CHAPTER 16

Choosing Tests and Writing Interpretations

Summary

Some of the most important tasks facing researchers are deciding on the relevant statistics for their research questions, conducting statistical tests, and interpreting and explaining results. The textbook gives you several ways to think about putting together all you've learned to help you choose appropriate statistical tests. To consolidate your new knowledge, you may want to re-read sections of the textbook, look back at notes and assignments, and construct a table or decision tree to help you approach new problems.

You can also approach the decision about the correct statistical technique for a research design by starting with a series of questions to help orient yourself to the critical details of the problem. You are encouraged you ask yourself these questions:

1. *Are you trying to describe the nature of something or make an inference?* If you are trying to describe something, you will be using one of the descriptive statistics you learned. Ask yourself if you are looking to understand central tendency (the mean, median, and mode), variability (range, IQR, and standard deviation), effect size, relative standing within or across populations (*z* score), or relationships in two sets of data (correlation). If you are making inferences, you are likely to be conducting a *t* test, an ANOVA, a chi-square, or a nonparametric test. (This said, it is also possible that you could be conducting a NHST analysis of correlation that is designed to make an inference or you could be using regression to make future predictions.)

2. *What scale(s) of measurement is(are) being used in your analyses*? This question can help you identify the most appropriate measure of central tendency for a set of data. When conducting inferential tests, continuous quantitative *dependent variables* on an interval or ratio scale are required to conduct *t* tests and ANOVAs. This question can be particularly helpful when you are conducting an inferential test and you have nominal or ordinal data (or data that has been converted to either scale). If your data is on a nominal scale of measurement and consists of frequency counts, you will be conducting a chi square analysis (with one variable, a goodness-of-fit test; with two variables, a test of independence). If your data is on an ordinal scale of measurement and consists of ranks, you will be conducting one of the other nonparametric tests described in the textbook (for independent samples, a Mann-Whitney *U* test; for paired samples, a Wilcoxon signed-rank *T* test; for one independent variable with more than two levels, a Kruskal-Wallis test; to create a correlation coefficient, a Spearman's correlation coefficient r_s).

3. *Do you have information about relevant populations*? If you know information about μ and σ, you will be finding a *z* score. If you have information about μ, but not σ, you are likely conducting a one-sample *t* test.

4. *If you are using inferential statistics, how many samples are included in your set of data*? For interval or ratio data, if you have one sample, you will be conducting a one-sample *t* test (if you know μ, but not σ) or *z* test (if you know μ and σ). If you have two samples, you could be finding a correlation to describe relationships, using regression to make predictions, or conducting a two-sample *t* test to determine the difference between two populations. If you have three or more samples, you will be conducting an ANOVA. For ranked data, remember, there are also parallels of nonparametric tests that depend on the number of samples as well.

5. *If you have multiple samples, are they paired or independent*? This will help in distinguishing between independent- and paired-samples *t* tests, between one-way ANOVA and one-factor repeated-measures ANOVA, and between Mann-Whitney U and Wilcoxon signed-rank T test.

6. *How many independent variables are there*? There are lots of possible tests if you have a study with one independent variable (single sample *t* and *z* tests, correlation and regression, two-sample

t tests, one-way ANOVA, one-factor repeated-measures ANOVA, chi-square goodness-of-fit test). If you have two or more, you will be conducting a factorial ANOVA. If you have two independent variables and all of the levels of the independent variables are independent, you will be conducting the 2-way factorial ANOVA you learned if the dependent variable is interval or ratio, or a chi-square test of independence if it is nominal or ordinal. However, if you have more than two independent variables, or if your independent variables do not have independent levels, you might need to consult an advanced text or colleague.

In this chapter of the study guide, you will first complete multiple choice problems that are designed to be a general refresher on the material you've learned in this course. Second, you will practice identifying the appropriate statistical test after being given a description of research. Third, you will practice writing conclusions about some of the statistical tests you've conducted in this study guide. These exercises are designed to help you see the way that the statistics you have learned work together to provide an essential tool for researchers.

To get the most from these exercises, read each problem and identify as many answers as possible. Then, go back to your notes, table, flow chart, or list of questions to help with the problems you couldn't complete on your first pass. This can be a helpful way to identify the topics you know well and the topics you will want to focus on refreshing in your memory.

Multiple Choice

1. The mean score of 25 participants who finished an anger management class was 95. The population mean for this test is 100. A one-sample *t* test produced a $t = 2.80$. The *best* conclusion is that ______.

a. there is no significant difference between participants in the anger management class and the population that has a mean of 100
b. the anger management class has a significant effect
c. participants in the anger management class scored significantly lower than the national mean, $p < .05$
d. cannot be determined from the information given

2. The correlation coefficient between two variables is $r = .30$. The proportion of variance the two variables have in common is ______.

a. .60
b. .30
c. .09
d. none of the above

3. Fifty children with a parent in the military were assessed for social functioning before the parent was deployed overseas (mean score = 75) and again just before the parent returned (mean score = 66). A Wilcoxon signed-rank *T* test produced a $T = 450$. For a one-tailed test with $\alpha = .05$, the *best* conclusion is that ______.

a. while a parent is deployed, children's social scores increase significantly, $p < .05$
b. while a parent is deployed, children's social scores decline significantly, $p < .05$
c. there is no effect of deployment on children's social scores
d. none of the above

<u>Research Description 16-1</u>

Does the moon really appear larger on the horizon than at its highest point in the sky (zenith)? Participants judged the size of the rising moon and later the zenith moon by comparing each to a standard circle. On average, the rising moon was judged 1.7 times larger than the circle; the zenith moon was judged 1.4 times larger than the circle. A paired-samples *t* test with *df* = 19 resulted in a *t* value of 2.60. Assume you were conducting a NHST with a two-tailed significance test and an $\alpha = .05$.

4. Based on Research Description 16-1, the *best* conclusion is that ______.

a. there is no difference in judged size between the rising moon and the zenith moon
b. the difference in judged size between the rising moon and zenith moon is not statistically significant, $p > .05$
c. there is a significant difference between the rising moon and the zenith moon
d. the rising moon is judged significantly larger than the zenith moon, $p < .05$

5. In Research Description 16 -1, the dependent variable is ______.

a. where the moon is in the sky
b. judgments about the size of the moon
c. the number of participants in the experiment
d. none of the above

6. How many participants were in the study described in Research Description 16-1?

a. 19
b. 20
c. 40
d. cannot be determined from the information given

Research Description 16-2

Shoppers judged the legibility of a flashing advertisement. For 16 of the shoppers, the ad was in capital letters and the mean legibility score was 1.62. For 26 of the shoppers, the ad was in lower case letters and the mean legibility score was 2.04. A two-tailed NHST analysis was conducted and the test value was 2.00.

7. The appropriate statistical analysis for Research Description 16-2 is ______.

a. an independent-samples *t* test
b. a paired-samples *t* test
c. a one-sample *t* test
d. none of the above

8. For this study, *df* = ______.

a. 38
b. 41
c. 40
d. 42

9. Based on Research Description 16-2, the *best* conclusion is that ______.

a. there is no statistically significant difference in the legibility of ads in capital letters and ads in lower case letters
b. ads in lower case letters are significantly more legible than ads in capital letters, $p < .05$
c. ads in lower case letters are significantly more legible than ads in capital letters, $p < .01$
d. cannot be determined from the information given

10. The independent variable in Research Description 16-2 is ______.

a. The 42 shoppers

b. legibility scores
c. the type of letters in the ad
d. none of the above

Research Description 16-3

Three groups of young rats were tested for learning ability after being raised in different living conditions. One group was raised in a standard environment, a second in a complex environment, and a third in a social environment with other rats. The mean scores on a learning test were: $\bar{X}_{standard} = 12$; $\bar{X}_{complex} = 15$; $\bar{X}_{social} = 20$. The ANOVA conducted on your data produced $F(2, 42) = 3.25$.

11. Based on the data provided in Research Description 16-3, the *best* conclusion is that ______.

a. living conditions do not produce a significant effect on learning ability of rats
b. living conditions produce a significant effect on learning ability of rats, $p < .05$
c. rats raised under social conditions have significantly higher scores than those raised in a complex environment, $p < .05$
d. rats raised under standard conditions have significantly lower scores than those raised in a complex environment, $p < .05$

12. For the probability figures to be correct for the *F* test in Research Description 16-3, which two characteristics of the population data must be true?

a. the population variances must be equal and the distributions must show a positive skew
b. the population variances must be equal and the distributions must be normally distributed
c. the population variances must not be equal and the distributions must show a positive skew
d. the population variances must not be equal and the distributions must be normally distributed

Research Description 16-4

Based on the genetics theory of "random assortment", fruit fly offspring will have a particular pattern of characteristics. When *actual* offspring were compared to those predicted by the theory of random assortment, a chi square value of 12.57 was obtained with $df = 7$.

13. For Research Description 16-4, the null hypothesis should be ______.

a. retained if $\alpha = .05$
b. rejected if $\alpha = .05$
c. rejected if $\alpha = .01$
d. rejected if $\alpha = .001$

14. The interpretation of Research Description 16-4 is that ______.

a. there is no significant difference between the data and the theory of random assortment, $p > .05$
b. the theory of random assortment does not fit these data, $p < .05$
c. the theory of random assortment does not fit these data, $p < .01$
d. none of the above

Research Description 16-5

A 2x2 chi square test of independence for gender and assertiveness produced a chi square value of 2.85.

15. The number of degrees of freedom for the chi square value in Research Description 16-5 is ______.

a. 1
b. 2

c. 4
d. cannot be determined from the information given

16. Based on Research Description 16-5, the *best* conclusion is ______.

a. males are more assertive than females, $p < .05$
b. females are more assertive than males, $p < .05$
c. gender and assertiveness are related to each other, $p < .05$
d. there is no significant relationship between gender and assertiveness, $p > .05$

Choosing Tests

Descriptions of 30 hypothetical studies follow. For each study description, choose the most appropriate statistical test for the data and identify the pieces of information in the question that helped you choose the test.

1. Idara was interested in examining the development of maturity from senior year in high school to college. She had her participants complete a maturity scale that yielded a score from 0 to 100, with higher numbers indicating greater maturity. Her participants were a group of seniors from an all-male high school and a group of first-year college students at an all-male college.

2. Salima wanted to know if athletic success is related to academic success at Central High School. She collected information about the performance of the students on the track team, by identifying the order in which runners finished a race against one another, as well as the students' class rank.

3. Lia and Oliver wanted to determine if packages of artificial sweetener actually contain the amount of product listed on the packaging (.035 oz). To test this, they weighed 30 packages of the sweetener that were available in their school cafeteria.

4. A group of faculty was interested in predicting the average incoming SAT scores of students, based on their high school GPA. The starting point of this prediction was using correlation data between these two variables from the previous 25 years of students on the college campus.

5. Carlos and Dylan wanted to know if auto mechanics and car salespeople have similar incomes. They gathered data about income from a group of auto mechanics and a group of car salespeople. Income data are known to be *strongly* positively skewed.

6. Josie and Kamal had participants read a series of short stories. Three of the stories were inconsistent with previously read material, three were consistent with previously read material, and three were unrelated to previously read material. Each participant read all of the stories and their reading time was measured. The researchers calculated an average reading time for inconsistent, consistent, and unrelated stories for each participant.

7. Nita wanted to know if male or female travelers were more likely to eat breakfast, lunch, or dinner in the airport food court. For three weeks, she counted the number of male and female travelers at each meal.

8. Elisa and Lin measured the reaction time of participants who read words as fast as they could from a computer screen and participants' performance was ranked. They tested equal numbers of men and women to determine which group had faster reaction times.

9. Mellie wanted to know if factory worker income was related to their years of service. Factory workers were asked to report how long they had worked for the company and their income.

10. Zahra wanted to know if the department in which people take the same class makes a difference in their content knowledge. She asked students who had taken a statistics course whether they took the course in the math department or the psychology department. Then, she gave students a questionnaire measuring knowledge of statistics.

11. Hasan conducts research on 2000 people and finds that women scored better on an IQ test than men. He wants to know *how much* greater the IQ score of women is than men.

12. Participants drank one cup of decaffeinated and one cup of caffeinated coffee over the course of two hours. Thirty minutes after each cup of coffee, James measured participants' reaction time. As a variable, reaction time is *strongly* positively skewed.

13. Bakari gave participants a test of their knowledge of U.S. history prior to taking an introductory college course in history to test the effect of the course on knowledge of history. After finishing the course, the participants took the same test again.

14. Catalina wants to know if the number of writing classes she has taken might predict her income when she is 40 years old. She collects income data from 100 40-year-old women who took different numbers of college writing classes to help her make this prediction.

15. Lindsay wanted to know if there was a difference in hearing ability between left- and right-handed students. She tested the hearing ability of left-handed and right-handed students by determining the smallest number of decibels the student could hear.

16. Rashida wanted to study if height is related to assertiveness. She asked her research participants to complete an assertiveness questionnaire (with possible scores ranging from 0-25) and to report their height.

17. Laila studied the relationship between gender and creativity. She had boys and girls build anything they wanted to with wood blocks and then the structures were ranked for creativity by an artist.

18. Maia wants to know whether time of day and mood will impact people's willingness to help. Half of her participants attend the study at 10 AM and half of the participants attend the study at 6 PM. In addition, half of the participants at each time watch a video clip designed to make them happy and the other half watch video clip designed to make them feel sad. As participants are leaving, they are asked to sign up for a number of community volunteer projects in their community in the coming week. Maia uses the amount of time they sign up for as the dependent variable.

19. Leslie asked her research participants to study a list of complicated words in a small group or a large group. After studying for five minutes, she administered a spelling test and scores on the test ranged from 0-10.

20. Masaki wants to study the relationship between socioeconomic status (SES) and college performance. Based on socioeconomic data, he splits his participants into three groups: high SES, middle SES, and low SES. He also collects information about each participant's college class rank.

21. Anusheh studied how physical activity impacted stress levels. Participants were assigned to do 15 minutes of vigorous exercise (like running on a treadmill), 15 minutes of gentle exercise (walking on a treadmill), or 15 minutes of sedentary reading. She then had participants give a saliva sample and cortisol (a hormone associated with stress) levels were measured.

22. Lei was interested in examining the pattern of visits to the school nurse among 3rd, 4th, and 5th grade students in an elementary school. She thought that 3rd graders would be twice as likely to visit the nurse as older students.

23. Jennifer wanted to know if there is a relationship between sleep and headaches. She asked 100 students how much they sleep on a daily basis and number of headaches in the past month.

24. Kwame paired participants with identical GPAs. One member of each pair napped regularly, the other didn't. Each participant took an SAT-style test.

25. Chris wanted to know if there is a relationship between quality of life and cost of living. He ranked 30 cities on both dimensions.

26. Kai wanted to study the relationship between music and test performance. The participants, who were matched on GPA, were assigned to listen to classical music, hip-hop, or heavy metal while they studied statistics. Later, the participants took a 30-item statistics test.

27. A manufacturer claimed their cans of corn weigh 206 grams. Cruz used the weights of 10 cans as a sample to test this manufacturer claim.

28. Claire examined the influence of sensory classroom experiences on memory. Students heard a five-minute presentation in a room that varied in terms of its temperature (65°F or 80°F) and its lighting (dim or normal). Participants then answered questions about what they remembered from the presentation.

29. Yael is interested in the importance of tutoring for students learning a foreign language. She has data for one semester about how many times students in Introductory Spanish visited the tutor, as well as their numerical course grade (out of 100) at the end of the semester. She wants to predict the grade of a student who went to tutoring 5 times during the semester.

30. Using income data, which are strongly positively skewed, Sara compared the incomes of doctors and lawyers in her town.

Writing Interpretations

The problems that follow will present you with summary data for analyses you may have already conducted in previous chapters of the study guide. Your task is use the summary data to write strong interpretations for each problem *without looking back on your work from previous chapters*.

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 used in this problem are PSS-10 scores for college students experiencing academic difficulty. The summary data are as follows:

N = 10
$\bar{X}$ = 18.60
t = 3.71
d = 1.18
Critical value: $t_{.05}(9) = \pm 2.262$

2. Yang and Urminsky (2018) studied how anticipated affective reactions, like smiling, can motivate different gift choices. In one part of their study, participants were asked to imagine the affective reactions that would be elicited by a personalized mug (X) and an award-winning ergonomic mug (Y). Participants were asked "How much of an affective reaction would receivers show in response to these gifts when receiving them?" and made a rating of each item. Below are summary data from Study 1 of Yang and Urminsky (2018) that closely replicate their findings.

N = 213
$\bar{X}$ = 6.00
$\bar{Y}$ = 5.00
t = 7.69
d = 0.55
Critical value: $t_{.05}(120) = \pm 1.980$

3. Hanna, Kee, and Robertson (2017) examined whether use of social media to connect with co-workers was associated with differences in job satisfaction. They assessed the amount of time participants connected with co-workers by asked each participating employee ''In a typical work week, approximately how much time per day (including during breaks, after work, during the weekend, etc.) do you spend on Facebook interacting with people from work?'' They divided their participants into three groups based on their answers to this question: less than 10 minutes (X_{Low}), 10-30 minutes (X_{Med}), and 31 minutes to 2 hours (X_{High}). They also assessed their job satisfaction using a set of five questions, with higher scores meaning greater job satisfaction. The data below are hypothetical job satisfaction scores for five participants in each group created so the conclusions and means that you reach will mimic those of Hanna and colleagues (2017).

$\bar{X}_{Low}$ = 3.00	Critical value: $\text{HSD}_{.05}(3, 12) = \pm 3.77$	$d_{low\ v.med} = -0.72$
$\bar{X}_{Med}$ = 3.80	$\text{HSD}_{low\ v.med} = -1.60$	$d_{low\ v.high} = -1.80$
$\bar{X}_{High}$ = 5.00	$\text{HSD}_{low\ v.high} = -4.00$	$d_{med\ v.high} = -1.08$
Critical value: $F_{.05}(2, 12) = 3.88$	$\text{HSD}_{med\ v.high} = -2.40$	η^2 = 0.64
F = 4.12		

4. T. S. Spatz (1991) conducted workshops teaching women to conduct breast self-examinations. She gave participants a 27-item true/false test of their knowledge before they were trained (pre-test), after they were trained (post-test), and again three months later (follow-up). The scores below mimic her results.

$\bar{X}_{Pre}$ = 12.0	Critical value: $F_{.05}(2, 10) = 7.56$	Critical value: $HSD_{.01}(3, 10) = \pm 5.27$
$\bar{X}_{Post}$ = 23.0	F = 37.25	$HSD_{pre\ v.post} = -11.46$
$\bar{X}_{Follow-up}$ = 21.00	η_p^2 = 2.01	$HSD_{pre\ v.follow\ up} = -9.38$
		$HSD_{post\ v.follow\ up} = 2.08$

5. Maglio and Polman (2014) studied the effects of spatial orientation on estimates of psychological distance. In their second study, researchers approached people at a subway station in Toronto, Ontario who were either on the eastbound platform (headed *toward* the University of Toronto campus) or the westbound platform (heading *away* from the University of Toronto campus). Researchers then asked each participant to either face toward or away from the University of Toronto and rate how far away they were from the campus on a scale of 1 (*very close*) to 7 (*very far*). Below are summary data from Study 2 of Maglio and Polman (2014; retrieved from https://osf.io/7rajd/ on July 13, 2018) that will allow you to replicate the results.

Means for Direction Traveling (A)	Means for Direction Facing (B)
$\bar{X}_{toward} = 3.63$	$\bar{X}_{toward} = 3.98$
$\bar{X}_{away} = 4.15$	$\bar{X}_{away} = 3.80$

Critical values: $F_{.05}(1, 70) = 3.98$ (same for all three tests)

6. In 2015, staff members of the American Psychological Association's (APA's) Center for Workforce Studies published a report indicating that as of 2013, for every male active psychologist, there were 2.1 female active psychologists in the workforce. A college professor had noticed that she had far more female students than male students and wondered if the gender makeup for her undergraduate students matched the findings from the APA's report or if by the time her students made it into the workforce the ratio would be even larger. She asked 310 of her students to self-identify their gender. If the students match the report, she is expecting 210 females and 100 males in her student sample. Below find her data. Let α = .05.

	Gender	
	Male	Female
Psychology Majors	77	233

Critical value: $x^2_{.05} = 3.841$
$x^2(1) = 7.81$

7. This problem uses a 2018 "Best Countries" dataset from U. S. News and World Report. Imagine that two young entrepreneurs were deciding whether to base their company in Europe or Asia. Using the "Best

Countries" report, the entrepreneurs selected the 10 highest-ranked countries from each continent and then gathered the "open for business" score for each (which was based on factors such as having a favorable tax environment and transparent government practices). Use these "open for business" scores (lower scores mean more open to business) to determine which continent is likely to be a better place to start their new company.

"Open for Business" Scores			
European Countries		Asian Countries	
Switzerland	2	Japan	26
Germany	19	Singapore	13
United Kingdom	20	China	33
Sweden	5	South Korea	25
France	27	United Arab Emirates	67
Netherlands	9	India	29
Denmark	4	Russia	80
Norway	8	Thailand	15
Finland	6	Israel	64
Italy	45	Malaysia	12

Critical values: $U_{.05} = 23.0$
$U = 18$

ANSWERS
Multiple Choice Questions

1. c. participants in the anger management class scored significantly lower than the national mean, $p < .05$
   Explanation: First look up a critical t value, here $t_{.05}$ (99) = 1.990. Determine that there is a significant effect. Option c. is better than option b. because it is better to include the probability value and describe the direction of the difference than to make the general conclusion reported in b.
2. c. .09
   Explanation: Remember, the coefficient of determination (r^2) tells you the proportion of variance that is shared between two variables.
3. b. while a parent is deployed, children's social scores decline significantly, $p < .05$
   Explanation: For a one-tailed test, $T_{.05}(50) = 466$ and your obtained value is smaller, so there is a significant effect. And, by comparing the means identified in the problem, you can see that scores were lower during deployment than before deployment.
4. d. the rising moon is judged significantly larger than the zenith moon, $p < .05$
   Explanation: For this experiment, $t_{.05}(19) = 2.093$. So, the probability of getting your obtained t value is < .05 if the null hypothesis were true. Although c. is accurate, conclusions are stronger when they specify the direction of the difference and the probability value.
5. b. judgments about the size of the moon
6. b. 20
   Explanation: Degrees of freedom for a paired-samples t test are N - 1, with N being the number of pairs. Adding 1 to the 19 degrees of freedom means there were 20 participants.
7. a. an independent-samples t test
   Explanation: One easy way to tell this is an independent-samples design is that the two samples have unequal Ns.
8. c. 40
   Explanation: The description states that one sample contained 16 individuals, whereas the other sample contained 26 individuals. Unequal sample sizes—plus no indication of pairing or repeated measures—tells you an independent-samples t test is appropriate. For that analysis, $df = N_1 + N_2 – 2$.
9. a. there is no statistically significant difference in the legibility of ads in capital letters and ads in lower case letters
   Explanation: For this experiment, $t_{.05}(40) = 2.021$. So, the probability of getting the obtained t value is < .05 if the null hypothesis were true.
10. c. the type of letters in the ad
11. b. living conditions produce a significant effect on learning ability of rats, $p < .05$
    Explanation: The critical value $F_{.05}(2, 42) = 3.22$, so you found a significant effect of learning conditions. You can rule out answers c. and d. because you do not have any information from Tukey HSD tests that would allow you to speak about pairwise comparisons.
12. b. the population variances must be equal and the distributions must be normally distributed
13. a. retained if $\alpha = .05$
    Explanation: The critical value for $\chi^2_{.05}(7) = 14.067$, which is larger than your obtained value, so you will retain the H_0.
14. a. there is no significant difference between the data and the theory of random assortment, $p > .05$
15. a. 1
    Explanation: Because you know this is a 2x2 test, you know there are two categories of each variable. As such, $df = (R – 1)(C – 1) = (2 – 1)(2 – 1) = 1$.
16. d. there is no significant relationship between gender and assertiveness, $p > .05$
    Explanation: For this study, $\chi^2_{.05}(1) = 3.841$. So, the probability of getting the obtained χ^2 value is > .05 if the null hypothesis were true.

Choosing Tests

1. Independent-samples *t* test

   You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (maturity score) on a ratio or interval scale, c) you do not have any population information, d) there are two samples of participants (high school and college students), e) there is no pairing, and f) there is one independent variable (level in school). This study doesn't have an experimentally manipulated independent variable, but you do have two different samples providing scores on the same dependent variable.
2. Spearman's *r*

   You know this is the correct test because a) you are describing a relationship between two variables, b) the data for both variables are ranks, which makes this a nonparametric test, c) you have bivariate data for each participant, and d) scores are paired. [Note: while correlations were introduced in the textbook as a descriptive statistic, we can also conduct an inferential NHST analysis using a correlation.]
3. One-sample *t* test

   You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (weight) on a ratio or interval scale, c) you have a population mean (μ) and not a population standard deviation (σ), and d) there is one sample.
4. Regression

   You know this is the correct test because a) you already have information that describes the relationships in question, b) you are trying to make an inference about future information (a prediction), c) you are measuring two continuous quantitative variables (SAT score and GPA) on a ratio or interval scale, and d) you have bivariate data that you are using to make a prediction.
5. Mann-Whitney *U* test (after converting income into ranks)

   You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (income) on a ratio or interval scale, c) your data is skewed, so you cannot assume that the variable is normally distributed (which would be required to use a *t* test or ANOVA), d) you do not have any population information, e) there are two samples of participants (auto mechanics and car salespeople), f) there is no pairing, and g) there is only one independent variable. This study doesn't have an experimentally manipulated independent variable, but you do have two different samples providing scores on the same dependent variable.
6. One-factor repeated-measures ANOVA

   You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (reading time) on a ratio or interval scale, c) you do not have any population information, d) there are three samples of participants (inconsistent, consistent, and unrelated stories), e) there is repeated-measures pairing, and f) there is one experimentally manipulated independent variable.
7. Chi Square test of independence

   You know this is the correct test because a) you are trying to make an inference, b) your data consists of frequencies, and c) you have two variables of interest (gender and meal) and you are testing whether they are related.
8. Mann-Whitney *U* test

   You know this is the correct test because a) you are trying to make an inference, b) the dependent variable is ranks, c) you do not have any population information, d) there are two samples of participants (males and females), e) there is no pairing, and f) there is only one independent variable. This study doesn't have an experimentally manipulated independent variable, but you do have two different samples providing scores on the same dependent variable.
9. Pearson product-moment correlation coefficient

You know this is the correct test because a) you are describing a relationship between two variables, b) you are measuring two continuous quantitative variables (income and years of service) on a ratio or interval scale, c) you have bivariate data for each participant, and d) scores are paired. [Note: while correlations were introduced in the textbook as a descriptive statistic, we can also conduct an inferential NHST analysis using a correlation.]

10. Independent-samples *t* test

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (statistics knowledge) on a ratio or interval scale, c) you do not have any population information, d) there are two samples of participants (students who took statistics in the math department and students who took statistics in the psychology department), e) there is no pairing, and f) there is only one independent variabale. This study doesn't have an experimentally manipulated independent variable, but you do have two different samples providing scores on the same dependent variable.

11. Effect size *d*

    You know this is the correct test because a) you want to describe something, b) you are measuring a continuous quantitative variable (IQ) on a ratio or interval scale, c) you have two samples, and d) you are looking for the size of the difference between the two samples.

12. Wilcoxon signed-rank *T* test

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (reaction time) on a ratio or interval scale, c) your data is skewed, so you cannot assume that the variable is normally distributed (which would be required to use a *t* test or ANOVA), d) you do not have any population information, e) there are two samples of data (performance after decaffeinated coffee and performance after caffeinated coffee), f) there is repeated-measures pairing, and g) there is one experimentally manipulated independent variable.

13. Paired-samples *t* test

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (test knowledge) on a ratio or interval scale, c) you do not have any population information, d) there are two samples of data (pre-test and post-test), e) there is repeated-measures pairing, and f) there is one experimentally manipulated independent variable.

14. Regression

    You know this is the correct test because a) you already have information that describes the relationships in question, b) you are trying to make an inference about future information (a prediction), c) you are measuring two continuous quantitative variables (income and number of courses) on a ratio or interval scale, and d) you have bivariate data for each participant that you are using to make a prediction.

15. Independent-samples *t* test

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (detectable audible decibels) on a ratio or interval scale, c) you do not have any population information, d) there are two samples of participants (left-handed and right-handed students), e) there is no pairing, and f) this is only one independent variable. This study doesn't have an experimentally manipulated independent variable, but you do have two different samples providing scores on the same dependent variable.

16. Pearson product-moment correlation coefficient

    You know this is the correct test because a) you are describing a relationship between two variables, b) you are measuring two continuous quantitative variables (assertiveness and height) on a ratio or interval scale, c) you have bivariate data for each participant, and d) scores are paired.

[Note: while correlations were introduced in the textbook as a descriptive statistic, we can also conduct an inferential NHST analysis using a correlation.]

17. Mann-Whitney *U* test

    You know this is the correct test because a) you are trying to make an inference, b) the dependent variable is ranks, c) you do not have any population information, d) there are two samples of participants (males and females), e) there is no pairing, and f) there is only one independent variable. This study doesn't have an experimentally manipulated independent variable, but you do have two different samples providing scores on the same dependent variable.

18. Factorial ANOVA

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (time volunteering) on a ratio or interval scale, c) you do not have any population information, d) there are four samples of participants (morning happy, morning sad, night happy, night sad), e) the samples are independent, and f) there are two experimentally manipulated independent variable (time and mood).

19. Independent-samples *t* test

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (spelling test score) on a ratio or interval scale, c) you do not have any population information, d) there are two samples of participants (small group and large group studying), e) there is no pairing, and g) there is one experimentally manipulated independent variable.

20. Wilcoxon-Wilcox multiple comparisons test

    You know this is the correct test because a) you are trying to make an inference, b) one of your variables is ranks, c) you do not have any population information, d) there are three samples of participants (high, medium, and low SES), and e) there is no pairing. This study doesn't have an experimentally manipulated independent variable, but you do have three different samples providing scores on the same dependent variable.

21. One-way ANOVA

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (cortisol level) on a ratio or interval scale, c) you do not have any population information, d) there are three samples of participants (vigorous exercise, gentle exercise, sedentary), e) the samples are independent, and g) there is one experimentally manipulated independent variable.

22. Chi Square goodness-of-fit test

    You know this is the correct test because a) you are trying to make an inference, b) your data consists of frequencies, c) you have one variable of interest (grade in school), and d) you are comparing your observed frequencies to a theory (that frequency of visit would be equal across grades).

23. Pearson product-moment correlation coefficient

    You know this is the correct test because a) you are describing a relationship between two variables, b) you are measuring two continuous quantitative variables (hours of sleep and number of headaches) on a ratio or interval scale, c) you have bivariate data for each participant, and d) scores are paired. [Note: while correlations were introduced in the textbook as a descriptive statistic, we can also conduct an inferential NHST analysis using a correlation.]

24. Paired-samples *t* test

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (GPA) on a ratio or interval scale, c) you do not have any population information, d) there are two samples of participants (those who nap and those who don't), and e) this is a matched pairs design. This study doesn't have an experimentally manipulated independent variable, but you do have two different samples providing scores on the same dependent variable.

25. Spearman's r_s

    You know this is the correct test because a) you are trying to make an inference about a relationship, b) the data for both variables are ranks, which makes this a nonparametric test, c) you have bivariate data for each city, and d) scores are paired. [Note: while correlations were introduced in the textbook as a descriptive statistic, we can also conduct an inferential NHST analysis using a correlation.]

26. One-factor repeated-measures ANOVA

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (test score) on a ratio or interval scale, c) you do not have any population information, d) there are three samples of participants (those who listen to classical music, hip-hop, and heavy metal), e) the samples are matched pairs, and f) there is one experimentally manipulated independent variable.

27. One-sample *t* test

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (weight) on a ratio or interval scale, c) you have a population mean (μ), and d) there is one sample.

28. 2 X 2 Factorial ANOVA

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (memory score) on a ratio or interval scale, c) you do not have any population information, d) there are four samples of participants (cold room/dim lighting, cold room/normal lighting, warm room/dim lighting, warm room/normal lighting), e) the samples are independent, and f) there are two experimentally manipulated independent variable (temperature and lighting) each with 2 levels.

29. Regression

    You know this is the correct test because a) you already have information that describes the relationship in question, b) you are trying to make an inference about future information (a prediction), c) you are measuring two continuous quantitative variables (tuition cost and rate of increase) on a ratio or interval scale, and d) you have bivariate data you are using to make a prediction.

30. Mann-Whitney *U* test

    You know this is the correct test because a) you are trying to make an inference, b) you are measuring a continuous quantitative variable (income) on a ratio or interval scale, c) your data is skewed, so you cannot assume that the variable is normally distributed (which would be required to use a *t* test or ANOVA), d) you do not have any population information, e) there are two samples of data (doctors and lawyers), f) there is no pairing, and g) there is only one independent variable. This study doesn't have an experimentally manipulated independent variable, but you do have two different samples providing scores on the same dependent variable.

Writing Interpretations

1. 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 (μ = 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.

2. According to the data, we reject the null hypothesis. The average expected affective reaction for a personalized mug ($\bar{X}$ = 6.00) was significantly greater than the expected affective reaction for an ergonomic mug ($\bar{Y}$ = 5.00). The effect size index, *d* = 0.55, was medium. Thus, participants expected a stronger positive affective reaction after receiving a personalized gift.

   APA format: The average expected affective reaction to a personalized mug (*M* = 6.00) is significantly higher than the expected affective reaction to an ergonomic mug (*M* = 5.00), *t*(212) = 7.69, *p* < .05. The size of the difference was medium, *d* = 0.55.

3. According to the data, we reject the null hypothesis. Time spent connecting with co-workers through Facebook was significantly related to differences in job satisfaction, *F*(2, 12) = 4.12, *p* < .05. The effect size index, *f* = 0.64, was large. Follow-up Tukey HSD tests showed that people spending the lowest amount of time connecting with co-workers on Facebook (less than 10 minutes per day) were significantly less satisfied with their jobs ($\bar{X}_{Low}$ = 3.00) than people spending the highest amounts of time (31 minutes to 2 hours per day) ($\bar{X}_{High}$ = 5.00), HSD = -4.00. People spending medium amounts of time connecting with co-workers through Facebook (10 to 30 minutes per day) ($\bar{X}_{Medium}$ = 3.80) were not significantly different in job satisfaction from either the low (HSD = -1.60) or high (HSD = -2.40) groups. The difference in job satisfaction between the low and high users was large (*d* = 1.80), as was the difference between medium and high users (*d* = 1.08), even though this difference was not significant. The difference between the medium and low users was medium to large (*d* = 0.72). These researchers might consider repeating their study with a larger sample size to increase power.

   APA format: Time spent connecting with co-workers through Facebook had a large significant association with job satisfaction, *F*(2, 12) = 4.12, *p* < .05, *f* = 0.64. Employees who spend the least amount of time connecting to co-workers through Facebook (less than 10 minutes per day) had significantly lower job satisfaction (*M* = 3.00) than employees who spend the most time (31 minutes to 2 hours) connecting in this way (*M* = 5.00), HSD (3, 12) = -4.00, *p* < .05, *d* = 1.80. This difference was very large. Medium users (10 to 30 minutes per day) fell between these two groups on job satisfaction (*M* = 3.80). However the medium-use group did not significantly differ in job satisfaction from either the low-use group, HSD (3, 12) = -1.60, *d* = -0.72, or the high-use group, HSD (3, 12) = -2.40, *d* = 1.08 (both *p*s > .05), though the magnitudes of these differences were medium to large.

4. According to the data, we reject the null hypothesis. Participants had significantly less knowledge before training ($\bar{X}_A$ = 12.0) than during the post-test ($\bar{X}_B$ = 23.0) or at the follow-up three months later ($\bar{X}_C$ = 21.0). The effect size index, *f* = 2.01, was large. Thus, training increased the amount of information participants had. There was a significant difference between knowledge pre-test and post-test (HSD = -11.46) and between knowledge pre-test and at the follow-up test three months later (HSD = -9.38). However, there was no significant difference between knowledge between the post-test and at the follow-up test three months later (HSD = 2.08).

   APA format: Training significantly affected participant knowledge about breast self-examination. Participants had less knowledge on the pre-test (*M* = 12.0) than on the post-test (*M* = 23.0) or the follow-up test three months later (*M* = 21.0), *F* (2, 10) = 37.25, *p* < .01. The size of the effect was large, *f* = 2.01. There was a significant difference between knowledge on the pre-test and post-test, HSD(3, 10) = -11.46, *p* < .01. There was a significant difference between knowledge on the pre-test and the follow-up test, HSD(3, 10) = -9.38, *p* < .01. But, there was not a significant difference between the knowledge post-test or at the follow-up, HSD (3, 10) = 2.08, *p* > .01.

5. $F_{traveling} = 4.80$
   $F_{facing} = 0.53$
   $F_{AB} = 0.27$

   $d_{traveling} = |-0.48|$
   $d_{facing} = 0.17$

   The interaction between the direction participants were traveling and the direction they were facing when estimating the distance of the University of Toronto was not significant ($p > .05$). The main effect of what direction they were facing was also not significant ($p > .05$). However, the direction participants were traveling did have a significant effect on their estimates of distance. Those traveling *toward* the University of Toronto campus estimated the campus to be closer to the subway station than those traveling *away* from the campus ($p < .05$). The effect size index, d = 0.48, indicates a medium effect.

   APA format: A 2 (direction traveling: away vs. toward) x 2 (direction facing: away vs. toward) factorial ANOVA was conducted to test the effects of spatial (direction traveling) and visual (direction facing) orientation on estimates of subjective distance. There was no significant interaction between direction traveling and direction facing on estimates of subjective difference, $F(1, 76) = 0.27$, $p > .05$, nor was there a significant main effect of direction facing, $F(1, 76) = 0.53$, $p > .05$, $d = 0.17$, However, there was a significant main effect of direction traveling, $F(1, 76) = 4.79$, $p < .05$ $d = 0.48$, such that those traveling toward the campus estimated the campus to be closer ($M = 3.63$) than those traveling away from the campus ($M = 4.15$). The magnitude of this main effect was medium.

6. Reject the null hypothesis; current students do not have the same sex makeup as the 2013 workforce reported in the APA report (x^2 (1) = 7.81; $p < .05$). There are fewer men and more women in the undergraduate sample than the 2013 workforce ratio of 2.1: 1 predicts. This suggests that when the professor's students get to the workforce there will be even fewer men compared to women.

   APA format: The student's sex makeup was significantly different from the 2013 workforce reported in the APA report (x^2 (1) = 7.81; $p < .05$). There were fewer men and more women than the 2013 workforce ratio of 2.1: 1 predicted. This suggests that when the professor's students get to the workforce there will be even fewer men compared to women.

7. The top-rated European countries have significantly more favorable business conditions than the top-rated Asian countries, according to data from U. S. News and World Report, $U = 18$, p < .05.