LEARNING OUTCOMES

After reading and reviewing Chapter 6, you should understand

- The use of inferential statistics to determine whether the finding of a study is unusual
- The importance of the sampling distribution
- How to carry out the hypothesis testing process
- When to reject or retain a null hypothesis and the types of errors associated with each of these decisions
- The distinction between statistical significance, effect size, confidence intervals, and practical significance

CHAPTER SUMMARY

The chapter is focused on inferential statistics that allow us to make inferences about a population from findings with a sample. Sampling distributions are a distribution of some statistic and provide the basis for deciding whether our sample is representative (or not) of the population—or when a difference makes a difference. The decision-making process is called hypothesis testing and involves multiple steps: stating null and alternative hypotheses, defining the sampling distribution, setting the criterion level (.05 or .01), computing a statistic, and deciding whether to reject or retain the null hypothesis. The alternative hypothesis determines the region of rejection in the sampling distribution and can be one-tailed (directional) or two-tailed (nondirectional). If the results of our analysis fall in the region of rejection, we reject the null hypothesis and support the alternative hypothesis, but if the statistic falls in the region of acceptance, we retain the null hypothesis.

We design studies to try to reject a false null hypothesis (called power), but because hypothesis testing is based on probability, the decision to reject or retain the hypothesis is associated with a probability of error. Type I error occurs when we reject a true null hypothesis, and Type II error occurs when we retain a false null hypothesis. Ways to decrease the probability of either type of error are described as well as ways to increase power. Finally, effect sizes (magnitude of an effect), confidence intervals (margin of error), and practical significance (everyday implications) are described as tools to enhance our understanding of the results of a study.
REVIEW AND APPLICATION OF KEY CONCEPTS
FROM CHAPTER 6

Exercise 6.1: Inferential Statistics (Review)

1. __________ statistics are the statistical analysis of data from a sample used to draw a conclusion about a population from which the sample is drawn.

2. Statistics from a population are called _____________________.

3. __________ is the population mean.

4. __________ is the population standard deviation.

5. Fill in the table:

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>Inferential</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample</td>
<td>parameter</td>
</tr>
<tr>
<td>mean (M)</td>
<td>sigma (σ)</td>
</tr>
</tbody>
</table>

6. Inferential statistics are based on ______________ theory, which examines random events, such as what card you will draw in poker.

7. ______________ is the decision-making process of determining the probability of finding a particular result.

8. A __________ describes a distribution of statistics while a __________ describes a distribution of scores.

Exercise 6.2: Hypothesis Testing (Review)

1. The __________ hypothesis predicts what you expect to find in your study.

2. The __________ hypothesis predicts no difference between groups.

3. A study is always designed to __________ the null hypothesis.

4. The typical region of rejection is the extreme ______% of the sampling distribution, but sometimes we use the more strict extreme _____% to define the region of rejection.

5. If your results fall in the region of rejection, you should __________ (reject/retain) the null hypothesis and you __________ (have, have not) found statistical significance.
6. If your results fall in the region of acceptance, you should ____________ (reject/retain) the null hypothesis and you ____________ (have/have not) found statistical significance.

7. Statistical significance suggests that your results are not due to ____________ and that the results ____________ (do/do not) belong to the sampling distribution implied by the null hypothesis.

8. A one-tailed test makes it ____________ (more/less) difficult to reject the null hypothesis because the region of rejection is located on one tail of the sampling distribution, while two-tailed tests divide the region of rejection between the two tails of the distribution.

9. A two-tailed test is a ____________ (more/less) conservative test than a one-tailed test because ____________.

10. Researchers typically compute a __________-tailed test because it is more conservative.

11. The value of a statistic that defines the region of rejection in a sampling distribution is called the ____________ value.

12. The percentage of the distribution that the researcher selects for the region of rejection is called the ________ level and typically is less than ______%.

13. When you reject the null hypothesis, there is a probability of a ____________ error and it is equal to the ____________.

14. When you retain the null hypothesis, there is a probability of a ____________ error.

15. Regardless of your decision to reject or retain the null hypothesis, there is always a chance of error because statistics are based on ____________.

16. Why can't you have a probability of both a Type I and a Type II error in a study?

17. Correctly rejecting a false null hypothesis is called ____________.

18. You can reduce the probability of a Type II error by (increasing/decreasing)
   a. ____________ sample size
   b. ____________ error in your research design
   c. ____________ strength of the effect
Exercise 6.3: Hypothesis Testing (Application)

1. “There will be no difference in the number of crimes committed by those in urban areas and those in rural areas” is an example of a(n) __________ hypothesis.

2. “The number of crimes committed in urban areas will differ from the number committed in rural areas” is an example of a(n) ________________ hypothesis and it is ________________ (directional/nondirectional).

3. “The number of crimes committed in urban areas will be greater than the number committed in rural areas” is an example of a(n) ________________ hypothesis and it is ________________ (directional/nondirectional).

4. “Males are more likely than females to physically bully their classmates.” This is a ___________________ (directional/nondirectional) alternative hypothesis.

5. Define a sampling distribution for the following hypothesis: There is no difference in the percentage of income spent on housing by elderly households and the percentage of income spent on housing by all households. Hint: First name the percentage spent by all households (estimate a reasonable percentage).

6. In the figures below, the numbers on the $x$-axis are standard deviation units away from the mean. For example, $-2 = 2 \text{SD}$'s below the $M$, while $1 = 1 \text{SD}$ above the $M$.

   Figure A: Draw the region of rejection and region of acceptance for a two-tailed test where $p < .05$.

   Figure B: Draw the region of rejection and region of acceptance for a one-tailed test where $p < .05$.

7. List the seven steps that you would follow to test the following hypothesis: “Males are more likely than females to physically bully their classmates.” Make your steps relevant to this specific study (e.g., state the null and alternative hypotheses for this study).

   a. __________________________________________

   b. __________________________________________
c. 

d. 

e. 

f. 

g. 

8. Which of the following p values would meet the criteria for statistical significance at \( p < .05 \)? Circle them.

\[ .001 \quad .05 \quad .006 \quad .50 \quad .047 \quad .20 \quad .02 \quad .07 \quad .70 \]

9. If you find that, in your sample, girls verbally bully their classmates more often than boys (\( p = .04 \)),

a. Are the results statistically significant? __________________

b. Would you reject or retain the null hypothesis? __________________

c. What is the probability of a Type I error? _____ Type II error? _____

d. Name two ways you could reduce your chance of a Type I error in the study.

i. __________________________________________

ii. __________________________________________

10. If you find, in your sample, that males physically bully their classmates more often than girls (\( p = .07 \)),

a. Are the results statistically significant? __________________

b. Would you reject or retain the null hypothesis? __________________

c. Is there a probability of a Type I error? _____ Type II error? _____

11. Two studies were conducted on bullying at an elementary school. Study 1 selected a convenience sample of 10 boys and 10 girls from grades K–5. Study 2 selected a convenience sample of 25 boys and 25 girls from the fourth grade.

Which study is likely to have more power? ______ Why? (Hint: There are two reasons.)

a. __________________________________________

b. __________________________________________
Exercise 6.4: Effect size, Confidence Intervals, and Practical Significance (Review)

1. The magnitude of an effect in a study is the ____________________.
2. The margin of error in a study is defined by the ____________________.
3. The everyday usefulness of results is the ________________________.
4. The effect size used to examine mean differences and measured in standard deviation units is called ________________________.
5. Another way to measure the effect size is the ________________________.
6. Fill in the table, but remember that these numbers are guidelines and not designed to be strict cutoffs.

<table>
<thead>
<tr>
<th>Interpretation of Effect Size</th>
<th>Effect Size: Cohen’s d</th>
<th>Effect Size: Proportion of Variability Accounted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small/Weak</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium/Moderate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large/Strong</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Exercise 6.5: Effect size, Confidence Intervals, and Practical Significance (Application)

1. Interpret each of the following effect sizes as weak, moderate, or strong:
   a. 2% of the variability accounted for ________________________
   b. 50% of the variability accounted for ________________________
   c.  $d = .48$ ________________________________________________
   d.  $d = 1.20$ ________________________________________________
   e.  $d = .12$ ________________________________________________
   f.  10% of the variability accounted for ________________________
   g.  $d = .25$ ________________________________________________

2. A teacher reports that exam grades ranged from 64% to 98%, with a mean of 79.85 ($SD = 10.18$), 95% CI [75.09, 84.61]. The confidence interval tells you that you are _______________ confident that the sample mean of _______________ represents a population mean that falls between _______________ and _______________.

Do not copy, post, or distribute
3. Another teacher reports exam scores based on academic year.
   First-year students: $M = 75.25, SD = 5.16, 95\% \text{ CI} [71.70, 78.88]$
   Second-year students: $M = 80.33, SD = 7.33, 95\% \text{ CI} [74.11, 86.55]$
   a. Do the confidence intervals for the population means overlap? ________
   b. A second-year student brags that their year outperformed first-year students. Based on the confidence intervals, how might you counter this student’s claim?
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
   c. What would you say about the practical significance of these results?
   __________________________________________________________
   __________________________________________________________
   __________________________________________________________

Exercise 6.6: The Big Picture: Making Sense of Results (Application)

1. What are the benefits of reporting each of the following in describing the results of a study?
   a. Statistical significance ________________________________________
   __________________________________________________________
   b. Effect size _________________________________________________
   __________________________________________________________
   __________________________________________________________
   c. Confidence interval __________________________________________
   __________________________________________________________
   __________________________________________________________
   d. Practical significance _________________________________________
   __________________________________________________________
   __________________________________________________________
2. Statistical significance, effect size, confidence intervals, and practical significance ____________ (can/cannot) vary independently.

3. A study examined whether convenience affected recycling behavior. The researchers operationally defined recycling behavior based on the weight (in pounds) of material recycled by each household. They found that households who were supplied a large recycling bin and weekly pickup ($M = 25.00, SD = 5.00, 95\% \text{ CI} [20.25, 29.75]$) recycled significantly more material than the households who were sent flyers encouraging them to recycle ($M = 10.00, SD = 4.00, 95\% \text{ CI} [6.50, 13.50], p = .003$). Convenience accounted for 22\% of the variance in recycling.
   a. Explain what the $p$ value indicates and how you would interpret it.
      ____________________________________________________________
      ____________________________________________________________
   b. Interpret the effect size.
      ____________________________________________________________
      ____________________________________________________________
   c. Interpret the confidence intervals.
      ____________________________________________________________
      ____________________________________________________________
   d. Interpret the practical significance of the study.
      ____________________________________________________________
      ____________________________________________________________

4. The researchers replicated the study in a different city and found that those supplied with a recycling bin ($M = 25.00, SD = 7.00, 95\% \text{ CI} [16.20, 33.80]$), recycled more than those who were just given an informational flyer ($M = 10.00, SD = 5.50, 95\% \text{ CI} [2.50, 17.50]$). This time, they found that $p = .10$, Cohen’s $d = .25$.
   a. Explain what the $p$ value indicates and how you would interpret it.
      ____________________________________________________________
      ____________________________________________________________
   b. Interpret the effect size.
      ____________________________________________________________
c. Interpret the confidence intervals.

__________________________________________________________

__________________________________________________________

d. Interpret the practical significance of the study.

__________________________________________________________

__________________________________________________________

5. Which of the two recycling studies

a. Found statistically significant results? Explain your answer.

__________________________________________________________

__________________________________________________________

b. Had a stronger effect size? Explain your answer.

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

c. Has more practical significance? Explain your answer.

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________

YOUR RESEARCH

Find a primary research article on your topic. If possible, find an article that has a simple design. If the article describes multiple studies, pick one of the studies as the focus of this exercise. Before beginning to work on this section, look through the entire article.

You will see that in order to answer these questions, you will have to read the article very carefully and probably several times. This practice should give you some idea of the care that you should take in digesting research. Even though you need to take a lot of time to read articles, you may end up summarizing the article in your own literature
review with a single sentence or by simply citing the article because of the measure you use from the study or for the procedure you adopt. But by attending to the details of published articles, you will learn a great deal about the format for reporting research, different methods for conducting research, various ways to analyze data, and appropriate interpretations of statistics. All of this will support you as you learn to design, carry out, analyze, and report your own studies.

1. Focus on the Introduction of the article.
   a. Identify one hypothesis that most interests you. (Remember that it is important to learn how to write about research in your own words. Be sure that you do not plagiarize in writing the hypothesis!)

   __________________________________________________________
   __________________________________________________________

   b. Is the hypothesis directional or nondirectional? How do you know?

   __________________________________________________________
   __________________________________________________________

2. Now examine the Method section.
   What do you learn about the sample and procedure of the study that may help the researchers to avoid Type II errors (e.g., sample size, error in the research design, strength of effect)? You may want to review the section “Reducing the Chance of a Type II Error” in Chapter 6 before answering this question.

   Sample:

   __________________________________________________________
   __________________________________________________________
   __________________________________________________________

   Research design:

   __________________________________________________________
   __________________________________________________________
   __________________________________________________________

   Strength of effect:

   __________________________________________________________
   __________________________________________________________
   __________________________________________________________
3. Review the Results section of the article. Even though there may be statistics you do not understand, focus on the statistics that you are familiar with: $M, SD$, range, possible and observed scores, $p$, percentage of variability accounted for (could be noted as $r^2$, $r_{pb}^2$, $\eta^2$), Cohen's $d$, confidence intervals. Consider the hypothesis you selected in question 1 and answer the following questions regarding the analyses used to test this hypothesis.

a. Did the author(s) report any descriptive statistics (e.g., $M, SD$) related to the hypothesis you selected? If yes, explain what the descriptive statistics tell you about the variable(s) under examination.

b. Did the researchers find statistically significant results for the hypothesis you noted in question 1? Explain.

c. Is there a probability of a Type I error? If yes, what is the exact probability? Is there a probability of a Type II error?

d. Did the authors report an effect size? If yes, list it here and interpret its meaning.

e. Did the authors include a confidence interval for their findings? If yes, list the confidence interval(s) and interpret the meaning.

4. Reread the Discussion section.

a. Did the authors indicate that they found support for the hypothesis you selected?
b. How do they interpret the meaning of the support or nonsupport for their hypothesis?


c. What do the researchers state or imply about the practical significance of their study? They may not use the term *practical significance*; instead, they may discuss the implications or usefulness of their results.


d. Researchers also discuss limitations of their study in the Discussion. Do you see any limitations they mention that you can relate to the possibility of Type I or Type II errors? How do the authors suggest that future researchers should address or overcome these limitations?


As you read other primary research articles on your topic, we recommend you use some or all of these questions to help you evaluate the research.
Calculating a Confidence Interval for a Mean Score

You can use IBM® SPSS® to calculate confidence intervals for means by running an analysis called a one-sample *t* test. You will learn more about the one-sample *t* test in Chapter 7. For our purposes here, we are only using it to calculate a confidence interval.

Data Entry

You need a variable measured on an interval or ratio scale to calculate the confidence interval for the mean of that variable. In the example below, we have 20 students and their final exam scores (out of 100), which is a ratio variable.

You can then calculate descriptive statistics and verify that the distribution meets the criteria for a normal curve, using either the Frequency or Descriptives command (see Chapter 5 of this study guide). The following analysis will repeat some of that information.

<table>
<thead>
<tr>
<th>studentID</th>
<th>finalexam</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98.00</td>
</tr>
<tr>
<td>2</td>
<td>85.00</td>
</tr>
<tr>
<td>3</td>
<td>78.00</td>
</tr>
<tr>
<td>4</td>
<td>64.00</td>
</tr>
<tr>
<td>5</td>
<td>70.00</td>
</tr>
<tr>
<td>6</td>
<td>88.00</td>
</tr>
<tr>
<td>7</td>
<td>72.00</td>
</tr>
<tr>
<td>8</td>
<td>88.00</td>
</tr>
<tr>
<td>9</td>
<td>70.00</td>
</tr>
<tr>
<td>10</td>
<td>70.00</td>
</tr>
<tr>
<td>11</td>
<td>77.00</td>
</tr>
<tr>
<td>12</td>
<td>85.00</td>
</tr>
<tr>
<td>13</td>
<td>70.00</td>
</tr>
<tr>
<td>14</td>
<td>95.00</td>
</tr>
<tr>
<td>15</td>
<td>85.00</td>
</tr>
<tr>
<td>16</td>
<td>77.00</td>
</tr>
<tr>
<td>17</td>
<td>85.00</td>
</tr>
<tr>
<td>18</td>
<td>65.00</td>
</tr>
<tr>
<td>19</td>
<td>80.00</td>
</tr>
<tr>
<td>20</td>
<td>95.00</td>
</tr>
</tbody>
</table>

Calculating the Confidence Interval for the Mean

On the Menu Bar, click **Analyze → Compare Means → One-sample *t* test**
1. Send the variable from the sample to “Test Variable(s).”

2. To calculate a confidence interval, leave the test value as 0. (Note: the default is a 95% confidence interval, to change this, click “Options.”)

3. Click “OK.”

**OUTPUT**

**One-Sample Statistics**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>finalexam</td>
<td>20</td>
<td>79.8500</td>
<td>10.17880</td>
<td>2.27605</td>
</tr>
</tbody>
</table>

The first table provides descriptive statistics for the sample.

**One-Sample Test**

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>finalexam</td>
<td>35.083</td>
<td>19</td>
<td>.000</td>
<td>79.85000</td>
<td>75.0862</td>
</tr>
</tbody>
</table>

When you are interested in the confidence interval for the mean, focus on the last two columns of the second table and ignore the rest.

**Writing Up Results**

When reporting a confidence interval for a mean, include the following:

- The mean and standard deviation ($M, SD$)
- What confidence level you are using (e.g., 95%, 99%)
- The confidence interval, typically but not always, is formatted as CI [lower, upper]

You may also opt to include other descriptive information, such as the sample size or minimum and maximum scores or other information you gained from conducting descriptive statistics (see Chapter 5 of this study guide).
Examples:

Twenty students completed a final exam. The mean score was 79.85 (SD = 10.18), 95% CI [75.09, 84.61].

OR

Student scores on the final exam ranged from 64 to 98 out of 100 possible (M = 79.85, SD = 10.18), 95% CI [75.09, 84.61].

OR

Final exam scores for 20 students indicated an average score of C+/B- (M = 79.85, SD = 10.18). At a 95% confidence level, we estimate that the population mean falls between 75.09 and 84.61.

Review and Practice Exercise

One evening, two waiters record what percentage of the bill their first 10 customers left as tips:

<table>
<thead>
<tr>
<th>Waiter 1</th>
<th>Waiter 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>.15</td>
<td>.18</td>
</tr>
<tr>
<td>.16</td>
<td>.12</td>
</tr>
<tr>
<td>.15</td>
<td>.25</td>
</tr>
<tr>
<td>.12</td>
<td>.30</td>
</tr>
<tr>
<td>.18</td>
<td>.15</td>
</tr>
<tr>
<td>.22</td>
<td>.10</td>
</tr>
<tr>
<td>.15</td>
<td>.20</td>
</tr>
<tr>
<td>.25</td>
<td>.18</td>
</tr>
<tr>
<td>.20</td>
<td>.22</td>
</tr>
<tr>
<td>.14</td>
<td>.15</td>
</tr>
</tbody>
</table>

1. Review: Enter the data into SPSS.
2. Review: Using SPSS, calculate the appropriate descriptive statistics to describe the tip percentage for the entire sample. Write the statistics below.

3. Review: Using SPSS, calculate the appropriate descriptive statistics to compare the two waiters’ tip percentages. Write the statistics below.
4. Practice with confidence intervals: Using SPSS, calculate the confidence interval for the entire sample mean and for each waiter separately. Write the intervals below.

_____________________________________________________________
_____________________________________________________________
_____________________________________________________________

5. Write up your results using correct APA format. Include a brief interpretation of the results based on the confidence interval.

_____________________________________________________________
_____________________________________________________________
_____________________________________________________________
_____________________________________________________________
_____________________________________________________________
_____________________________________________________________
_____________________________________________________________

6. What is the practical significance of the results? What other information would be useful in understanding the practical significance?

_____________________________________________________________
_____________________________________________________________