# ANALYZING DATA

## Comparing Means

**8**

> I note the obvious differences between each sort and type, but we are more alike, my friends, than we are unalike.
>
> *—Maya Angelou, poet*

## LEARNING GOALS

This chapter shows you how to assess differences between means using t-tests and analysis of variance. As with other bivariate data analysis, knowing when to use these statistical procedures and how to interpret them is central to testing hypotheses. By the end of the chapter, you should be able to understand what t-tests and ANOVA are and when they are suitable for data analysis (see Figure 8.1).

**Figure 8.1   Statistical Decision Steps (also see Statistical Analysis Decision Tree in Appendix)**

Every day it seems people are making comparisons among groups of people. Women are from Venus, and men are from Mars, or something like that. Millennials listen to different music than baby boomers, Asians don't vote for the same candidates as Hispanics, and certain cities have higher crime rates than others.

One of the fundamental questions asked in research is how alike or different people are. Good questionnaires ask demographic information that can provide important ways of breaking down survey data into interesting comparison categories. For example, you might be curious about the differences in the amount of television viewing among racial/ethnic groups or comparisons between people under 30 and people over 60 in the number of text messages they send per day.

If the data you want to compare are interval/ratio measures or at least ordinal measures with equal-appearing intervals, then there are several basic statistical techniques that can be used to answer the question of meaningful differences. For example, when you compare the average number of hours studied per week between first-year students and seniors, it is unlikely that the means will be exactly the same for both groups. Hence, the question is whether the difference that exists in hours studied between first-year students and seniors is a statistically significant one, as opposed to just any difference that could easily occur by chance alone.

When seeking to evaluate differences in means, we can use two very important and popular statistics: the *t-test* and *analysis of variance* (*ANOVA*).

## T-TESTS

When what you want to know (the dependent variable) is a ratio/interval measure, an ordinal measure with equal-appearing intervals, or a dichotomy, and a mean and standard deviation can be calculated for this dependent variable for each of two categories of comparison (often two categories of the independent variable), you can select a *t-test* for data analysis. T-tests ask whether the difference between two means is significantly different from zero. If both means were identical, then subtracting one from the other would result in zero. When data come from a sample of respondents, you are technically asking whether the difference in the true means in the populations from which these samples are derived is zero. To put it another way, you are asking a null question: There is no difference between the two means in the sample; they are the same and are from the same population.

Remember, you are comparing information that you collected from just one sample of respondents. If all possible samples of a certain size were taken from the population and all possible differences in means between two categories for all those samples were calculated (for example, mean number of tweets posted per day by urban people versus rural people) and plotted on a graph, the figure would approximate a normal

distribution, as discussed in earlier chapters. Actually, they would form a t-distribution, slightly different in shape from a perfect normal curve, or so "Student"—the nom de plume for W. S. Gosset, an employee in Ireland's Guinness brewery who developed the "Student's t"—realized.

What we are asking is the core question of statistical research in the social sciences: How much do our sample means deviate from the norm? Is the difference in means we found between two categories in our sample of respondents significantly different from zero? Where on the distribution of all possible mean differences is our mean difference? If it is 1.96 standard deviation units away (actually called *the standard error of the differences between means* when discussing distributions of mean differences), then you can conclude that the results are significant at the .05 level using a two-tailed or two-directional hypothesis. If we wanted to see whether one of the means was significantly larger than the other, or conversely one was smaller than the other, we are asking a one-directional question, and we would see whether the mean difference was 2.58 units away from the norm of zero difference in order to be significant at the .05 level.

Let's take an example in which we want to learn whether older or younger people download more movies on their smartphones in a typical month. First, we have to create two categories in order to make the comparison suitable for a t-test when there does not already exist a dichotomy. In this case, what is probably an interval/ratio measure (age) needs to be made into a simple dichotomy by deciding what constitutes "younger" and "older." We can look at the frequency distribution and cut the sample at the median, so that half the respondents are now above the median (older) and half are below (younger). Or we can decide ahead of time, on the basis of previous research or theory, and make 40 or under the "younger" respondents and those 41 and above the "older" people, even though we may not get a 50 percent split this way.

Now we have two groups, and we can calculate the average (mean) number of movies downloaded in a typical month. Remember, the questionnaire item has to be written as an interval/ratio measure, or an ordinal one with equal-appearing intervals, so we can accurately calculate a mean and standard deviation. We hypothesize that the difference in average number of downloaded movies between the younger and older people is zero, and assuming we have no preconceived idea about which group is more likely to be heavier downloaders, we proceed with a two-directional t-test to assess the differences in means. If we assume, on the basis of previous research, that younger people are more likely to download movies to their smartphones, then we will have a one-tailed hypothesis. We will definitely get some difference in means because it is unlikely that the two groups download exactly the same average number of movies, but we want to know if that difference is *statistically significant*.

The statistical software calculates a t-test value and presents us with the significance level. If the probability of obtaining that t-value by chance alone is less than .05, then we conclude that the category with the higher mean downloads more movies per month on average than the group with the lower mean. The difference between these two means is declared to be significantly different from zero; that is, if we wrote a hypothesis in the null form, we would reject no difference between the two groups and declare a significant finding (see Box 8.1). Rather than use the double negative of rejecting a null, we can more easily say that we are accepting a finding of difference between the two categories and then state which group uses their smartphones to watch movies more or less than the other group.

## BOX 8.1
# EXPLAINING THE T-TEST

Just as we can tell how far away an individual score is from the group mean in terms of standard deviation units (that is, z is a score minus the mean divided by the standard deviation), so, too, t is a mean minus another mean divided by the standard error of the differences in means.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$

In this formula, the two sample means are subtracted to find a difference, and the two population means are subtracted to find their difference. But because the null hypothesis is that the two population means are the same, the difference between them is zero. Hence, the second part of the numerator (where the $\mu$s are being subtracted) becomes zero and is not used in the calculations. The formula essentially is asking how large the difference between the two sample means is from zero, so all you need to know are the two sample means. It doesn't really matter what the real population means are since they are hypothesized to be the same, and therefore equal to zero when subtracted. Besides, if we knew what they were, we wouldn't be doing a sample survey in the first place!

The denominator is the *standard error of the differences between means*, and it is calculated differently depending on whether the variances of the two samples are the same or different and whether the data come from paired samples and how correlated they are, that is, the same people measured twice. Advanced statistics books and websites provide the formulas for the denominators. The main thing to remember is that the t is a ratio of the difference between two means (the numerator) divided by the standard error (the denominator).

When calculating the t-value for the difference between a sample mean and some known population mean, the formula is similarly a ratio of the mean differences divided by the standard error or deviation:

$$t = \frac{\bar{X} - \mu_{\mathrm{o}}}{s_x}$$

## Understanding T-Tests

When I was a kid, I often got frustrated when I did not reach the special line on the sign at the entrance to a really cool ride at the amusement park. I was too short to get on. I was sure I was shorter than everyone else in the world. But I guess I did not have it as bad as my sister and all her girlfriends, who were shorter than I. After all, we all know that girls are shorter than boys! Or do we? Let's take a look at this and see if indeed there is a significant difference between the average height of women and the average height of men in a survey of 151 college students.

Because height is an interval/ratio measure, you can calculate a mean and standard deviation for each category. And because you are comparing means between two groups within your sample, the t-test is the most appropriate statistic to use. You are testing to see if the difference between the two means is significantly greater than zero. The difference would be zero if the two means were the same (that is, the null hypothesis of no difference in heights). Let's just assume for the moment that you do not have any preconceived ideas about which group is taller or shorter. You are just interested in seeing if there is a difference and do not have any real interest in which group is taller or shorter. Therefore, you employ a two-directional test of significance.

## Independent Samples T-Test

When the two subsamples are independent of one another—that is, it is impossible that someone could select both categories of the questionnaire item—we use the *independent samples t-test*. The first part of the SPSS output in Table 8.1 indicates that the average height of the 97 women who answered the question is 65.39 inches (approximately 5′5″) and 70.72 inches (approximately 5′11″) for the 54 men. (Note that other statistical software programs may use different formats and provide other information in their outputs, although the same concepts should apply.) There definitely is a difference in height, but is it a statistically significant difference? What is the probability that this difference could have happened by chance and not be a result of real differences between men and women? Is 5 or 6 inches a meaningful difference, given the way the samples of men and women are distributed and the size of the samples? If the male and female respondents each are widely dispersed, smaller differences will be less meaningful. If the distributions are narrowly dispersed—that is, there is a lower standard deviation—then a small difference will be much more meaningful.

The standard deviations in the first part of Table 8.1 tell us that the heights for the men in the study are more dispersed than those for the women (3.36 versus 2.94) and

**Table 8.1 Example of Independent Samples T-Test, SPSS**

**GROUP STATISTICS**

| | Gender | N | Mean | Standard Deviation | Standard Error Mean |
|---|---|---|---|---|---|
| Height | Female | 97 | 65.3918 | 2.9353 | 0.2980 |
| | Male | 54 | 70.7222 | 3.3614 | 0.4574 |

**INDEPENDENT SAMPLES TEST**

| | LEVENE'S F-TEST FOR EQUALITY OF VARIANCES | | T-TEST FOR EQUALITY OF MEANS | | | | | 95% CONFIDENCE INTERVAL OF THE MEAN | |
|---|---|---|---|---|---|---|---|---|---|
| | F | Significance | t | df | Significance (two-tailed) | Mean Difference | Standard Error Difference | Lower | Upper |
| Height Equal variances assumed | 1.148 | 0.286 | −10.148 | 149 | .000 | −5.3305 | 0.5252 | −6.3684 | −4.2946 |
| Equal variances not assumed | | | −9.764 | 97.817 | .000 | −5.3305 | 0.5459 | −6.4139 | −4.2470 |

that the distribution of all possible means for men is wider than it is for women (the standard error of the means shows that). The calculation of the t-value is affected by these distributions.

The SPSS output for independent samples includes the results of the calculations for two different t ratios, depending on which standard error formula is used in the denominator. To determine which t-test to look at, you first must take a side trip and look at the Levene F-test for equality of variance (hereafter called the F-value). The second part of Table 8.1 shows whether the two variances (that is, the standard deviation squared) are equal to one another. If the F-value calculated is significant ($p < .05$), then you conclude there is a difference in variances and reject the null hypothesis of no difference; equal variances are not assumed. If the significance level for the F-value is not significant ($p > .05$), then you accept the fact there is no difference in variances and assume equal variances between the two categories. In the second part of the table, you can see that the significance level for the Levene F-value is greater than .05 (it is .286), so you assume that the variance of height for our sample of men and the variance of height for women are approximately equal.

But you're not done yet. You just took a side trip and finished figuring out if there was a difference in *variances*. Differences in variances affect the denominator in the calculation of the t-value, so it might be important to know this information, as you will see. Now you have to decide the main question as to whether there is a difference in *means*. Note that there are two t-values calculated: The one following the phrase "equal variances assumed" is −10.148, and the one following "equal variances not assumed" is −9.764. Because variances are assumed equal, you follow the t-value in the first line and see that its significance level is less than .05. In fact, it is so small, it's below .000; you see only the first few digits, and the actual number could be something like .00059, for example. You can now say that the probability of obtaining a t-value of 10.148, with samples of this size and with these standard deviations, by chance alone is less than 1 in 1,000 ($p < .001$).

This is certainly good enough for you to upload a photo of your results on Instagram and state that indeed men are taller than women in your sample. The difference between means is −5.33 inches. The minus signs are simply a function of which mean was subtracted first. Because the men's mean was subtracted from the women's, it results in a minus sign. If women's mean height were subtracted from the men's, then the result would still be 5.33 inches, but positive. The number is important for determining the significance level, not the minus or plus sign (see Box 8.2 for an example from an academic publication).

**BOX 8.2**
# T-TESTS IN AN ACADEMIC ARTICLE

For a study assessing body image perceptions and their relationship with mental health among 593 randomly selected secondary school adolescents in Nigeria, Otakpor and Ehimigbai (2016) used t-tests to compare mean scores for their measures. They used a well-validated self-report inventory to assess feelings of physical attractiveness, with high scores indicating a positive and satisfied "appearance evaluation." Another measure was "body areas satisfaction," with higher scores suggesting that respondents were generally content about the size and appearance of most areas of their bodies.

Respondents were divided into positive and negative groups for each of these variables. Mean scores on a "general health questionnaire" which focuses on psychiatric issues (scores over 5 indicate probable mental impairment) were then compared. Table 8.2 shows some of the results:

**Table 8.2   Body Image Study Results**

| VARIABLE | GENERAL HEALTH SCORE MEAN AND (SD) | df | t | P |
|---|---|---|---|---|
| Appearance evaluation: positive | 3.66 (0.94) | 591 | −4.26 | 0.00 |
| Appearance evaluation: negative | 3.97 (0.82) | | | |
| Body areas satisfaction: positive | 4.04 (1.05) | 591 | −2.51 | 0.01 |
| Body areas satisfaction: negative | 4.24 (0.86) | | | |

Both t-test analyses indicate a statistically significant result, with $p = .00$ and $p = .01$ (both less than .05) for the t-values. These findings suggest that those Nigerian teenagers who hold more positive feelings of their physical attractiveness and are more satisfied with the appearance and size of their body areas tend to have lower scores on the psychiatric mental health measure (general health score).

Note that the findings do not demonstrate a cause-and-effect relationship, just an association. By the way, a t-value greater than 2 is usually statistically significant, or as I like to remember, it takes "t for two" to be significant!

**Confidence Interval.**  If you were to draw 100 samples of the size of the sample in Table 8.1, 95 of them would have a mean difference in heights somewhere between 4.29 and 6.37 inches; this is called the *confidence interval* and is reported in the column "95% Confidence Interval of the Mean." In other words, you are 95 percent confident that the true difference in mean heights between men and women in the

population from which you drew this particular random sample is within that range. Remember you are inferring population information on the basis of what you found in one particular sample, so there is likely some error and a range within which the true information lies. And it also assumes you are drawing a random probability sample from that population.

## Paired Samples T-Test

The previous analysis was based on the assumption that the two categories being compared do not have any overlapping membership. These are independent samples. Sometimes, however, you need to compare two means for the same sample of people. For example, you might want to compare graduating seniors' average grades with their own first-year GPAs. Or you might want to look at the before-and-after effects of a new diet plan and compare the subjects' weights when they first started the diet with their current weights. When you do this, you must also take into account that there is a connection or correlation between the two measures; after all, it is the same pool of people being studied at two different points in time. You need to use a t-test formula that assesses the difference in means while taking into account the correlation between the two measures. In such cases, we calculate a *paired samples t-test*. Consider the following SPSS output from the 2014 General Social Survey about parents' education (Table 8.3). The same 1,773 respondents reported the highest year of both their mother's and father's educational attainment. Because we have pairs of data from each respondent measured on the same interval/ratio scale, we can use a paired samples test to see if there is a significant difference in schooling between mothers and fathers. On average, mothers completed 11.76 years of schooling, and fathers finished 11.82 years (12 years would indicate a high school graduate). It appears to be a small difference, but we still need to see if it is a statistically significant one or not. We also can see that the Pearson r correlation between mothers' and fathers' school attainment is 0.693, a very high association (reinforcing some research that suggests people tend to marry others of similar educational levels).

The difference in means is −.055 and (based on 1,772 degrees of freedom [$n - 1$], a strong correlation, and a .076 standard error of the mean) the t-value of −.724 is not statistically significant ($p = .469$, which is not less than .05). The lower and upper confidence interval also tells us that we are 95 percent confident that the true difference between mothers' and fathers' highest year of schooling in the population from which this sample is drawn is somewhere between −.203 and .094 years.

## Table 8.3 Example of Paired Samples T-Test, SPSS

**PAIRED SAMPLES STATISTICS**

| | | Mean | N | Standard Deviation | Standard Error Mean |
|---|---|---|---|---|---|
| Pair 1 | Highest year school completed, Mother | 11.76 | 1,773 | 3.860 | .092 |
| | Highest year school completed, Father | 11.82 | 1,773 | 4.225 | .100 |

**PAIRED SAMPLES CORRELATIONS**

| | | N | Correlation | Significance |
|---|---|---|---|---|
| Pair 1 | Highest year school completed, Mother and Highest year school completed, Father | 1,773 | 0.693 | .000 |

**PAIRED SAMPLES TEST**

| | PAIRED DIFFERENCES | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 95% CONFIDENCE INTERVAL OF THE DIFFERENCE | | | | Significance (two-tailed) |
| | Mean | Standard Deviation | Standard Error Mean | Lower | Upper | t | df | |
| Pair 1 Highest year school completed, Mother minus Highest year school completed, Father | −.055 | 3.183 | .076 | −.203 | .094 | −.724 | 1,772 | .469 |

## One-Sample T-Test

There are also a few occasions when we might want to compare a mean to some standard that already is known. For example, you wonder if the average GPA of a group of students in each major is really different from the average GPA of all students at the university. In this case, you would calculate a *one-sample t-test* (Table 8.4). Let's say you find out the grade point average for students in one particular major is 3.10 (out of 4.0). According to the registrar's office, the average GPA for the entire college is 2.95, so this number is entered as the one to compare with the average for your sample. But is that difference of 0.1513 GPA units statistically significant? It looks so small; it's difficult to tell if it is a meaningful difference unless you perform a test of statistical significance. Using the one-sample t-test, you learn that the probability of obtaining a t-value of 4.44 by chance is less than .001 and is therefore statistically significant. You can conclude that the sample of 150 has a larger mean GPA than the university as a whole. There's something special about this group—they must be sociology students!

Table 8.4  Example of One-Sample T-Test, SPSS

ONE-SAMPLE STATISTICS

| | N | Mean | Standard Deviation | Standard Error Mean |
|---|---|---|---|---|
| GPA | 150 | 3.1013 | 0.4171 | 0.034 |

ONE-SAMPLE TEST

TEST VALUE = 2.95

| | | | | | 95% CONFIDENCE INTERVAL OF THE DIFFERENCE | |
|---|---|---|---|---|---|---|
| | t | df | Significance (two-tailed) | Mean Difference | Lower | Upper |
| GPA | 4.444 | 149 | .000 | 0.1513 | 0.084 | 0.2186 |

## ANALYSIS OF VARIANCE (ANOVA)

With many research questions and hypotheses, we find ourselves wanting to compare more than two means at a time. For example, we might be interested in studying racial/ethnic differences in the amount of television viewing. When there are more than two categories for race/ethnicity, which is typically the case, we should use analysis of variance (ANOVA), a technique that asks whether the differences *within* a category are larger or smaller than those *between* (or, more grammatically, *among*) three or more categories. Like the t-test, ANOVA requires that the dependent variable be interval/ratio, a dichotomy, or an ordinal measure with equal-appearing intervals, so that a mean and variance can be calculated. The categories of the independent variable (called a *factor* in ANOVA) should be mutually exclusive in order to perform a one-way ANOVA. When there are multiple factors or independent variables, and interactive effects among those factors need to be assessed, or there are repeated measures on the same sample to be compared, use more sophisticated versions of ANOVA as discussed in advanced statistics books and on websites.

### Understanding ANOVA

ANOVA works by looking at the dispersion (variation) of scores around several means: the overall total dependent variable mean and the individual means for each of the categories of the main factor or independent variable. The computer program calculates an F-value, which, if significantly high enough, suggests that there is indeed a difference among the three or more means. In short, it tells you that the independent variable or factor has some effect on the dependent or outcome variable.

Consider the example in Table 8.5: Your hunch is that students in different years of college have different grade point averages. Everyone likes to think that their class is smarter than the others! Because grade point average is an interval/ratio measure, you calculate a mean and standard deviation for students in each level of college. As can be seen, there is some difference, but you need to be sure that the difference is not one that could have occurred merely by chance alone more than five times out of every 100 samples ($p > .05$). If it is, then those differences are not statistically significant, and we accept the null hypothesis of no difference.

The first part of the ANOVA output is descriptive data that show that seniors have higher GPAs (3.14) than others and higher than the overall average for all 150 students (3.10). Staying in college must make you smarter. But you clearly can't make any causal conclusions until you can first demonstrate that there is a relationship and rule out alternative explanations (such as those with lower grades were kicked out). Since you are comparing means for more than two categories of the main factor (four levels of the independent variable are measured), you use ANOVA, which calculates

**Table 8.5  Example of Analysis of Variance, SPSS**

| DESCRIPTIVES | | |
|---|---|---|
| GPA | Mean | Standard Deviation |
| First-years | 3.0714 | 0.53763 |
| Sophomores | 2.9130 | 0.46251 |
| Juniors | 3.1361 | 0.40867 |
| Seniors | 3.1424 | 0.38201 |
| Total | 3.1013 | 0.41708 |

| ANOVA | | | | | |
|---|---|---|---|---|---|
| GPA | Sum of Squares | df | Mean Square | F | Significance |
| Between groups | 0.995 | 3 | 0.332 | 1.942 | .125 |
| Within groups | 24.925 | 146 | 0.171 | | |
| Total | 25.920 | 149 | | | |

an F ratio comparing the variation between the four groups relative to the variation within each of them.

According to the second part of the table, the F-value of 1.942 is not significant at .125, which is greater than the minimum .05 we require for statistical significance. Therefore, we accept the null hypothesis of no difference and conclude that students' average grades cannot be attributed simply to what year of college they are in. One class does not appear to be earning better grades than the others.

As a rule of thumb, if the between groups variation is the same as the within groups variation (a null hypothesis of no difference), then the F statistic will be 1, since dividing a number by itself always results in one (see Box 8.3 for a mathematical explanation of ANOVA). In this situation, there is no influence of the independent variable categories on the dependent variable. The larger the ratio is, the more likely it will be statistically significant. However, it is not a measure of strength like a Pearson r correlation, so the size of F does not tell you much. More variation among the three or more groups in comparison to the variation within the groups results in an F increasingly larger than 1 and likely to be statistically significant. In other words, the independent

## BOX 8.3
# EXPLAINING THE F-TEST FORMULA

Sometimes following the mathematics underlying the calculation of these statistics can help you understand the concepts better. So for those of you who want to follow along, here's a look at the ANOVA math. The F is based on a formula that breaks down *the total sum of squares* (which is the sum of squaring each score subtracted from the overall mean) into the *within groups sum of squares* and the *between groups sum of squares*. Within groups sum of squares is calculated by taking each score, subtracting its category (or sub-sample) mean from it, squaring the difference, and summing those numbers:

$$\sum_{i=1}^{K} \left[ \sum_{j=1}^{n_i} \left( X_{ij} - \bar{X}_i \right)^2 \right]$$

Or it can be calculated by subtracting the between sum of squares from the total sum of squares:

$$SS_w = SS_{tot} - SS_B$$

The between groups sum involves subtracting the overall mean from each category or group mean, squaring the difference, and summing those numbers:

$$SS_B = \Sigma N_i (\bar{x}_i - \bar{x}_{tot})^2$$

The total sum of squares formula is

$$SS_{tot} = \Sigma x_{tot}^2 - \left( \frac{(\Sigma X_{tot})^2}{N} \right)$$

The F-value is the ratio of the between groups variation (the between group sum of squares divided by the *df*, or what is called the *mean squares*) to the within groups variation (the within groups sum of squares divided by the *df*):

$$F = \frac{ms_{Between}}{ms_{Within}}$$

Using the GPA example in Table 8.5, the F is not statistically significant because the variation (between groups mean squares of 0.332, which comes from dividing the between sum of squares, 0.995, by 3, the *df*) of the category means (3.07 for first-year students, 2.91 for sophomores, etc.) from the overall mean (3.10) is not larger than the variation (within groups) of the first-year students' GPAs from their category mean of 3.07, plus the variation of the sophomores' GPAs from their group's mean of 2.91, and so on.

variable categories make a difference in understanding the dependent variable when the between groups variation is larger than the within groups variation.

Like the Levene F-test for equality of variance used in the independent samples t-test, ANOVA involves analyzing (the "AN" of ANOVA) whether the two variances

(the between and within ones, the "VA" of ANOVA) are equal or not. If the probability of obtaining that F by chance is less than .05, we conclude that there is a significant difference in variances; we reject the null of no difference and agree that the between groups variance is larger. We then state that this is due to at least two of the categories having very different means from each other, because the original formula calculated how much these means differed from the overall mean. And that is why something called an analysis of *variance* is used to understand the difference among three or more *means*!

## DIFFERENCES NOT STRENGTH

With both the t-test and the F-test, we learn about the differences between means among several subgroups or samples. These are alternative ways of providing information about the relationship between two variables. They do not tell us, like Pearson r, what the strength of the relationship is, or how much variance is explained in the dependent variable by the independent as PRE statistics do. Many researchers make use of several statistical techniques at the same time to gather as much information as they need about their hypothesized relationships. It is not unusual then to see a study employing crosstabs, correlation coefficients, and differences in means tests for the same research questions and hypotheses. Each provides alternative insights about the relationships. Your job as a researcher is to make ethical decisions about which statistics best assist you in accurately answering your questions and learning what the statistics differently contribute to your understanding of the relationships.

We turn our attention in the next chapter to what happens when we want to analyze three or more variables at the same time. Social reality is such that behavior and attitudes are rarely explained by simple bivariate relationships. We often need to look at more complex relationships among multiple variables.

# REVIEW: WHAT DO THESE KEY TERMS MEAN?

ANOVA (F Value)

Between groups and within groups variances

Confidence interval

Independent samples t-test

Levene's F for equality of variance

One-sample t-test

Paired samples t-test

Standard error of the differences between means

T-tests

## TEST YOURSELF

1. What *statistic* would be most useful to test whether you reject or accept each of the following hypotheses? (Choose the *best* one if more than one can be used.)
   a. There is a difference in GPA between students living on-campus in Mead Hall, Atherton Hall, and Sanborn Hall.
   b. There is no difference in employees' mean scores on a scale (from 1 = terrible to 10 = wonderful) used to measure opinions about the food in the company's cafeteria. The survey was given both before and several months after a new company was hired to prepare the food.
   c. There is no difference in average number of books read in the past year between people whose zodiac sign is Scorpio versus Aries.
2. The General Social Survey measured attitudes toward classical music (where 1 = like very much and 5 = dislike very much) and the differences between those who graduated from college and those who didn't (see Table 8.6).
   a. Why is a t-test being used here?
   b. What does the Levene test tell you (in this SPSS output)?

### Table 8.6   Example From General Social Survey

| GROUP STATISTICS | | N | Mean | Standard Deviation | Standard Error Mean |
|---|---|---|---|---|---|
| Classical music | No college degree | 1,080 | 2.85 | 1.224 | 0.037 |
| | College degree | 344 | 2.06 | 0.983 | 0.053 |

| INDEPENDENT SAMPLES TEST | | LEVENE'S F-TEST FOR EQUALITY OF VARIANCES | | T-TEST FOR EQUALITY OF MEANS | | |
|---|---|---|---|---|---|---|
| | | F | Significance | t | df | Significance (2-tailed) |
| Classical music | Equal variances assumed | 47.774 | .000 | 10.913 | 1,422 | .000 |
| | Equal variances not assumed | | | 12.205 | 709.993 | .000 |

c. Which t-value do you use then?

d. Interpret the results for those who understand statistics.

e. How do you put the results into words for the general public?

# INTERPRET: WHAT DO THESE REAL EXAMPLES TELL US?

1. Meyer et al. (2010) report on how credibility is viewed by people using the Internet for the news. Although they studied various types of websites (blogs, straight news, etc.) on a variety of perceptions (expertise, opinionated, credible, etc.), let's take a look at just a few of the results.

    "Expertise" meant reporters sounded like experts, knew what they were talking about, and had done their homework for various types of Internet news sites. A "straight news" site represented standard objective news writing, a "blog" was a conversational format of personal views, a "collaborative" was cowritten by a reporter and local citizens, and an "opinionated" site was strong in its opinions, like Fox News.

    Each of the 140 college students read four different stories and was asked to rate his or her perception of expertise. Higher means indicate stronger belief in the expertise level of the news site. Paired t-tests were calculated between "straight news" and each of the other three website formats.

| Expertise | Mean | T-Value |
|---|---|---|
| Straight news | 3.53 | |
| Opinionated | 2.81 | 6.85** |
| Collaborative | 3.24 | 3.39** |
| Blog | 2.82 | 7.45** |

*df = 106, ** p < .01*

a. Why is a paired t-test used for these data?

b. State the null hypothesis for this study.

c. Explain to someone knowledgeable about statistics what is going on with the t-values and significance levels. And put into words for a general audience the conclusions you would draw from these results.

2. Here is some SPSS output from the General Social Survey assessing whether people with different outlooks on life (exciting, routine, dull) have different years of education on average. In short, does education make a difference in how people feel about life?

## DESCRIPTIVES

### HIGHEST YEAR OF SCHOOL COMPLETED

| | N | Mean | Standard Deviation | Standard Error | 95% Confidence Interval for Mean | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| Exciting | 433 | 13.64 | 3.108 | 0.149 | 13.35 | 13.94 |
| Routine | 503 | 12.44 | 2.726 | 0.122 | 12.20 | 12.68 |
| Dull | 41 | 10.49 | 3.257 | 0.509 | 9.46 | 11.52 |
| Total | 977 | 12.89 | 3.022 | 0.097 | 12.70 | 13.08 |

### ANOVA

### HIGHEST YEAR OF SCHOOL COMPLETED

| | Sum of Squares | df | Mean Square | F | Significance |
|---|---|---|---|---|---|
| Between groups | 582.721 | 2 | 291.360 | 34.077 | .000 |
| Within groups | 8,327.779 | 974 | 8.550 | | |
| Total | 8,910.499 | 976 | | | |

a. Why is an F-test (ANOVA) the best statistic to use here?

b. State a two-tailed null hypothesis that these statistics test.

c. Interpret the results in words. What do you conclude (a) for the statistical experts, and (b) for members of the general public who don't understand these numbers?

# CONSULT: WHAT COULD BE DONE?

1. You've been asked by the university's financial aid office to compare the cost of textbooks used in five majors: sociology, psychology, biology, history, and women's studies. Explain the steps you would take to do this and what statistic you would use to help you make a conclusion.

2. You've been asked by the Human Resources office at a company to assist in resolving a rumor going around the workplace. It seems that people believe that higher educated employees are getting larger salaries than less educated ones. Describe

what research you would do and the statistical procedures you would use to gather information to address the rumor.

# DECIDE: WHAT DO YOU DO NEXT?

For your study on how people develop and maintain diverse friendships, especially on social media, respond to the following items:

1. Look over your hypotheses and list the ones that could be assessed using a t-test. If none, develop a research question that uses t-tests. What would the independent variable look like, and what information would you need for the dependent variable?
2. List the hypotheses that could be assessed using ANOVA. If none, develop a research question that uses ANOVA. What would the independent variable look like, and what information would you need for the dependent variable?
3. If you have actual data from a survey, develop a few hypotheses suitable to analyze with at least one t-test and one ANOVA. Interpret the results and put into words what you have discovered.