CHAPTER 9

Effect Size and NHST: One-Sample Designs

Summary

This chapter is all about analyzing data from a research design with only one sample.

This chapter introduced the basics of ***null hypothesis significance testing (NHST)***. NHST is a statistical tool that helps you make decisions about populations when you have sample data. This chapter introduces you to techniques that are appropriate when data are gathered from only *one* sample. Techniques that apply data from two or more samples are covered in later chapters. For a one-sample *t* test, the *effect size index, d,* tells you about the *size* of the difference between the mean of the population from which the sample was drawn and the mean of the null hypothesis population. As before, *d* values of 0.20, 0.50, and 0.80 characterize small, medium and large effect sizes, respectively, for a one-sample *t* test. An effect size estimate is an increasingly popular way to describe differences between populations.

There are several steps involved in NHST:

1. Prior to collecting data, set your ***alpha*** (α) level. *Alpha* is typically .05 or less and is also called the ***significance level***.
2. Set your ***null hypothesis*** (H_0), which states that your sample statistic (for example, $\bar{X}$ or *r)* came from a particular population with a specified parameter (μ_0 or ρ). This is a hypothesis of equality.
3. Set your ***alternative hypothesis*** (H_1) by deciding whether you want to conduct a ***two-tailed test of significance*** or a ***one-tailed test of significance***. The *alternative hypothesis* represents a researcher's "best guess" about the results of the study and is a hypothesis of difference.
4. Choose a sampling distribution that gives the probability of different sample statistics when samples are drawn from the population that has the particular parameter. In this chapter, the proper sampling distribution is the ***t distribution***.
5. Use the proper sampling distribution to determine the probability of your sample mean or one more extreme. In this chapter, this involves comparing your calculated *t* value to the appropriate ***critical value*** of *t* from Table D. A NHST analysis for a sample mean uses $df = N - 1$, where *N* is the number of scores. A NHST analysis for a correlation coefficient uses $df = N - 2$, where *N* is the number of *pairs* of scores. The probability value provided by the sampling distribution is correct *if the null hypothesis is true.*
6. If the probability is small (equal to or less than α; your calculated value falls within the ***rejection region***), conclude that the null hypothesis is probably incorrect and that your sample data have likely come from some other population. Here, you *reject the null hypothesis.*
7. If the probability is large (greater than α; your calculated value does not fall in the *rejection region*), conclude that the data are consistent with the null hypothesis and perhaps with other hypotheses as well. In statistical terms, you are now left with insufficient evidence to reject the null hypothesis and, instead, *retain the null hypothesis*.
8. Write an informative conclusion about the results of your NHST analysis that a) uses the terms of the experiment; b) tells the direction of the difference between the population means; and c) identifies whether or not this difference is ***statistically significant***.

Expanding on Step 3, the most common *alternative hypothesis* is *two-tailed* (H_1: $\mu_1 \neq \mu_0$ or H_1: $\rho \neq .00$), an alternative that places half the rejection region in each tail of the sampling distribution. A *two-tailed test* permits the rejection of the null hypothesis for means that are either larger or smaller than the null hypothesis mean and for correlation coefficients that are either positive or negative. One version of a *one-tailed t test* permits rejection of the null hypothesis only if the sample mean is larger than the null hypothesis mean (H_1: $\mu_1 > \mu_0$). Such a test cannot detect a sample mean that is smaller, no matter how small. A second version reverses the sign to permit rejection only if the sample mean is *smaller* than the null hypothesis mean. This chapter only reviewed the two-tailed version of the alternative hypothesis for testing the significance of correlation coefficients.

Like all decision-making aids, NHST analyses can lead to wrong conclusions. If the *null hypothesis* is indeed true, and the data lead you to reject it, you have made a ***Type I error***. The probability of a *Type I error* is never greater

than α, which is set by the researcher. If the *null hypothesis* is actually false and the data lead you to retain it, you have made a ***Type II error***. The probability of a *Type II error* is symbolized by ***beta (β)***. Among several factors that determine β are a) α - the smaller α is, the larger β is, and vice versa; and b) the actual difference between the population from which the sample was drawn and the null hypothesis population—the larger the difference, the smaller β is. Thus, not having sufficient evidence to reject the null hypothesis does not mean that the experiment is a failure. There are several reasons why a test may fail to reject the null hypothesis. You will learn more factors that determine β and about the related concept of power in later chapters.

Finally, when working through the problems at the end of this page (and in future study guide chapters), note that the interpretations provided in the answers will be presented in two ways. The first version will be a simple, verbal interpretation of the NHST results. The second version will mirror formatting you would see in journals that follow APA style, which dictates how descriptive and inferential statistics are reported. However, your instructor may prefer you write your interpretations using different formatting. Please take this into consideration when checking your work.

Multiple-Choice Questions

1. In NHST, the hypothesis that is tested is about a ________________.

a. sample
b. population
c. both a. and b.
d. neither a. nor b.

2. Using NHST, you can conclude that the null hypothesis is ________________.

a. probably true
b. probably false
c. both a. and b. are possible conclusions
d. neither a. nor b. are possible conclusions

3. The *t* distribution, as a sampling distribution, gives the probability of events ________________.

a. when the null hypothesis is true
b. when the null hypothesis is false
c. regardless of whether the null hypothesis is true or false
d. when the alternative hypothesis is not identified

4. In statistics, the word “significant” indicates ________________.

a. the null hypothesis was rejected
b. results are important
c. a large effect size is present
d. a small effect size is present

5. Which phrase goes with “in the rejection region”?

a. the probability of the difference is small, if the null hypothesis is true
b. reject the null hypothesis
c. both a. and b.
d. neither a. nor b.

6. When an experimenter uses $\alpha = .05$, the rejection region is ________________.

a. 95% of the curve
b. 5% of the curve
c. 5% of the curve for a one-tailed test and 10% of the curve for a two-tailed test
d. 10% of the curve for a one-tailed test and 5% of the curve a two-tailed test

7. When we reject the null hypothesis, we have evidence that the sample mean ________________.

a. is very large
b. is very small
c. comes from a different population than the null hypothesis population
d. comes from the same population as the null hypothesis population

8. An effect size index is most closely associated with which phrase below?

a. the α level chosen by the researcher
b. the probability of a Type II error
c. the size of the difference between the sample mean and the null hypothesis mean
d. the size of the rejection region

9. Suppose you obtained a sample from a population for which the null hypothesis was true. On the basis of a *t* test, you failed to reject the null hypothesis. You have made a ________________.

a. Type I error
b. Type II error
c. correct decision
d. cannot be determined from the information given

10. Suppose you obtained a sample from a population different from the one specified by the null hypothesis. On the basis of a *t* test, you failed to reject the null hypothesis. You have made a ________________.

a. Type I error
b. Type II error
c. correct decision
d. cannot be determined from the information given

11. "$p > .05$" means that ________________.

a. the null hypothesis should be rejected
b. the difference between the statistic and the null hypothesis parameter is statistically significant
c. both a. and b.
d. neither a. nor b.

12. We reject the null hypothesis when ________________.

a. $p > \alpha$
b. $p < \alpha$
c. $d = 0.00$
d. cannot be determined from the information given

13. Which of the following shows a correct match-up of an alternative hypothesis and its one- or two-tailed test?

a. H_1: $\mu_1 = \mu_0$, one-tailed
b. H_1: $\mu_1 > \mu_0$, two-tailed
c. H_0: $\mu_1 = \mu_0$, two-tailed
d. H_1: $\mu_1 \neq \mu_0$, two-tailed

14. The choice of an alternative hypothesis has an effect on ________________.

a. conclusions you may draw
b. α level
c. which hypothesis you are testing
d. all of the above

15. Which answer below belongs with the concept of a two-tailed test of significance?

a. H_1. $\mu_1 > \mu_0$
b. H_0: $\mu_1 < \mu_0$
c. $d > 0.80$
d. a divided rejection region

16. A one-tailed test is appropriate when ________________.

a. you do not have enough data for a two-tailed test
b. you have only one sample, not two
c. you are only interested in finding out if a new treatment reduces symptoms
d. you want to test for either a positive or a negative correlation

17. When using the *t* distribution to determine the significance of a correlation coefficient, the null hypothesis is that the population correlation coefficient is ________________.

a. -1.00
b. 0.00
c. 1.00
d. equal the coefficient obtained from the sample

18. A calculated *t* test value that is larger than the critical value in the table ________________.

a. leads to the rejection of the null hypothesis
b. has a probability of less than α if the null hypothesis is true
c. both a. and b.
d. neither a. nor b.

19. The reason for not reducing the alpha level to one in a million or even less is ________________.

a. tables do not exist for such small levels
b. α would be too large
c. β would be too large
d. β would be too small

20. If the calculated *t* test value is less than the critical value in the table, ________________ the null hypothesis even though you could be making a ________________ error.

a. reject, Type I
b. retain, Type II
c. reject, Type II
d. retain, Type I

Short-Answer Questions

1. Distinguish between Type I and Type II errors.

2. Distinguish between $\alpha = .05$ and $p = .05$.

3. Distinguish between rejecting and retaining the null hypothesis.

4. Write an interpretation of each of the following situations.
   a. Stanley Milgram found in the early 1960s that a cross section of Americans was willing to administer an average of 285 volts to other participants in an experiment (Milgram, 1963). Have times changed? Suppose that the study was replicated today with 20 participants who were willing to administer an average of only 240 volts. A *t* test produced a value of -2.00, with an effect size index of $d = |-0.45|$. Write the null hypothesis and a two-tailed alternative hypothesis. Find the appropriate critical value. Then, write a conclusion about the difference between today and the early 1960s in people's willingness to administer shock to others.
   b. Han, Kim, and Kim (2017) studied the relationships between smartphone separation anxiety and various psychological factors. For example, they found a correlation of .57 between the degree to which smartphone users associate their phones with positive personal memories and levels of nomophobia (the fear of being without one's mobile phone). Write the null hypothesis and the two-tailed alternative hypothesis. Set alpha at .05 and find the appropriate critical value. If the correlation between these two variables was based on a sample of 301 participants, what conclusion would be appropriate?

5. For each of the correlations below, determine whether you would *reject* or *retain* the null hypothesis under the conditions specified. Explicitly state whether or not the correlation is significantly different from .00 for a two-tailed test.
   a. $r = .27$, $N = 32$, $\alpha = .05$
   b. $r = .20$, $N = 102$, $\alpha = .05$
   c. $r = .40$, $N = 17$, $\alpha = .01$
   d. $r = -.11$, $N = 63$, $\alpha = .05$
   e. $r = -.38$, $N = 52$, $\alpha = .01$

Problems

1. The 10-item Perceived Stress Scale (PSS-10; Cohen et al., 1983; Cohen & Williamson, 1988) is a common measure of everyday experiences of stress. Scores on this scale can range from 0 to 40. Across three years (1983, 2006, and 2009), the weighted average score for individuals aged 18-25 was 16.04 (Cohen & Janicki-Deverts, 2012). The hypothetical data in this problem are PSS-10 scores for college students experiencing academic difficulty. State the null hypothesis, set a two-tailed alternative hypothesis, perform a *t* test, calculate the effect size index, and write a conclusion about feelings of stress among students experiencing academic difficulty, relative to the population average for those aged 18-25 years. Set alpha at .05.

16	17	17	19	18	23	21	20	17	18

2. Every year, thousands of college-bound American high school seniors take the Scholastic Aptitude Test (SAT). According to the College Board's SAT Suite of Assessments Annual Report, the mean total score in 2017 was 1060. Summary statistics for one small high school follow. How do the students compare to the national norm? State the null hypothesis, choose an alternative hypothesis, perform a *t* test, calculate the effect size index, and write a conclusion. Set alpha at .05.

$\Sigma X = 27{,}456$ $\Sigma X^2 = 29{,}900{,}000$ $N = 26$

3. Some psychologists study self-interested behavior and the factors that affect it. For example, Zhong, Bohns, and Gino (2010) tested the effects of darkness on cheating behavior. In the first study, a group of participants worked on 20 sets of math problems in a dimly-lit room. The participants were told they would be paid $0.50 for each math problem they correctly solved in a 5-minute period. After 5 minutes, the participants self-reported how many math problems they correctly solved. Prior research established that people can solve 7 of the 20 problems, on average, during 5 minutes. Did these participants cheat? Below are summary statistics that partially mimic those reported in Study 1 of Zhong et al. Perform a two-tailed *t* test that compares the average number of math problems reportedly solved by the participants in the dimly-lit room to the known norm. Calculate the effect size and write a conclusion that summarizes the results. Set alpha at .05. For the purposes of your calculations, assume the standard deviation reported in the table is $\hat{s}$.

	Dimly-lit Room
Mean	11.47
Standard deviation	4.32
N	42

4. In a follow-up study, Zhong, Bohns, and Gino (2010) tested the effects of wearing dark sunglasses on sharing behavior. In this study, the participants played a "dictator game," in which they were given $6 and told to divide that money between themselves and a stranger in a different room however they wanted. In fact, there was no stranger in a different room and the main dependent measure was how much money participants chose to share. Below are partial summary statistics from Study 2 of Zhong et al. (2010). Approximate the results of Zhong et al. by performing a two-tailed *t* test that compares the average amount of money shared by those wearing sunglasses to the fair division of money ($3.00). Calculate the effect size and write a conclusion that summarizes the results. Set alpha at .05. For the purposes of your calculations, assume the standard deviation reported in the table is $\hat{s}$.

	Sunglasses Group
Mean	$1.81
Standard deviation	$1.30
N	26

ANSWERS

Multiple-Choice Questions

1. b
   Explanation: Although we use samples to represent populations, the hypotheses we test are *always* about populations.
2. b
   Explanation: All probability values in NHST are assumed to be accurate when the null hypothesis is true. Thus, when the probability of a sample mean or one more extreme is low (typically, $p < .05$), there is strong evidence that the null hypothesis is false and we *reject* it. However, when the probability of a sample mean is higher (typically, $p > .05$), the null hypothesis being true becomes just one of several probable conclusions. Thus, we *retain* the null or *fail to reject it,* but we never claim it to be true.
   Understanding what conclusions can be drawn about the null hypothesis is one of the most difficult ideas you'll tackle in a basic statistics course. Now might be a good time to review "The Logic of Null Hypothesis Significance Testing (NHST)" in your textbook.
3. a
   Explanation: The probability values provided by sampling distributions always assume that the null hypothesis is true.
4. a
   Explanation: In statistics, "significant" is short for "statistically significant," which means that a difference is unlikely to be due to sampling error. Whether or not a difference or a relationship is statistically significant does not tell us anything about the effect size, nor does it tell us about the importance of the finding.
5. c
6. b
   Explanation: When conducting a one-tailed test, the entirety of the rejection region (here, all 5%) falls in one tail of the distribution. When conducting a two-tailed test, the rejection region gets split between both tails of the distribution, with 2.5% in each tail.
7. c
8. c
9. c
10. b
    Explanation: If your sample comes from a population that is different from the one specified by the null hypothesis, the correct decision is to *reject* the null hypothesis. A Type II error is failing to reject a false null.
11. d
    Explanation: If $p > .05$, then the null hypothesis should be retained, which means a statistically significant difference is *not* present.
12. b
13. d
    Explanation: Only null hypotheses should indicate equality (=) of two means. The notation used in b. indicates a one-tailed test, not a two-tailed test.
14. a
    Explanation: Whether or not a null hypothesis is rejected or retained depends, in part, on the form the alternative hypothesis takes. A one-tailed test places alpha in one tail of the distribution (one larger rejection region taking up, say, 5% of the curve), whereas a two-tailed test splits alpha between *both* tails of the distribution (two smaller rejection regions, each taking up 2.5% of the curve). The α level is set by the researcher and does not depend on the alternative hypothesis.
15. d
    Explanation: A two-tailed test of significance involves splitting the rejection region between both tails of the distribution. The hypothesis presented in a. denotes a one-tailed alternative hypothesis and the hypothesis presented in b. denotes a null hypothesis. Effect size is independent of tests of significance.
16. c
    Explanation: A one-tailed test always specifies a *directional* prediction. In other words, a one-tailed test of significance can *only* determine if the difference between means is positive *or* negative.
17. b
    Explanation: For correlation problems, the null hypothesis states there is no relationship and a correlation of zero means there is no relationship.
18. c
19. c
    Explanation: The probability of making a Type I error (α) is inversely related to the

probability of making a Type II error (β). Thus, reducing α increases β.

20. b

Short-Answer Questions

1. A Type I error is committed when you *reject* the null hypothesis is false when, in fact, you should retain it. A Type II error is committed when *retain* the null hypothesis when, in fact, you should reject it. The probability of committing a Type I error is quantified by alpha (α) and the probability of committing a Type II error is quantified by beta (β).
2. Alpha (α) is the significance level chosen by the researcher. p is the probability of the observed data (e.g., a sample mean), if the null hypothesis is true. If $\alpha = .05$ and the p associated with the observed data is less than .05 ($p < .05$), then we reject the null hypothesis. If $p > .05$, we retain the null hypothesis.
3. A null hypothesis is rejected when the difference between the sample statistic and the population parameter is unlikely to be observed *if the null hypothesis is true.* The null hypothesis is retained when the difference between the sample statistic and the population would be expected to occur *if the null hypothesis is true.* When you reject the null hypothesis, you are left with strong support for the alternative hypothesis. When you retain the null hypothesis, you are left with both the possibility that the null hypothesis is true *and* the possibility that other hypotheses (including the alternative hypothesis) are true.
4.
   a. **H_0: $\mu_1 = \mu_0 = 285$; H_1: $\mu_1 \neq \mu_0$**
   Critical value: $t_{.05}(19) = \pm 2.093$
   According to the data, we retain the null hypothesis. The average level of shock people were willing to administer to others today was 240 volts, which was not significantly different from the average population level of 285 volts in the early 1960s ($p > .05$). The effect size index, $d = 0.45$, was medium.

   APA format: The results of the *t* test revealed no statistically significant difference between people's average willingness to shock others in the 1960s ($\mu = 285$ volts) and people's average willingness to shock others today ($M = 240$), $t(19) = -2.00$ $p > .05$, $d = 0.45$. The magnitude of the difference was medium.

   b. **H_0: $\rho = .00$; H_1: $\rho \neq .00$**
   Critical value: $r_{.05}(120) = \pm .1779$
   According to the data, we reject the null hypothesis. There is a positive correlation ($r = .57$) between the degree to which smartphone users associate their phones with positive personal memories and levels of nomophobia and the correlation is significantly different from .00, $p < .05$.

   APA format: There is a significant positive correlation between the degree to which smartphone users associate their phones with positive personal memories and levels of nomophobia ($r(299) = .57$, $p < .05$). Thus, the more positive personal memories people associate with their phones, the higher they tend to score on a measure of nomophobia.

   For this problem, note that Table A in the textbook does not provide a critical value for $df = 299$. When that happens, always use the next lowest *df* option, which is $df = 120$ for this problem. Although we used the critical value associated with $df = 120$, report the actual *df* in the APA-formatted conclusion. At $\alpha = .05$ for a two-tailed test at 120 *df*, the absolute value of the correlation must equal or exceed .1779. Thus, the correlation is significant at the .05 level. It is also important to remember that the context of the research question is important when interpreting the magnitude (effect size) of a correlation coefficient.
5.

a. retain the null hypothesis, the correlation is not significantly different from .00
b. reject the null hypothesis, the correlation is significantly different from .00
c. retain the null hypothesis, the correlation is not significantly different from .00
d. retain the null hypothesis, the correlation is not significantly different from .00
e. reject the null hypothesis, the correlation is significantly different from .00

Recall that the null hypothesis for testing the significance of a correlation coefficient is H_0: $\rho = .00$ and $df = N - 2$. If a correlation coefficient (absolute value) equals or exceeds the critical value reported in Table A (based on df and α), the correlation is significantly different from .00 and we reject the null hypothesis. If a correlation coefficient (absolute value) is less than the critical value reported in Table A, the correlation is *not* significantly different from .00 and we retain the null hypothesis.

a. ±.3493
b. ±.1946
c. ±.6055
d. ±.2500
e. ±.3542

Problems

1. H_0: $\mu_1 = \mu_0$
   H_1: $\mu_1 \neq \mu_0$

   $t = 3.71$; $d = 1.18$

   According to the data, we reject the null hypothesis. The average perceived stress level of college students experiencing academic difficulty was 18.6, which was significantly greater than the population average of 16.04 for those aged 18-25 ($p < .05$). The effect size index, $d = 1.18$, was large. Thus, students experiencing academic difficulty also experience more stress than a typical 18-25 year old.

   APA format: The average perceived stress level of college students experiencing academic difficulty ($M = 18.6$) is significantly higher than the average perceived stress level of individuals aged 18-25 ($\mu = 16.04$), $t(9) = 3.71$, $p < .05$. The size of the difference was large, $d = 1.18$, suggesting students experiencing academic difficulty also experience a lot more stress than a typical 18-25 year old.

X	X^2
16	256
17	289
17	289
19	361
18	324
23	529
21	441
20	400
17	289
18	324
$\Sigma X = 186$	$\Sigma X^2 = 3{,}502$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{186}{10} = 18.6$$

$$\hat{s} = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}} = \sqrt{\frac{3{,}502 - \frac{(186)^2}{10}}{10-1}}$$
$$\sqrt{\frac{3{,}502 - 3{,}459.6}{9}} = \sqrt{\frac{42.4}{9}} = 2.17$$

$$s_{\bar{X}} = \hat{s}/\sqrt{N} = \frac{2.17}{\sqrt{10}} = 0.69$$

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{X}}} = \frac{18.6 - 16.04}{0.69} = \frac{2.56}{0.69} = 3.71$$

$$d = \frac{\bar{X} - \mu_0}{\hat{s}} = \frac{18.6 - 16.04}{2.17} = \frac{2.56}{2.17} = 1.18$$

$df = N - 1 = 10 - 1 = 9$
Critical value: $t_{.05}(9) = \pm 2.262$

2. H_0: $\mu_1 = \mu_0$
   H_1: $\mu_1 \neq \mu_0$

   $t = -0.11$; $d = 0.002$

   According to the data, we retain the null hypothesis. The average SAT score for students at this high school was 1,056, which was not significantly different from the population average of 1,060 ($p > .05$). The effect size index, $d = 0.002$, was very small. Thus, students at this high school performed at the same level as the national average.

   APA format: There is not a statistically significant difference between the average SAT score of the small high school ($M = 1{,}056$) and the national average ($\mu = 1{,}060$), $t(25) = -0.11$, $p > .05$. The size of the difference was very small, $d = 0.002$. Thus, students at this high school performed at the same level as the national average.

$$\bar{X} = \frac{\Sigma X}{N} = \frac{27{,}456}{26} = 1{,}056$$

$$\hat{s} = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}} = \sqrt{\frac{29{,}900{,}000 - \frac{(27{,}456)^2}{26}}{26-1}}$$
$$\sqrt{\frac{29{,}900{,}000 - 28{,}993{,}536}{25}} = \sqrt{\frac{906{,}464}{25}} = 190.42$$

$$s_{\bar{X}} = \hat{s}/\sqrt{N} = \frac{190.42}{\sqrt{26}} = 37.34$$

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{X}}} = \frac{1{,}056 - 1{,}060}{37.34} = \frac{-4}{37.34} = -0.11$$

$$d = \frac{\bar{X} - \mu_0}{\hat{s}} = \frac{1{,}056 - 1{,}060}{190.42} = \frac{-4}{190.42} = |-0.002|$$

$df = N - 1 = 26 - 1 = 25$
Critical value: $t_{.05}(25) = \pm\ 2.060$

A two-tailed alternative hypothesis is best here in order to be able to test whether the high school is better *or* worse than the national average. If you found 1.708 as your critical value, that is the appropriate critical value for a *one*-tailed test, but it would not change your conclusion.

3. $t(41) = 6.67$; $d = 1.03$

For participants in the dimly-lit room, we reject the null hypothesis that they accurately reported on the number of math problems they solved. The average number of math problems they reported solving was 11.47, which was significantly higher than the norm of 7 ($p < .05$). The effect size index, $d = 1.03$, was large. Thus, being in a dark room resulted in cheating behavior.

APA format: On average, participants in the dimly-lit room reported completing significantly more math problems ($M = 11.47$, $SD = 4.32$) than the norm of 7, $t(41) = 6.67$, $p < .05$, $d = 1.03$. The size of this effect was large. Thus, being in a dark room resulted in cheating behavior.

$$s_{\bar{X}} = \hat{s}/\sqrt{N} = \frac{4.32}{\sqrt{42}} = 0.67$$

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{X}}} = \frac{11.47 - 7.00}{0.67} = \frac{4.47}{0.67} = 6.67$$

$$d = \frac{\bar{X} - \mu_0}{\hat{s}} = \frac{11.47 - 7.00}{4.32} = \frac{4.47}{4.32} = 1.03$$

$df = N - 1 = 42 - 1 = 41$
Critical value: $t_{.05}(40) = \pm\ 2.021$

For this problem, note that Table D in the textbook does not provide a critical value for $df = 41$. When that happens, always use the next lowest df option, which is $df = 40$ for this problem. Although we used the critical value associated with $df = 40$, report the actual df in the APA-formatted conclusion.

4. $t(25) = -4.76$; $d = 0.92$

We reject the null hypothesis that participants wearing sunglasses fairly divided money between themselves and a stranger. The average amount of money these participants gave to a stranger during the dictator game was \$1.81, which was significantly lower than the fair division of \$3.00 ($p < .05$). The effect size index, $d = 0.92$, was large. Thus, wearing dark sunglasses resulted in self-interested behavior.

APA format: On average, participants who wore sunglasses while playing the dictator game gave significantly less than \$3.00 (the fair division of money) to a stranger ($M = \$1.81$, $SD = \$1.30$), $t(25) = -4.76$, $p < .05$, $d = 0.92$. The size of this effect was large. Thus, wearing dark sunglasses resulted in self-interested behavior.

$$s_{\bar{X}} = \hat{s}/\sqrt{N} = \frac{1.30}{\sqrt{26}} = 0.25$$

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{X}}} = \frac{1.81 - 3.00}{0.25} = \frac{-1.19}{0.25} = -4.76$$

$$d = \frac{\bar{X} - \mu_0}{\hat{s}} = \frac{1.81 - 3.00}{1.30} = \frac{-1.19}{1.30} = |-0.92|$$

$df = N - 1 = 26 - 1 = 25$
Critical value: $t_{.05}(25) = \pm 2.060$

References

Cohen, S., & Janicki-Deverts, D. (2012). Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. *Journal of Applied Social Psychology, 42*(6), 1320-1344. doi: 10.1111/j.1559-1816.2012.00900.x

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of psychological stress. *Journal of Health and Social Behavior*, *24*, 385-396.

Cohen, S., & Williamson, G. (1988). Psychological stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health: Claremont symposium on applied social psychology* (pp. 31–67). Newbury Park, CA: Sage.

College Board. (2017, September 26). *SAT suite of assessments annual report*. Retrieved from https://reports.collegeboard.org/pdf/2017-total-group-sat-suite-assessments-annual-report.pdf

Han, S., Kim, K. J., & Kim, J. H. (2017). Understanding Nomophobia: Structural equation modeling and semantic network analysis of smartphone separation anxiety. *Cyberpsychology, Behavior, and Social Networking, 20*(7), 419-427. doi: 10.1089/cyber.2017.0113

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67,* 371-378.

Zhong, C., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science, 21*(3), 311-314. doi: 10.1177/0956797609360754