coefficient is statistically significant by using a critical values table.

In Chapters 10 to 12, we will continue our discussion of inferential statistics, looking at statistical procedures appropriate for experimental designs with two or more equivalent groups and those appropriate for designs with more than one independent variable.

**KEY TERMS**

- **z test**
- **sampling distribution**
- **standard error of the mean**
- **central limit theorem**
- **critical value**
- **region of rejection**
- **statistical power**
- **confidence interval**
- **estimated standard error of the mean**
- **chi-square ($\chi^2$) goodness-of-fit test**
- **observed frequency**
- **expected frequency**

**CHAPTER EXERCISES**

(Answers to odd-numbered exercises appear in Appendix C.)

1. A researcher is interested in whether students who attend private high schools have higher average SAT scores than students in the general population. A random sample of 90 students at a private high school is tested and has a mean SAT score of 1050. The average score for public high school students is 1000 ($\sigma = 200$).
   a. Is this a one- or two-tailed test?
   b. What are $H_0$ and $H_a$ for this study?
   c. Compute $z_{obt}$.
   d. What is $z_{cv}$?
   e. Should $H_0$ be rejected? What should the researcher conclude?
   f. Determine the 95% confidence interval for the population mean, based on the sample mean.

2. The producers of a new toothpaste claim that it prevents more cavities than other brands of toothpaste. A random sample of 60 people used the new toothpaste for 6 months. The mean number of cavities at their next checkup is 1.5. In the general population, the mean number of cavities at a 6-month checkup is 1.73 ($\sigma = 1.12$).
   a. Is this a one- or two-tailed test?
   b. What are $H_0$ and $H_a$ for this study?
   c. Compute $t_{obt}$.
   d. What is $t_{cv}$?
   e. Should $H_0$ be rejected? What should the researcher conclude?
   f. Determine the 95% confidence interval for the population mean, based on the sample mean.

3. Why does $t_{cv}$ change when the sample size changes? What must be computed to determine $t_{cv}$?

4. Henry performed a two-tailed test for an experiment in which $N = 24$. He could not find his table of $t$ critical values, but he remembered the $t_{cv}$ at $df = 13$. He decided to compare his $t_{obt}$ with this $t_{cv}$. Is he more likely to make a Type I or a Type II error in this situation?

5. A researcher hypothesizes that people who listen to music via headphones have greater hearing loss and will thus score lower on a hearing test than those in the general population. On a standard hearing test, the overall mean is 22.5. The researcher gives this same test to a random sample of 12 individuals who regularly use headphones. Their scores on the test are 16, 14, 20, 12, 25, 22, 19, 17, 17, 21, 20.
   a. Is this a one- or two-tailed test?
   b. What are $H_0$ and $H_a$ for this study?
   c. Compute $t_{obt}$.
   d. What is $t_{cv}$?
   e. Should $H_0$ be rejected? What should the researcher conclude?
f. Determine the 95% confidence interval for the population mean, based on the sample mean.

6. A researcher hypothesizes that individuals who listen to classical music will score differently from the general population on a test of spatial ability. On a standardized test of spatial ability, \( \mu = 58 \). A random sample of 14 individuals who listen to classical music is given the same test. Their scores on the test are 52, 59, 63, 65, 58, 55, 62, 63, 53, 59, 57, 61, 60, 59.
   a. Is this a one- or two-tailed test?
   b. What are \( H_0 \) and \( H_a \) for this study?
   c. Compute \( t_{obt} \).
   d. What is \( t_{cv} \)?
   e. Should \( H_0 \) be rejected? What should the researcher conclude?
   f. Determine the 95% confidence interval for the population mean, based on the sample mean.

7. When is it appropriate to use a \( \chi^2 \) test?

8. A researcher believes that the percentage of people who exercise in California is greater than the national exercise rate. The national rate is 20%. The researcher gathers a random sample of 120 individuals who live in California and finds that the number who exercise regularly is 31 out of 120.
   a. What is \( \chi^2_{obt} \)?
   b. What is \( df \) for this test?
   c. What is \( \chi^2_{cv} \)?
   d. What conclusion should be drawn from these results?

9. A teacher believes that the percentage of students at her high school who go on to college is higher than the rate in the general population of high school students. The rate in the general population is 30%. In the most recent graduating class at her high school, the teacher found that 90 students graduated and that 40 of those went on to college.
   a. What is \( \chi^2_{obt} \)?
   b. What is \( df \) for this test?
   c. What is \( \chi^2_{cv} \)?
   d. What conclusion should be drawn from these results?

CRITICAL THINKING CHECK ANSWERS

8.1
1. A sampling distribution is a distribution of sample means. Thus, rather than representing scores for individuals, the sampling distribution plots the means of samples of a set size.
2. \( \sigma_X \) is the standard deviation for a sampling distribution. It therefore represents the standard deviation for a distribution of sample means. \( \sigma \) is the standard deviation for a population of individual scores rather than sample means.
3. A \( z \) test compares the performance of a sample with the performance of the population by indicating the number of standard deviation units the sample mean is from the population mean. A \( z \)-score indicates how many standard deviation units an individual score is from the population mean.

8.2
1. Predicting that psychology majors will have higher IQ scores makes this a one-tailed test.

8.3
1. The \( z \) test is used when the sample size is greater than 30, normally distributed, and \( s \) is known. The \( t \) test is used when the sample size is smaller than 30.
KEY TERMS

independent-groups t test  
correlated-groups t test  
standard error of the  
difference scores  
standard error of the  
difference scores  
Wilcoxon matched-pairs signed-ranks T test  
Wilcoxon rank-sum test  
effect size  
Cohen’s d  
difference between means  
difference scores  

chi-square (χ²) test of independence  
phi coefficient  

CHAPTER EXERCISES

(Answers to odd-numbered exercises appear in Appendix C.)

1. A college student is interested in whether there is a difference between male and female students in the amount of time they spend studying each week. The student gathers information from a random sample of male and female students on campus. The amounts of time spent studying are normally distributed. The data are:

<table>
<thead>
<tr>
<th>Males</th>
<th>Females</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>25</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>25</td>
<td>29</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>19</td>
<td>18</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. What statistical test should be used to analyze these data?
b. Identify H₀ and Hₐ for this study.
c. Conduct the appropriate analysis.
d. Should H₀ be rejected? What should the researcher conclude?
e. If significant, compute and interpret the effect size.
f. If significant, draw a graph representing the data.
g. Determine the 95% confidence interval.

2. A student is interested in whether students who study with music playing devote as much attention to their studies as do students who study under quiet conditions (he believes that studying under quiet conditions leads to better attention). He randomly assigns participants to either the music or no-music condition and has them read and study the same passage of information for the same amount of time. Subjects are given the same 10-item test on the material. Their scores appear next. Scores on the test represent interval-ratio data and are normally distributed.

<table>
<thead>
<tr>
<th>Music</th>
<th>No Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

a. What statistical test should be used to analyze these data?
b. Identify H₀ and Hₐ for this study.
c. Conduct the appropriate analysis.
d. Should H₀ be rejected? What should the researcher conclude?
e. If significant, compute and interpret the effect size.
f. If significant, draw a graph representing the data.
g. Determine the 95% confidence interval.

3. A researcher is interested in whether participating in sports positively influences self-esteem in young girls. She identifies a group of girls who have not played sports before but are now planning to begin participating in organized sports. The researcher gives them a 50-item self-esteem inventory before they begin playing sports and administers the same test again after 6 months of playing sports. The self-esteem inventory is measured on an interval scale, with higher numbers indicating higher self-esteem. In addition, scores on the
inventory are normally distributed. The scores follow.

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>46</td>
</tr>
<tr>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>39</td>
<td>41</td>
</tr>
<tr>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>42</td>
<td>43</td>
</tr>
<tr>
<td>43</td>
<td>45</td>
</tr>
</tbody>
</table>

a. What statistical test should be used to analyze these data?
b. Identify $H_0$ and $H_a$ for this study.
c. Conduct the appropriate analysis.
d. Should $H_0$ be rejected? What should the researcher conclude?
e. If significant, compute and interpret the effect size.
f. If significant, draw a graph representing the data.
g. Determine the 95% confidence interval.

4. The researcher in exercise 2 decides to conduct the same study using a within-participants design to control for differences in cognitive ability. He selects a random sample of subjects and has them study different material of equal difficulty in both the music and no-music conditions. The study is completely counterbalanced to control for order effects. The data appear next. As before, they are measured on an interval-ratio scale and are normally distributed; he believes that studying under quiet conditions will lead to better performance.

<table>
<thead>
<tr>
<th>Music</th>
<th>No Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

a. What statistical test should be used to analyze these data?
b. Identify $H_0$ and $H_a$ for this study.
c. Conduct the appropriate analysis.
d. Should $H_0$ be rejected? What should the researcher conclude?
e. If significant, compute and interpret the effect size.
f. If significant, draw a graph representing the data.
g. Determine the 95% confidence interval.

5. A researcher is interested in comparing the maturity level of students who volunteer for community service versus those who do not. The researcher assumes that those who perform community service will have higher maturity scores. Maturity scores tend to be skewed (not normally distributed). The maturity scores appear next. Higher scores indicate higher maturity levels.

<table>
<thead>
<tr>
<th>No Community Service</th>
<th>Community Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>41</td>
</tr>
<tr>
<td>41</td>
<td>48</td>
</tr>
<tr>
<td>54</td>
<td>61</td>
</tr>
<tr>
<td>13</td>
<td>72</td>
</tr>
<tr>
<td>22</td>
<td>83</td>
</tr>
<tr>
<td>26</td>
<td>55</td>
</tr>
</tbody>
</table>

a. What statistical test should be used to analyze these data?
b. Identify $H_0$ and $H_a$ for this study.
c. Conduct the appropriate analysis.
d. Should $H_0$ be rejected? What should the researcher conclude?
e. If significant, compute and interpret the effect size.
f. If significant, draw a graph representing the data.
g. Determine the 95% confidence interval.

6. Researchers at a food company are interested in how a new spaghetti sauce made from green tomatoes (and green in color) will compare to their traditional red spaghetti sauce. They are worried that the green color will adversely affect the tastiness scores. They randomly assign subjects to either the green or red sauce condition. Participants indicate the tastiness of the sauce on a 10-point scale. Tastiness scores tend to be skewed. The scores follow.

<table>
<thead>
<tr>
<th>Red Sauce</th>
<th>Green Sauce</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

a. What statistical test should be used to analyze these data?
b. Identify $H_0$ and $H_a$ for this study.
c. Conduct the appropriate analysis.
d. Should $H_0$ be rejected? What should the researcher conclude?

7. Imagine that the researchers in exercise 6 want to conduct the same study as a within-subjects design. Participants rate both the green and red sauces by indicating the tastiness on a 10-point scale. As in exercise 6, researchers are concerned that the color of the green sauce will adversely affect tastiness scores. Tastiness scores tend to be skewed. The scores follow.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Red Sauce</th>
<th>Green Sauce</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

a. What statistical test should be used to analyze these data?
b. Identify $H_0$ and $H_a$ for this study.
c. Conduct the appropriate analysis.
d. Should $H_0$ be rejected? What should the researcher conclude?

8. You notice in your introductory psychology class that more women tend to sit up front, and more men sit in the back. To determine whether this difference is significant, you collect data on the seating preferences for the students in your class. The data follow.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front of the Room</td>
<td>15</td>
<td>27</td>
</tr>
<tr>
<td>Back of the Room</td>
<td>32</td>
<td>19</td>
</tr>
</tbody>
</table>

a. What is $\chi^2_{\text{obs}}$?
b. What is $df$ for this test?
c. What is $\chi^2_{\text{cv}}$?
d. What conclusion should be drawn from these results?

9. Identify the statistical procedure that should be used to analyze the data from each of the following studies:

a. A study that investigates whether men or women (age 16 to 20) spend more money on clothing. Assume the amount of money spent is normally distributed.
b. In the (a) study, it has since been determined that the amount of money spent really is not normally distributed.
c. A study that investigates the frequency of drug use in suburban versus urban high schools.
d. A study that investigates whether students perform better in a class that uses group learning exercises versus a class that uses the traditional lecture method. Two classes that learn the same information are selected. Performance on a 50-item final exam at the end of the semester is measured.

**CRITICAL THINKING CHECK ANSWERS**

10.1

1. Effect size indicates the magnitude of the influence of the experimental treatment, regardless of the sample size. A result can be statistically significant because the sample size is very large, even if the effect of the independent variable is not so large. Effect size indicates whether this is the case because, in this situation, effect size should be small.

2. In the long run, it means that the obtained $t$ is more likely to be significant. In terms of the formula used to calculate $t$, increasing the sample size will decrease the standard error of the difference between means ($s_{\bar{X}_1 - \bar{X}_2}$). This, in turn, will increase the size of the obtained $t$. A larger obtained $t$ means that the obtained value is more likely to exceed the critical value and be significant.

3. Decreasing variability also makes a $t$ test more powerful (likely to be significant) because decreasing variability also means that $s_{\bar{X}_1 - \bar{X}_2}$ (the standard error of the difference between means) will be smaller. Again, this increases the size of the obtained $t$, and a larger obtained $t$ means that the obtained value is more likely to exceed the critical value and be significant.