Single-Sample and Two-Sample *t* Tests

6

The simple *t* test has a long and venerable history in psychological research. Whenever researchers want to answer simple questions about one or two **means** for **normally distributed variables** (e.g., neuroticism, daily caloric intake, height, rainfall), a *t* test will often provide the answer to such questions. In the simplest possible case, a researcher might want to determine whether a *single* mean is different from some known standard (e.g., an established population mean). For example, a researcher interested in the academic qualifications of Dartmouth students might want to see if the SAT scores of Dartmouth students are higher than the national average. If the researcher had access to a small but representative sample of Dartmouth students, she might obtain their SAT scores (let's say the mean was 1190) and conduct a statistical test to see if this mean is significantly higher than the national average (it fluctuates around 1000; let's say it's 1023). As you may recall, the formula for the single-sample *t* test is

$$t = \frac{M - \mu}{S_M},$$

where *M* is the sample mean about which you would like to draw an inference (in this case, 1190), μ is the hypothetical population mean (1023) against which you would like to compare your observed sample mean, and S_M is the **standard error of the mean.** The standard error of the mean is closely related to the estimate of the population **standard deviation,** but it also takes the size of your sample into consideration. More specifically, it's

$$S_M = \frac{S_x}{\sqrt{N}}.$$

As your sample size gets *larger,* your estimate of the standard error of the mean gets *smaller,* and this is why single-sample t tests (like almost all other statistics) become more powerful (i.e., more likely to detect a true effect, if there is one) as sample size increases. As a concrete example, suppose that you only had access to a measly sample of four Dartmouth students, whose mean SAT score was 1190 (as noted previously). Furthermore, suppose you didn't know that the population standard deviation for the SAT is 100 points—and thus you had to estimate the standard error of the mean for the SAT by first estimating the population standard deviation for the SAT (based on your sample of four randomly sampled Dartmouth students). Let's say this estimate of the standard deviation turned out to be 114 points. According to the formula for a single-sample t test, the t value corresponding to your mean of 1190 would be

$$\frac{1190-1023}{57},$$

which turns out to be 167/57, which is 2.93.

The critical value of t with 3 df ($\alpha = .05$) is 3.18. Thus, based on this single-sample t test, we would *not* be able to conclude that the mean SAT score of Dartmouth students is significantly greater than the national average. However, if we obtained the same mean and the same estimate of the standard deviation in a random sample of nine rather than four Dartmouth students, our new estimate for the standard error of the mean would be 38 rather than 57 (you should be able to confirm why this is so with some very quick calculations). The t value corresponding to this new test would be 4.39, and the new critical value for t with 8 df is 2.31. Even if we set alpha at the more stringent .01 level, the new critical value would be 3.36, and we'd be able to conclude (quite correctly, of course) that Dartmouth students have above-average SAT scores. Finally, if we had a random sample of 100 Dartmouth students at our disposal, and we observed the same values for the mean and the standard deviation, the resulting value for a single-sample t test would be 14.65, a very large value indeed. We would have virtually no doubt that Dartmouth students have above-average SAT scores.

So the single-sample t test can answer some important questions when you have sampled a single mean. To be blunt, however, many questions of this type are pretty darn boring. Most interesting psychological questions usually have to do with at least *two* sample means. In an experiment, these two means might be an experimental group and a control group. In a survey study of gender differences, these two means might represent men's and women's problem-solving strategies or the strength of their preferences for the initials in their first names (by the way, women like their first initials even more than men do).

Back in the good old days, a *lot* of psychological studies could be analyzed using a two-sample t test. That is, many studies focused on (a) a single

independent variable and (b) a normally distributed, or quasi-normally distributed, dependent variable. For example, a classic social psychological study by Hastorf and Cantril (1954) examined people's reactions to a bitterly contested football game between Princeton and Dartmouth (back in the days when these schools had highly respectable teams). The only independent variable in this classic study was whether the observers of the game were enrolled at Princeton or at Dartmouth. One of the main dependent variables was students' estimates of the total number of penalties Dartmouth committed during the game (the researchers also looked at estimates of the number of penalties Princeton committed, but we'll focus on Dartmouth because these results were the most dramatic). Because you live in the era of controversial murder trials, controversial NFL playoff games, and controversial presidential election counts, you probably won't be too surprised to learn that the Princeton students and Dartmouth students strongly disagreed about just how many penalties Dartmouth had committed—*even after reviewing a film of the game.* However, back in the 1950s, many people subscribed to the view that there is a single reality out there that most honest, self-respecting people can easily detect. How hard is it, really, to keep track of blatant infractions *on film* during a football game? Hastorf and Cantril documented that it could be quite hard indeed. Whereas the average Dartmouth student estimated that Dartmouth committed 4.3 infractions during a sample film clip, the average Princeton student put the total number of infractions at 9.8 after observing the same film clip. A two-sample *t* test confirmed that these different estimates were, in fact, different. This *t* test thus indicated that, even after viewing the same segments of game film, the student bodies of Princeton and Dartmouth strongly disagreed about the number of infractions committed by Dartmouth.

In this chapter, we are going to examine the *t* test and explore some of its most common uses. We begin by examining a simple data set that requires us to conduct a one-sample *t* test. Because the single-sample *t* test is conceptually similar to certain kinds of χ^2 tests, we will compare and contrast the single-sample *t* test with a χ^2 analysis based on the same simple study. Next, you will examine some simulated data from a classic field experiment using both an independent samples *t* test and a two-groups χ^2 test. After warming up on these two activities, you will use *t* tests to examine three other data sets, including a name-letter liking study, an archival study of heat and violence, and a blind cola taste test. Let's begin with the data involving inferences about a single score or mean.

Bending the Rules About Happiness

A cognitive psychologist was interested in the function rule that relates happiness to the objective favorability of life outcomes. Because of the preliminary nature of her research, she wanted to begin by testing her most

basic assumption about the perception of happiness. In particular, the researcher assumed that as the *objective* favorability of a positive outcome increases in a linear fashion, people's *subjective,* psychological responses to the outcome will *not* increase in a linear fashion. Instead, she felt that subjective happiness will follow what might be termed a *law of diminishing returns.* As events become objectively more favorable in a linear fashion, people will only experience them as more favorable in a diminishing, curvilinear fashion. For example, based on past research, this researcher knew that if we wish to double the *perceived* intensity of a light presented in an otherwise dark room, we will have to increase its physical (i.e., objective) intensity by a factor of about 8. To be perceived as twice as bright as Light A, Light B must emit about eight times the amount of light energy as Light A (e.g., see Billock & Tsou, 2011; Stevens, 1961). Does this simple rule of diminishing returns apply to emotional judgments such as happiness? As a first test of this idea, the researcher asked people to think about how happy they would be if they were given 10 dollars. She then asked them to report exactly how much money it would take to make them exactly twice this happy (see Thaler, 1985). Her reasoning was simple. If the law of diminishing returns applies to happiness, it will take *more* than $20 to make the average person twice as happy as he or she would be if he or she were to receive $10. Image 1 depicts an SPSS data screen containing data from 20 hypothetical participants who took part in this simple study. *You should convert the screen capture in Image 1 to an SPSS data file of your own.*
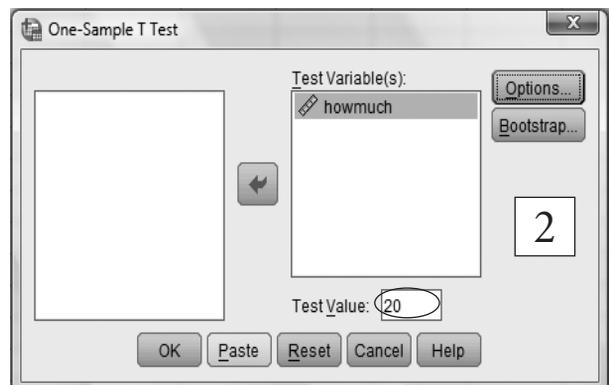
After making sure the scores in your data file look exactly like those in Image 1, click "Analyze" and then go to "Compare Means" and "One-Sample T-Test. . . ." Clicking on this analysis button will open a dialog box like the one you see in Image 2. Send your only variable ("howmuch") to

the Test Variable(s) list, and enter 20 as your Test Value (circled in Image 2). Now click the "Paste" button to send this one-sample *t* test command to a newly created syntax file. Your syntax file (especially the text) should look very much like the one you see in Image 3.


```
DATASET ACTIVATE DataSet0.
T-TEST
  /TESTVAL=20
  /MISSING=ANALYSIS
  /VARIABLES=howmuch
  /CRITERIA=CI(.95).
```

⇨ **T-Test**

[DataSet0]

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
howmuch	20	43.7500	26.18733	5.85567

One-Sample Test

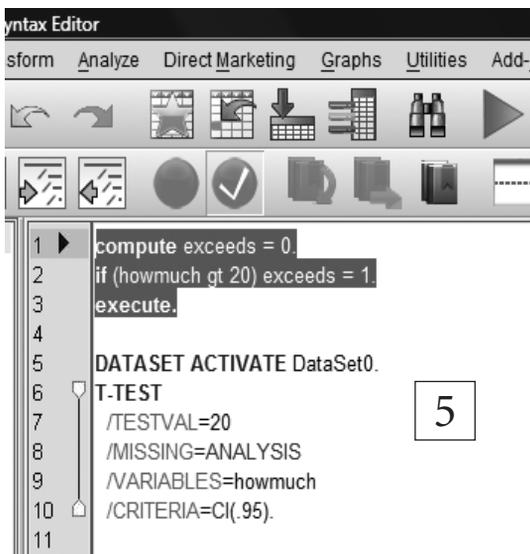|	Test Value = 20					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
howmuch	4.056	19	.001	23.75000	11.4940	36.0060

Running this one-sample *t* test command (by clicking the "Run" button) should produce an output file that looks a lot like the one you see in Image 4. We will leave it up to you to decide exactly what each piece of information in this output file tells you about the results of the researcher's study.

QUESTION 6.1a. Summarize the results of this study as if you had conducted the study yourself. Make sure to include everything you would want readers to know about your results. First, briefly explain why you set 20 (vs. 0, 100, etc.) as your comparison standard. Defend this choice against the (incorrect) criticism that one-sample means are supposed to be compared with some kind of *population value* (e.g., population SAT scores). Second, be sure to describe your findings relative to this comparison standard (e.g., "Dartmouth students exceeded the population mean on the SAT by more than 30%."). In other words, what, exactly, did you find? Third, was this finding statistically significant? Report your observed *t* value the way you would report it in a research paper. In case you are not familiar with the format for reporting the results of a one-sample *t* test, some examples appear in Appendix 6.1.

QUESTION 6.1b. Just to keep you on your toes, work from the information given in your printout and check to be sure that SPSS (a) correctly converted the standard deviation to the standard error of the mean and (b) correctly calculated the value of the *t* test itself. Make sure to show your work for these hand calculations. (Do not start from scratch. Instead, merely use the values for the sample mean and standard deviation that you are given in your SPSS output file.)

Simplifying the Outcome

Imagine that you showed your data to a collaborator who insisted that the most important question in your study is not the exact value of people's monetary self-reports but whether these reports *exceed* the hypothetical standard of $20. In short, the collaborator wants to know what happens if you simplify people's responses by coding them for whether they exceed the $20 standard. The syntax file depicted in Image 5 provides an example of how you might perform this recode. Be sure to highlight the entire set of new syntax commands (including "execute," as we did in Image 5) before you click the "Run" arrow. After you have run this "compute" command, your revised SPSS data file should look something like the one depicted in Image 6.


Now you want to analyze this new "exceeds" variable. Because this new variable is dichotomous (0 or 1), Karl Pearson would be quick to remind you that you can run a simple χ^2 test to see if more than half of your participants exceeded this critical value. After all, if people's scores just represent random noise around a true mean of 20, then half of the scores should exceed 20—just as about half of all coin tosses land tails up and about half of all z scores exceed the hypothetical mean of zero. In Images 7 through 10, you'll get a peek at how to do this. By the way, the way you select and modify nonparametric statistical tests changed pretty radically in Version 20 of SPSS, so if you have an older version of SPSS, you may wish to consult Appendix 6.2 for a couple of screenshots based on an earlier version of SPSS. However you do it, though, you'll want to conduct a simple chi-square test in which you compare your observed frequency for "exceeds" with the frequency you'd expect if "exceeds" had a probability of .50 in the population.

As you can see in Image 7, you begin a one-sample chi-square test by selecting "Analyze," "Nonparametric Tests," and then "One Sample. . . ." As

shown in Image 8, after you click "One Sample. . . " SPSS will open a dialog box that has an Objective tab, a Fields tab, and a Settings tab (circled near the top of Image 8). You should also notice that the default objective (also circled in Image 8) is to "automatically compare observed data to hypothesized." This is great because this default is exactly what you want. From here, you *could* select run, but SPSS would assume that you want to analyze every single variable in your data file! This is rarely what people want to do. So let's *not* do that. Instead, let's focus on the newly created "exceeds" variable.
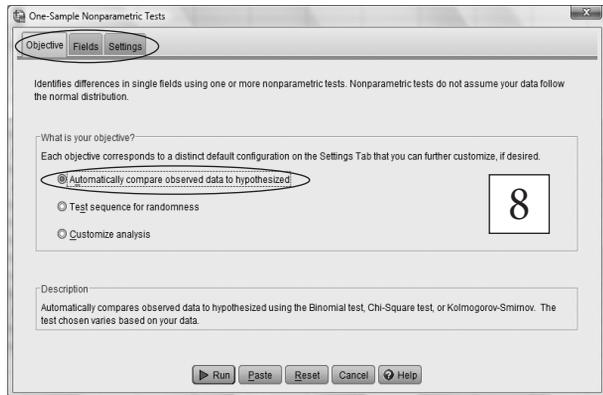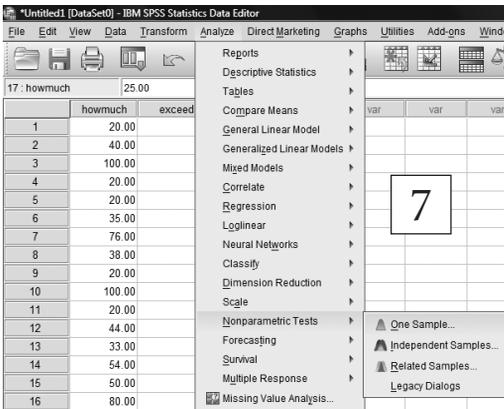
To do this, click the Fields button (the second of the three circled buttons at the top of Image 8), and this will take you to the screen you see in Images 9 and 10. Notice in Image 9 that the default is to include both of your variables in the "Test Fields." Of course, you don't want to analyze the original "howmuch" variable, and if you click on this variable as we did in Image 9, you can use the arrow button (circled for you in Image 9) to send it back from "Test Fields:" to "Fields:" (see Image 10)—where SPSS will now politely ignore it.
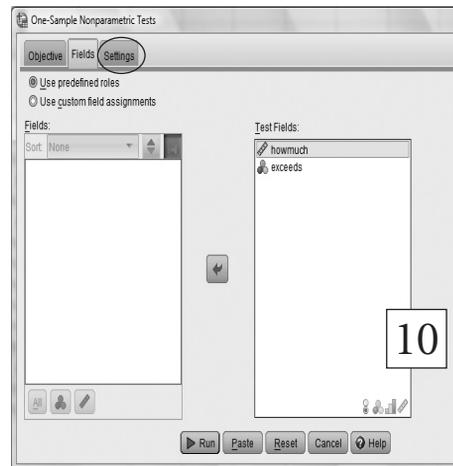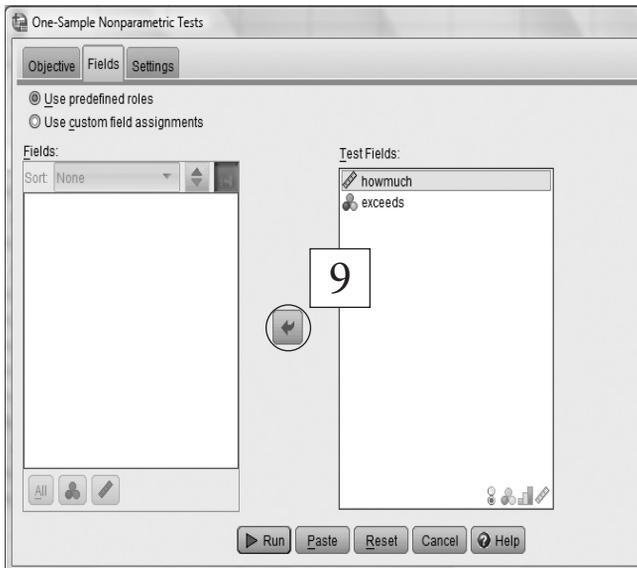
Unfortunately, you still are not quite ready to run your analysis. You'll now want to click the "Settings" button (circled in Image 10) to open the "One-Sample Nonparametric Tests" dialog window, which you can see in Image 11.
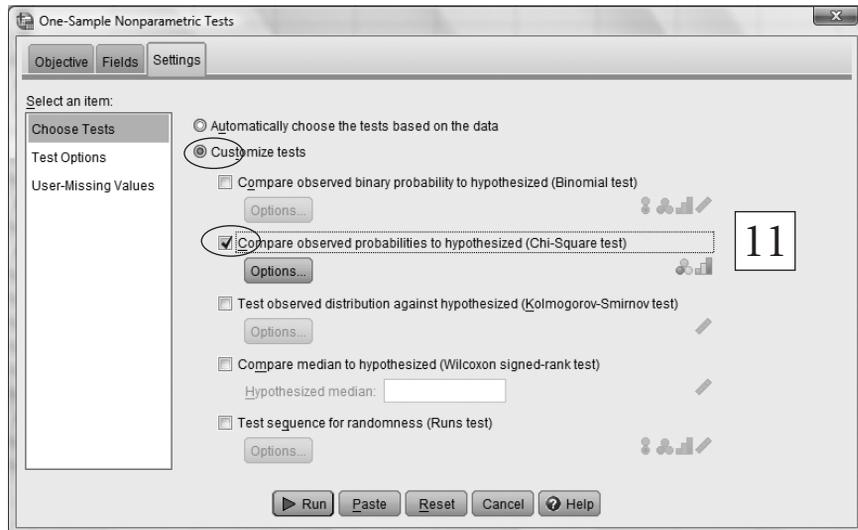
Instead of accepting the really interesting default, which is to let SPSS choose what test to run based on the properties of the "exceeds" variable that you chose in the last ("Fields") dialog box, click "Customize tests" (circled near the top of Image 11). This will allow you to decide, for example, whether to run a binomial test or a chi-square test. To keep things simple, let's only ask for the chi-square test, which you can see we requested by checking the appropriate box in Image 11. If you now click "Paste," you should send the appropriate chi-square analysis command to your syntax file, and the crucial portion of that syntax file should look a lot like the screen capture you see in Image 12.
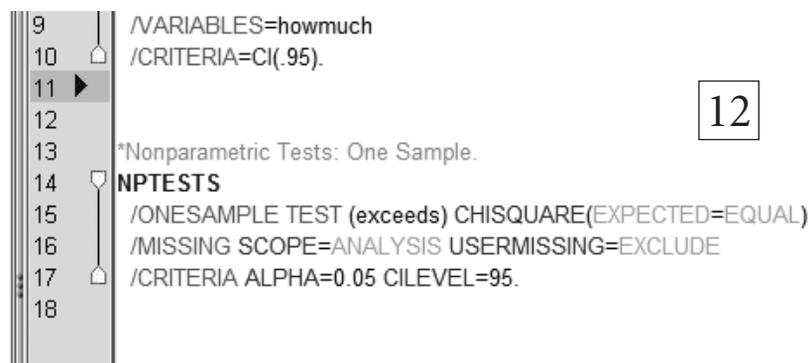
Running the appropriate portion of this syntax command should produce the portion of the SPSS output file you see in Image 13.
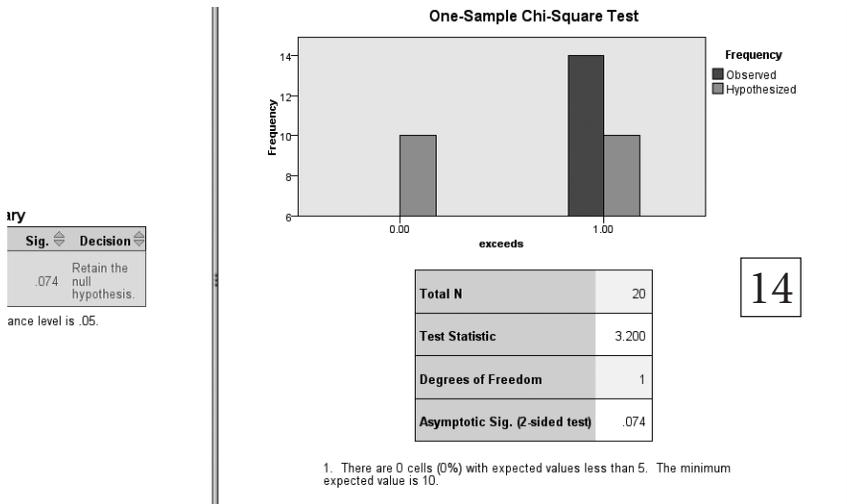
Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories of exceeds occur with equal probabilities.	One-Sample Chi-Square Test	.074	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

From here you can probably already see that the results of this test are *not* statistically significant (*p* = .074). However, it would also be nice to know the exact value of the chi-square statistic (rather than merely knowing its *p* value). To see the actual chi-square value associated with this *p* value, just double click anywhere inside the "Hypothesis Test Summary" box you see in Image 13. This will open up the more detailed part of your output file you see selectively captured in Image 14.

Total N	20
Test Statistic	3.200
Degrees of Freedom	1
Asymptotic Sig. (2-sided test)	.074

1. There are 0 cells (0%) with expected values less than 5. The minimum expected value is 10.

You now have all the information you need to make a complete report of your findings using this alternate measure.

QUESTION 6.2a. Compare the results of your new χ^2 analysis with your original findings based on the one-sample *t* test. We'll leave it up to you to figure out the output file and to draw your own conclusions about the pros and cons of this alternate analysis.

But we can't resists making one helpful suggestion. To figure out if any differences between your new (χ^2) results and your old (*t* test) results are

due to the nature of the *statistical tests* themselves (versus being due to the way you *recoded the data*), we suggest that you copy the entire *t* test syntax command that you initially conducted and paste it below the χ^2 command. Now edit the copied *t* test command by hand in two ways. First, replace "howmuch" with "exceeds." Then replace "20" in the TESTVAL = line with "0.50." This will mean that the new *t* test analyzes the "exceeds" variable rather than "howmuch" and compares it with the standard of 0.50 rather than the old standard of 20. Here is how that new command should read

DATASET ACTIVATE DataSet0.

T-TEST

/TESTVAL=0.50

/MISSING=ANALYSIS

/VARIABLES=exceeds

/CRITERIA=CI(.95).

Running the edited *t* test command will generate a new section in your SPSS output file.

QUESTION 6.2b. Compare the results of the χ^2 test and the *t* test when you perform the two tests on *the same scores*. Based on this additional comparison, what is the best way of coding your data in this study (continuously or dichotomously)? Be sure to defend your answer.

Incidentally, you may be having a kneejerk reaction that says it's a big mistake to analyze dichotomous data using a *t* test. Like most other jerks, this particular jerk is wrong. Remember that Question 6.2b is more of a diagnostic exercise in seeing how statistics work than it is a recommended way of analyzing dichotomous data. If that isn't good enough for you, we'd like to reassure you that many parametric tests (e.g., *t, z, F*) are more robust than most people think to the assumption that dependent variables are normally distributed, especially when the nonnormal distribution is platykurtic (i.e., flat). At any rate, it certainly is *not* the case, for example, that merely switching from a parametric statistic to a nonparametric statistic when your data are not normally distributed will solve all of your distributional dilemmas (e.g., see Harwell, Rubinstein, Hayes, & Corley, 1992; Zimmerman, 1987; Zumbo & Zimmerman, 1993). Nonetheless, if you wish to avoid painful and unnecessary debates with reviewers and editors when trying to publish your own research with nominal variables, we recommend—for purely pragmatic reasons—that you lean toward the default of using nonparametric tests such as χ^2.

Before you move on to the next data set, make sure that you save both your data file and your syntax file using the same descriptive name stems. We recommend "perceived happiness study (one-sample t-test)." If you

haven't done so already, you should also give your variables very clear labels in your data file before you save everything one more time.

## The Independent Samples (Two-Samples) *t* Test

One of the most widely cited studies in social and educational psychology is Rosenthal and Jacobson's (1966) classic study of how teachers' expectations can influence children's academic performance. Like a lot of other classic studies, this study involved a very simple manipulation (with only two levels), and thus, it qualifies as a good illustration of the two-sample *t* test. Rosenthal and Jacobson went into an elementary school and received permission to give all of the children in the school a new intelligence test. The test was a recently developed test of nonverbal intelligence that none of the students or teachers in the school would have been likely to have ever seen before. Thus, Rosenthal and Jacobson were able to create teacher expectancies regarding the test without worrying much about any naturally existing expectancies the teachers may have already had. In particular, they informed the teachers (erroneously, of course) that the new test was a "test for intellectual blooming." They further explained to the teachers that children who perform well on this test typically experience what might be thought of as an academic growth spurt. That is, the researchers led teachers to believe that students who did well on this test possessed a special kind of intelligence—one that might normally go undetected in the classroom. And then they informed teachers of exactly which students had achieved exceptional scores on the test. Of course, these students were *not* identified on the basis of their actual performance on the test. Instead, Rosenthal and Jacobson *randomly* chose about a dozen students from each of six large classrooms (composed of about 50 students each). In other words, they conducted a **field experiment** on teacher expectancies.

   Eight months later, the researchers returned to administer the same test again in each of the classrooms. Back in those days, people typically analyzed change by simply examining difference scores. In this case, the researchers simply developed a single score for IQ change by subtracting children's (true) initial IQ scores from their IQ scores on the same test administered 8 months later. Not surprisingly, most children showed increases in performance on the nonverbal IQ test (after all, they had completed a year of formal schooling). The question, of course, was whether the children who were randomly dubbed intellectual bloomers would show greater increases in IQ than the remaining children (who had not been labeled as bloomers). Your data for this part of the assignment are a *very close approximation* of the Rosenthal and Jacobson (1966) data for first-grade children only (if you want to see their findings for children in other grades, see the original article). The SPSS data file is called [**simulated teacher expectancy data (2-sample t-test).sav**]. A small portion of the data file appears in Image 15.

Before you go any further, you should check out the variable labels for these data. Doing this will show you, for example, that for the independent variable "bloomer," scores of 0 refer to kids in the control group and scores of 1 refer to kids randomly labeled "bloomers." You can also see a reminder that "iqchange" is the difference between children's nonverbal IQ scores at Time 2 and these same IQ scores at Time 1 (it's Time 2 minus Time 1 IQ, so more positive scores indicate greater increases in IQ).

Because you are probably becoming quite adept now at selecting statistical analyses, we offer you only a couple of clues in Images 16 through 18 about how to select and run an independent samples (two-samples) *t* test. However, one novel aspect of the two-sample *t* test window is that SPSS requires you to tell it exactly how you defined (coded) the two levels of your independent variable. In this case, you probably recall that the codes for the control and "bloomer" groups, respectively, were 0 and 1. So after defining bloomer as the "Grouping Variable" and "iqchange" as the "Test Variable," as we did in Image 17, click on "Define Groups. . ." from the box you see in Image 17, and then enter these values of 0 and 1 in the "Define Groups" box you see in Image 18.
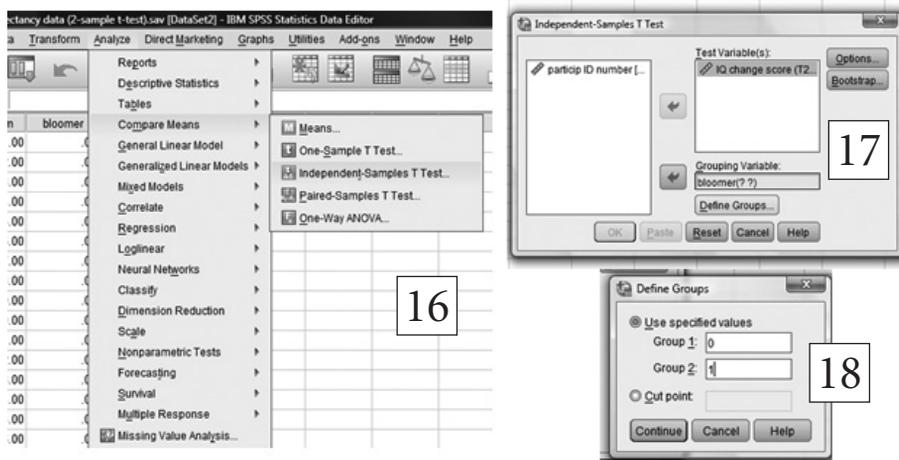
The "Continue" button will now take you back to the main analysis box (Image 17), and if you click "Paste" from this box, you should generate the specific syntax command that appears below:

DATASET ACTIVATE DataSet2.

T-TEST GROUPS = bloomer(0 1)

/MISSING = ANALYSIS

/VARIABLES = iqchange

/CRITERIA = CI(.95).

Running this command in the usual way will produce the results of this independent samples *t* test.

Results of the Teacher Expectancy Study

The part of your output file that matters most should strongly resemble what you see in Image 19.

This output file will tell you everything you need to know about one set of findings from this classic study.

Group Statistics

	was child labeled bloomer?	N	Mean	Std. Deviation	Std. Error Mean
IQ change score (T2 - T1)	no	19	12.0000	16.39444	3.76114
	yes	11	27.3636	12.57197	3.79059

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Differe
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	
IQ change score (T2 - T1)	Equal variances assumed	.744	.396	-2.678	28	.012	-15.36364	5.73623	-27.11378	
	Equal variances not assumed			-2.877	25.599	.008	-15.36364	5.33992	-26.34839	

QUESTION 6.3. Summarize the results of this study. Include a table of your findings and summarize your results in the same way you might in a very short journal article or research talk. Make sure that you include a brief discussion of the meaning and significance (social, educational, etc.) of these findings.

More Simplification

The same colleague with whom you collaborated on the happiness study argues that, to simplify this study, you should simply code for whether the students showed an increase (of any size) in IQ over the course of the

academic year. Based on what you recently learned about the syntax commands for computing and recoding data (in the happiness activity), write some coding statements that will create a new variable called "gain." Code gain as 1 if students showed any increase at all in IQ, no matter how small. Code it as 0 if a student failed to show such an increase (including a loss or no change at all). Now using the skills you have developed in this text, conduct both a χ^2 test and a new independent samples t test (on the same simplified "gain" variable) to see if your results are any different when you use this much simpler way of coding and analyzing the data. You might be tempted to conduct a χ^2 test using the same set of commands that you just used for the one-sample t test. You cannot easily do this, however, because this specific test is designed for the χ^2 equivalent of one-sample t tests. The χ^2 test you want to run now is the kind of χ^2 contingency test you conducted a couple of times in Chapter 4 (the tests having to do with name-letter preferences in relation to choices about colleges and marriage). As you may recall, the easiest way to run this χ^2 test is to start at "Analyze," then to go to "Descriptive Frequencies" and then "Crosstabs." You can always return to Chapter 4 if you want more detailed instructions.

QUESTION 6.4. To get back to the point of this exercise, summarize the results of both the χ^2 test and the t test for the "gain" variable. Be sure to include a brief statement about which coding scheme (original or simplified) you think is better for this study.

   If you haven't already done so, remember to save your syntax file, ideally giving it the same prefix that your data file has (your data file should be OK the way it was when you opened it; so there is no need to save it). There is really no need to save your output file, by the way, unless your instructor asks you to submit it as proof that you ran the analyses that will support your answers.

Yet Another Name-Letter Preference Study

The famous Chilean social psychologist Mauricio Carvallo (he says he's *particularly* famous in Santiago but refuses to say why) once convinced the first author of this text to collaborate on a name-letter study (i.e., a study of implicit egotism) in which we sampled every person named Cal or Tex who lived in either California or Texas. (His most persuasive argument for doing the study, by the way, is that he would do all the hard work.) Mauricio thus sampled about 900 men from an Internet email directory, asking them to report how much they liked each of the 26 letters of the alphabet on a 9-point scale (where higher numbers indicate higher levels of letter liking). In addition to replicating the finding that people were disproportionately likely to live in states that resembled their names, we wanted to see whether men who lived in states that resembled their names would like their initials more than men who lived in states that did *not*

resemble their names. The idea would be to gather some direct evidence that strength of name-letter preferences plays a role in the strength of behavioral preferences based on name letters.

Here are some hypothetical data that very strongly resemble our actual data. The SPSS file containing these data is called [**Cal Tex name letter liking study.sav**].

Analyze these data to see if we were correct that relative to people living in states that do not resemble their first names, people who live in states that do resemble their first names have stronger preferences for their first initials. Here's the tricky part: Find a way to do this that allows you to analyze a *single score* for each participant, and code the score so that higher numbers reflect higher levels of name-letter liking. Although there are more complex ways to analyze the data, you can and should be able to conduct a simple, two-groups *t* test when you complete all the recoding for this activity. Table 6.1 gives you some ideas about how to think about and analyze these data. As shown in Table 6.1, each participant (all of them named Cal or Tex) has two important scores: liking for the letter *C* followed by liking for the letter *T*.

Table 6.1   Liking for the Letters *C* and *T* Among People Named Cal Versus Tex Who Live in California Versus Texas

	FIRST NAME	
STATE	**Cal**	**Tex**
California	8 4, 9 5, 9 8, 3 4, 9 9, 3 4, 8 5, 7 8, 7 4, 9 8, 4 4, 6 5, 8 5, 6 5	5 6, 7 9, 9 9, 7 4, 8 9, 9 9
Texas	7 8, 8 8, 3 4, 9 7	5 7, 7 7, 9 9, 3 7, 5 8, 8 7, 3 4, 9 9, 5 8, 8 9, 3 4, 5 6, 7 7, 5 5, 6 8, 6 7

Here are some further clues: First, you'll need to create a new variable that reflects whether men do or do not live in a state that resembles their own first name. For this purpose, a Cal living in California is no different from a Tex living in Texas. But they are both different from Cals living in Texas and Texes living in California (who belong in the same group). Second, you will also need to create a new variable that reflects the strength of people's liking for their first initials. To do so, you need to take a difference score (e.g., liking for *C* minus liking for *T*), but this difference score will be calculated differently for people with different names. Because we have not asked you to do a lot of independent syntax programming in this text, we include a brief primer on SPSS COMPUTE statements and IF statements in Appendix 6.2. This will probably come in very handy for this data coding and data analysis activity.

QUESTION 6.5. Report what you did and what you found. Be sure to include (a) exactly how you coded your data to allow for a test of our hypothesis. This should include a copy of the part of your syntax file that makes the crucial calculations (i.e., that creates the two new variables). In addition, (b) discuss the *results* of your *t* test, including a discussion of whether these findings support the basic assumptions behind the idea of implicit egotism. Finally, (c) be sure to discuss any critiques or limitations of your findings. If it seems relevant, discuss the issue of one-tailed versus two-tailed hypothesis testing.

An Archival Study of Heat and Aggression

A researcher interested in aggression identified the 10 hottest and coldest U.S. cities and assessed murder rates in each city. His hypothesis was that because heat facilitates aggression, the hottest cities would have higher murder rates than the coldest cities. Each city name below is followed by the number of murders per 100,000 people committed in that city in 2006. Conduct a *t* test with city as the unit of analysis ($n = 20$) to see if there is support for the researcher's hypothesis. (You'll need to create your own SPSS data file for these real data.)

Hot Cities	Murder Rate	Cold Cities	Murder Rate
1. Key West, Florida	4.1	1. International Falls, Minnesota	0.0
2. Miami, Florida	19.6	2. Duluth, Minnesota	1.2
3. W. Palm Beach, Florida	17.1	3. Caribou, Maine	0.0
4. Ft. Myers, Florida	25.2	4. Marquette, Michigan	0.0
5. Yuma, Arizona	3.4	5. Sault Ste Marie, Michigan	0.0
6. Brownsville, Texas	2.9	6. Fargo, North Dakota	2.2
7. Orlando, Florida	22.6	7. Williston, North Dakota	0.0
8. Vero Beach, Florida	0.0	8. Alamosa, Colorado	0.0
9. Corpus Christi, Texas	7.2	9. Bismarck, North Dakota	1.7
10. Tampa, Florida	7.5	10. St. Cloud, Minnesota	0.0

Note: Murder rate data were taken from http://fargond.areaconnect.com/crime/compare .htm. The hottest and coldest U.S. city list can be found at http://web2.airmail.net/danb1/ usrecords.htm.

QUESTION 6.6. Summarize the results of this study. Did it support the researcher's hypothesis? Be sure to point out any methodological concerns you have about the study. A good answer will use the word *confound*. A good source that may address some of the confounds that are relevant to this research is Anderson (2001).
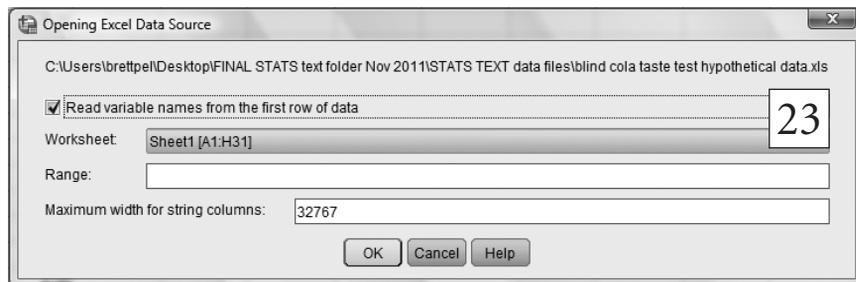
A Blind Cola Taste Test

During about 20 years of teaching research methods, the first author of this text has frequently conducted blind taste tests in which he served people different popular colas and allowed the taste testers to see whether they can tell the different colas apart. People in such taste tests are always told, truthfully, that they will get two or three different colas (e.g., Coke, Pepsi, and a generic cola) without being told which cola is which. I always use sequentially numbered cups, and I serve the colas to participants in different random (counterbalanced) orders. I've now done this blind taste test so many times that I don't usually bother to save the data anymore, and so I am going to give you some hypothetical data that reflect the typical findings when I serve three colas (more often I serve only two). Fortunately for you, I entered these data into an Excel file rather than an SPSS data file, and so you get to learn a new skill (reading an Excel file into SPSS). A screen capture from the Excel file appears in Image 20.

Microsoft Excel - blind cola taste test hypothetical data.xls

idnum	gender	predcoke	predpepsi	predgeneric	gotcoke	gotpepsi	gotgeneric
1	1	0.75	0.75	0.75	0	1	0
2	2	0.75	0.75	0.75	1	1	1
3	1	0.8	0.9	0.6	1	1	1
4	1	0.55	0.33	0.22	1	0	0
5	2	0.35	0.33	0.22	1	0	0
6	1	0.33	0.48	0.4	1	0	0
7	1	0.99	0.99	1	1	0	0
8	1	1	1	0.8	0	0	0
9	2	0.6	0.7	0.6	1	1	1
10	2	0.58	0.58	0.44	1	1	1
11	2	0.66	0.88	0.5	0	0	1
12	1	0.33	0.33	0.33	0	0	0
13	1	0.33	0.33	0.34	0	0	1
14	2	0.75	0.9	0.8	1	1	1
15	2	0.33	0.33	0.33	0	0	0
16	1	0.68	0.5	0.5	0	1	0
17	2	0.33	0.33	0.34	0	0	0

Fortunately, SPSS reads Excel data files almost as fluently as it reads SPSS data files. Thus, to convert an Excel data file to SPSS, you do much the same thing you would do to open an SPSS data file. The screen captures for this simple series of steps follow:
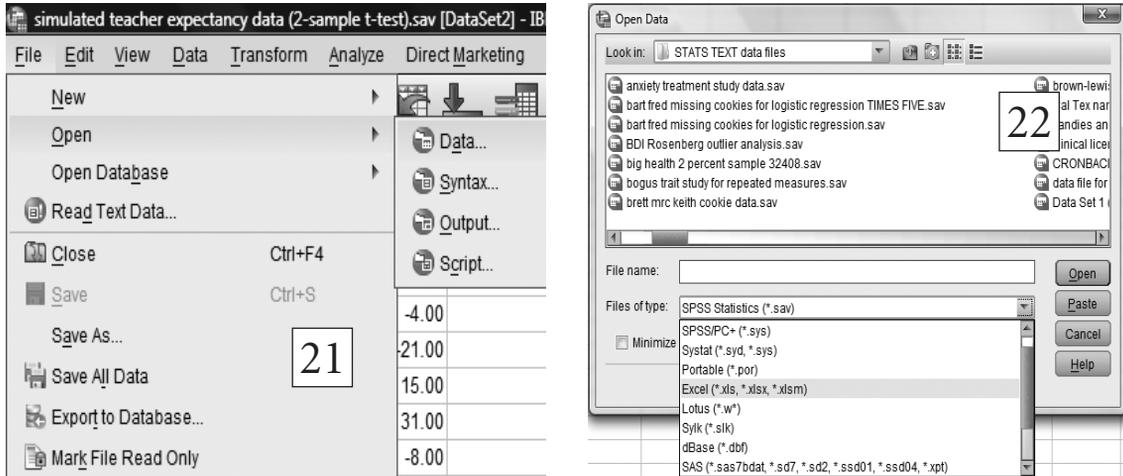
Once you've clicked on the "Data. . ." button shown in Image 21, you'll see a dialog window that resembles the one in Image 22. From this window, you'll need to let SPSS know, using the "Files of type:" box, that you want to open an Excel file rather than the default SPSS data (.sav) file. Once you've done this, just click "Open" and then accept the default (shown in Image 23), which is to allow SPSS to read the variable names from the first row of data in an Excel file. Of course, if your Excel file had no such variable names in the first row, you'd have another job to do, but luckily for you, the Excel file you are reading was created by a caring nurturer who wanted your data analysis life to be as wonderful as possible.

The Excel file for the blind cola taste test data is called [**blind cola taste test hypothetical data.xls**]. So you merely need to go to the folder where this Excel file is located and then read the data directly into SPSS. Incidentally, because this Excel file came from a caring nurturer who is

not very organized, there are no variable labels in this file. Thus, we should tell you that gender is coded 1 = M, 2 = F. Also, the three variables that begin "pred" are people's *predicted* accuracy scores (assessed before people got feedback) for Coke, Pepsi, and a generic cola. Thus, the answer .75 for "predcoke" for the first participant means that he said the chances that he had correctly labeled what the experimenter knew to be Coke was 75%. Finally, the last three variables (e.g., "gotcoke") reflect whether a person actually did label a given brand of cola correctly. Thus, the first participant labeled Pepsi correctly but mixed up the Coke and the generic cola (thus mislabeling them both). Now that you understand what all the scores indicate, you should enter all of your variable labels and save the SPSS data file you created from the original Excel file. Then you'll be ready to analyze the data so that you can answer the questions that are soon to follow. Before you do so, however, you will need to use your developing variable creation (COMPUTE) skills to create two new variables. One is the average *predicted accuracy* score for each participant. The other is the average accuracy score for each participant. If you do these calculations correctly, the first participant, who we are pretty sure was the well-known psychologist Keith Maddox, will have an average predicted accuracy score of .75 (the average of .75, .75, and .75) but will have a true accuracy score of only .33 (the average of 0, 1, and 0).

QUESTION 6.7a. Write up these results for labeling accuracy as if your data were real. Specifically, can people identify the three colas at an above chance level in a blind taste test? A tricky issue is how to define the chance standard used for purposes of statistical comparison in your one-sample *t* test. To decide on this, assume that everyone always got all three brands of cola (they always do in our three-cola tests) and that the experimenter presented each of the six possible orders of the three colas (counterbalanced across all of the participants). Now assume that a typical participant was *arbitrarily guessing* the identities of the colas and label her guesses 1, 2, and 3 (or C, G, and P, for the three brands). If you compare this random answer pattern with the list of every possible order of delivery, you should be able to see how many correct answers the average guesser would get across all of the trials. In case you are not sure what we mean by this, the six possible orders of cola delivery are indicated for you in Table 6.2. All you need to do to figure out the average accuracy score produced by chance guessing is to always assume that any one constant set of guesses (e.g., P, G, C) is used in all six scenarios (for all six orders). Just make sure to use the *same* constant guess pattern for all six orders. Incidentally, to prove that the arbitrary guess order you choose doesn't matter, try the exercise with more than one order (e.g., first with C, G, P and then with P, G, C). Once you have converted this average to a proportion, you will know what value to enter as your comparison standard in the one-sample *t* test.

Table 6.2 Six Possible Orders of Cola Delivery in a Fully Counter-balanced Blind Cola Taste Test

Actual Cola Delivery Order	Number of Correct Answers
C, G, P	_____
C, P, G	_____
G, C, P	_____
G, P, C	_____
P, C, G	_____
P, G, C	_____
Average:	_____

QUESTION 6.7b. Do people have an accurate appreciation of their ability to do this task? That is, is the degree of *confidence* the average person reports for the cola-labeling task appropriate for the average level of *accuracy* people show during the blind taste test? A clue: Check out research on "overconfidence" by Fischhoff, Slovic, and Lichtenstein (1977). Another clue: To really answer this question thoroughly, you should compare the average confidence score not only with chance performance (in one analysis) but also with actual accuracy (in a different analysis). Your standards of comparison then (against which you should compare confidence rates) in these one-sample *t* tests are (a) chance performance in one case and (b) observed accuracy in the other case (whatever it proves to be).

QUESTION 6.7c. Are men and women equally accurate? Are they equally confident? Answering this will require two separate *t* tests.

Appendix 6.1: Reporting the Results
of One-Sample and Two-Sample *t* Tests

This appendix contains the results of some studies, both hypothetical and real, that make use of *t* tests.

> The results of this marketing study showed that consumer liking for the card game *PRIME* (*M* = 6.3 on the 7-point Coopersmith liking scale) significantly exceeded the established population estimate for *UNO* (*M* = 5.1), *t*(438) = 138.43, *p* < .001.

The value in parentheses (438) right after the *t* is the degrees of freedom (*df*) for the *t* test. For a one-sample *t* test, this value is always *n* − 1. So the above analysis tells you that these marketing researchers sampled 439 consumers. It is traditional to report actual *t* values to two decimals and to report *p* values to three decimals. So even though a *t* value this large would certainly have a *p* value with many leading zeros after the decimal point, it would be sufficient just to say *p* < .001.

Here are two examples of how you might report a *nonsignificant* result:

> The results of a one-sample *t* test did *not* support our prediction that sexual attraction to the middle-aged experimental social psychologists would exceed the negative scale endpoint of 1.0. Recall that on this 9-point scale, a score of 1 corresponds to *I'm about as attracted to the target as I am to a pile of smelly gym socks.* The sample mean was only *M* = 1.2, which did not significantly exceed 1.0, *t*(19) = 1.29, *p* = .214. In fact, only 2 of the 20 supermodels studied failed to choose 1 as their answer, and all 20 of the attraction scores were below the scale midpoint of 5 (*I'm about as attracted to the target as I am to my kind but unattractive cousin*).

> There were only six U.S. presidents for whom we were able to obtain WAIS-R intelligence test scores. The average IQ score for these six presidents (*M* = 116.3) was descriptively higher than the U.S. population mean of 100. Perhaps due to the very small sample size, however, this value did not reach conventional levels of significance, *t*(5) = 2.16, *p* = .083.

Here is an example of how to report a result for a **two-sample (i.e., independent samples) *t* test.** Unlike the hypothetical examples above, this real example comes from an early draft of a research paper by Shimizu, Pelham, and Sperry (2012). The manipulation in the study was whether participants who tried to take a nap were subliminally primed with sleep-related or neutral words immediately prior to the nap session. That is, prior to trying to nap, some participants were *unknowingly* exposed (for just a few milliseconds per word) to words such as *relaxed, restful,* and

slumber. In contrast, others were subliminally exposed to neutral words such as *bird, chair,* and *paper.* (The means, incidentally, represent minutes of self-reported sleep.)

> For self-reported sleep duration, the effect of prime condition was significant, $t(71) = 2.17$, $p = .03$. Participants primed with sleep-related words reported having slept about 47% longer ($M = 9.11$, $SD = 6.38$) than did those primed with neutral words ($M = 6.18$, $SD = 5.14$).

Because this was a simple two-groups experiment, analyzed by means of a two-sample t test, the *df* value of 71 you see in parentheses above tells you that there were 73 participants. In the case of a two-samples t test, $df = n - 2$ rather than $n - 1$.

Appendix 6.2: Some Useful SPSS Syntax Statements and Logical Operands

There are many times in data analysis when researchers need to modify a variable or variables from an SPSS data file. For example, in Chapter 5, you computed a single self-esteem score from the 10 separate self-esteem questions that make up the Rosenberg Self-Esteem Scale. At other times, a researcher might want to break a continuous variable (such as age) into meaningful categories rather than dealing with 80 or 90 specific scores. The two most useful SPSS syntax commands for this purpose are the COMPUTE statement and the IF (. . .) statement, each of which we briefly review here, along with some very useful logical or mathematical operators.

COMPUTE: The COMPUTE command creates a new variable in the active SPSS data file according to whatever logical or arithmetic rules follow the command. The new variable name always follows the COMPUTE command.

COMPUTE quiz_1 = q1+q2+q3+q4+q5.

EXE.

For every case with complete data on q1 to q5 (items on a quiz), this syntax command will produce a new variable called quiz_1 and add it to the SPSS data file.

COMPUTE quizp_1 = (q1+q2+q3+q4+q5)/5.

EXE.

This researcher decided to divide by 5 to calculate the **proportion** of correct quiz responses rather than the number of correct responses. Thus, you can see that mathematical commands can be used with COMPUTE.

IF (. . .): The IF (. . .) command creates a new variable based on whether some already existing variable (or variables) satisfies some condition or conditions. For example, you can use the IF statement to create age groups from a continuous "AGE" score. The syntax statement series that follows would create a new variable consisting of four age groups. If you highlighted and ran this entire statement set, the new variable would become the last variable in your SPSS data file.

IF (AGE LE 34) age_group = 1.

IF (AGE GE 35 and AGE LE 49) age_group = 2.

IF (AGE GE 50 and AGE LE 64) age_group = 3.

IF (AGE GE 65) age_group = 4.

EXE.

Notice that this IF set uses "LE" and "GE" ("le" and "ge" would also be acceptable) to indicate "is less than or equal to" and "is greater than or equal to," respectively. SPSS also recognizes other logical operands. Here is a list of some very useful ones along with their meanings:

"**LT**"   "is less than"

"**GT**"   "is greater than"

"**EQ**"   "is equal to"

"**NE**"   "is not equal to"

"**AND**" "and"

"**OR**"   "or"

Thus, the following syntax command

IF (GENDER EQ 2 AND NUM_KIDS GE 1) MOTHER = 1.

IF (GENDER EQ 2 AND NUM_KIDS EQ 0) MOTHER = 0.

EXE.

would only create "MOTHER" scores for women (assuming 2 means women for the gender variable). Furthermore, women with kids (regardless of how many) would be coded as mothers (1) and those without kids would be coded as nonmothers (0). Men would not receive a score at all for this new variable.

As a very different example, imagine that a researcher wanted to identify people who have been victims of any of three common crimes in the past 6 months. She might use the following IF statements:

IF (assaulted EQ 1 OR theft_victim EQ 1 or vandal_victim EQ 1) crime_victim = 1.

IF (assaulted EQ 0 AND theft_victim EQ 0 AND vandal_victim EQ 0) crime_victim = 0.

EXE.

This would create two mutually exclusive groups: (a) recent crime victims (regardless of how many times or in how many ways they were victimized) and (b) recent nonvictims.

Here is an example that combines both the IF (. . .) and the COMPUTE commands.

COMPUTE TALL = 0.

IF (SEX EQ 1 AND HEIGHT GE 74) TALL = 1.

IF (SEX EQ 2 AND HEIGHT GE 69) TALL = 1.

EXE.

This combination of COMPUTE and IF commands creates a new variable called TALL that has a default value of zero—and then replaces the zero with a value of 1 for all men who are 6′2″ (74″) or taller and all women who are 5′9″ (69″) or taller. A danger of this otherwise elegant approach is that if a person failed to report his or her gender and/or height, the variable TALL will still be set to zero for cases for which there are no data for height and/or gender. To avoid this problem, the researcher could have revised the command set as follows:

IF (SEX EQ 1 AND HEIGHT LT 74) TALL = 0.

IF (SEX EQ 1 AND HEIGHT GE 74) TALL = 1.

IF (SEX EQ 2 AND HEIGHT LT 69) TALL = 0.

IF (SEX EQ 2 AND HEIGHT GE 69) TALL = 1.

EXE.

Only people who report both their gender and their height will now receive a score on the dichotomous variable TALL.

RECODE: Sometimes (as was the case in Chapter 5 for some of the self-esteem items) you want to recode an item (e.g., to reverse score it). The RECODE statement is a very useful way to do this.

RECODE sex (0=2) (1=1) INTO gender.

EXECUTE.

In this example, the researcher may have wished to recode the variable "sex" from someone else's data file to use the familiar variable name and value labels that she prefers. As you will see later, it would also be easy to add a VALUE LABELS command that would specify exactly which gender is 1 and which is 2.

Here is an IF (. . .) command that uses the EQ command and the OR operand with alphanumeric (i.e., "String" or word) variables. Notice that you must use quotations marks with the EQ operand to specify alphanumeric variables.

IF (country EQ "Japan" OR country EQ "China" OR country EQ "India") East_West = 1.

IF (country EQ "Italy" OR country EQ "Germany" OR country EQ "Spain") East_West = 2.

EXE.

Although the IF (. . .) and COMPUTE commands are very flexible, they cannot do everything. Often, researchers need to perform mathematical, rather than logical, calculations on variables. When this is the case, you'll usually want to combine the COMPUTE command—and occasionally the IF (. . .) command—with some kind of mathematical operand. By far, the most common computations researchers use are simple addition (+), subtraction (−), division (/), and multiplication (*). However, many other mathematical calculations are possible, including, for example, log transformation, squaring or cubing variables, and taking the absolute value of a number. Here are a few quick examples of some useful mathematical operands along with brief explanations of what they do.

MEAN: The MEAN command is extremely useful because you cannot always get a mean of a set of scores by simply adding up the scores and dividing by the number of scores. For example, suppose you asked people whether they had been victims of three different crimes over the past year. What if a substantial number of respondents left one of the three crime questions blank? What if some respondents were only asked two of the three questions in the first place? For people missing only *one* crime question response, it would probably be reasonable to use the data they provided you for the two questions they did answer. The easiest way to do this would be to use the SPSS MEAN command, as shown below:

COMPUTE crime_index = MEAN.2 (assault, theft_vic, vandal_vic).

EXE.

This version of the mean.x command would give you a "crime_index" score for everyone who answered at least two (any two) of the crime questions—because of the ".2" suffix. So if Bubba *failed to answer* the vandalism question and said that he had *not* been assaulted but said that he *had* been a theft victim, his crime index score would be 0.50. Importantly, if you just use the MEAN command without a ".x" suffix, SPSS will calculate a mean score even if a respondent only has one score on the specific scale items listed in parentheses. So be careful what you specify and be especially careful not to omit the suffix (unless you're willing to live with a single score to represent a composite or index).

ABS: The ABS mathematical operand can also be very useful. Here is an example:

COMPUTE age_diff = ABS (wife_age − husb_age).

EXE.

This researcher seems to be studying married couples. The command that creates the new variable "age_diff" will take the **absolute value** of the difference between a husband's age and his wife's age, *without regard for which member of the couple is older.*

LG10: The base-10 logarithm command returns the logarithm of whatever value or variable is placed in parenthesis after the LG10 command (the power to which you need to raise the number 10 to yield the number in question). This can be useful for reducing skew so that extreme outliers are not weighted disproportionately in an analysis. The base-10 log of 100 is 2. The base-10 log of 1,000 is 3.

COMPUTE log10_height = LG10 (height).

EXE.

Note that it is the LG10 command (*after* the equal sign) that is recognized by SPSS.

SD or SD.X: The SD.X command calculates the standard deviation of whatever scores or variables are listed in parentheses. A researcher who collected negative mood scores on eight different occasions might be interested in the variability of people's negative moods. Imagine that she only wanted mood standard deviation scores for participants who had mood data on at least six occasions. Note that if you use SD (rather than the SD.X), SPSS will make $n = 2$ scores per participant the default requirement to produce a score.

COMPUTE st_dev_mood=SD.6 (mood1, mood2, mood3, mood4, mood5, mood6, mood7, mood8).

EXE.

SQRT: The square root command returns the square root of whatever value or variable is placed in parenthesis after SQRT. This can be useful for reducing the skew of a variable so that extreme outliers are not weighted disproportionately in an analysis.

COMPUTE sqrt_height = SQRT (height).

EXE.

VALUE LABELS: When you create new variables, it's often useful to remind users of the variables exactly what different scores on the variables indicate. The VALUE LABELS command allows you to do this (although you can also do it manually when in the SPSS Variable View mode). An example for body mass index (BMI) follows.

COMPUTE BMI = (weight*703)/ (height*height).

EXE.

IF (BMI LT 21) BMIcat4 =1.

IF (BMI GE 21 and BMI LT 25) BMIcat4 = 2.

IF (BMI GE 25 and BMI LT 30) BMIcat4 = 3.

IF (BMI GE 30) BMIcat4 = 4.

EXE.

VALUE LABELS BMIcat4 1 'underweight' 2 'normal' 3 'overweight' 4 'obese'.

EXE.

Note that the VALUE LABELS command stands on its own. It does not have to follow any IF (. . .) statements, and it could be used, for example, to label the unlabeled values of a variable that someone else previously created.

Needless to say there are many other things you can do to perform calculations or logical operations in SPSS syntax files. Recall, however, that whatever you plan to do with these statements, the statements will only create new variables in your SPSS data file once you highlight them and then run them (including the EXE commands, complete will all of your periods). Finally, we should add that anything you can do by typing in the syntax commands we reviewed here can *also* be done from SPSS drop-down menus. For example, you could do a COMPUTE command by clicking "Transform," then "Compute Variable. . . ." Assuming your instructor is a seasoned SPSS user, he or she could lead you through this alternate approach to variable manipulation and variable creation. For the time being, though, we hope this primer will get you started.

Appendix 6.3: Running a One-Sample Chi-Square Test in Older Versions of SPSS (SPSS 19 or Earlier)

After you select "Chi-square..." as shown in Image 24, you'll open the dialog box you see in Image 25, where you'll need to (a) send "exceeds" to the "Test Variable List," (b) click the "Options" button that you can see in the lower right-hand corner of Image 25, (c) click "Descriptives" from the Options box depicted in Image 26, (d) click "Continue" from that same Options box, and (e) click "Paste" when you get back to the "Chi-Square Test" box shown in Image 25.
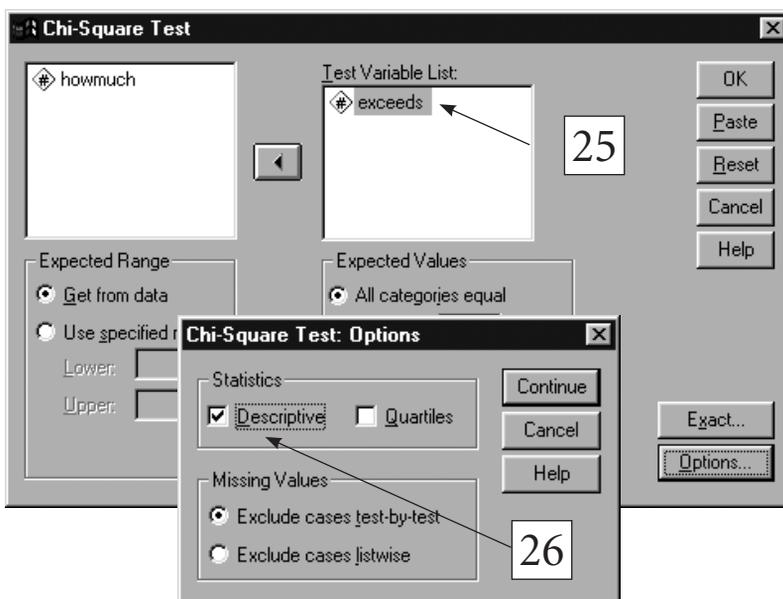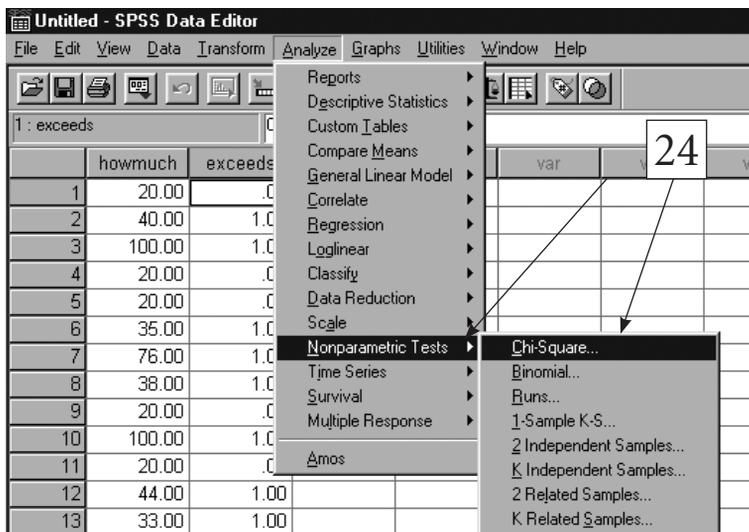
The changes this will create in your syntax file are shown in Image 4. As you know by now, you can run this new analysis by highlighting *only* the section that begins with "NPAR TEST" and then clicking the "Run" arrow (circled in Image 27).


```
Syntax1 - SPSS Syntax Editor
File  Edit  View  Analyze  Graphs  Utilities  Run

compute exceeds = 0.
if (howmuch gt 20) exceeds = 1.
execute.


T-TEST
 /TESTVAL=20
 /MISSING=ANALYSIS
 /VARIABLES=howmuch
 /CRITERIA=CIN (.95) .

NPAR TEST
 /CHISQUARE=exceeds
 /EXPECTED=EQUAL
 /STATISTICS  DESCRIPTIVES
 /MISSING ANALYSIS.
```

27