CHAPTER 5

Hypothesis Tests with Means of Samples

Chapter Outline

In Chapter 4, we introduced the basic logic of hypothesis testing. The studies we used as examples had a sample of a single individual. As we noted, however, in actual practice, psychology research almost always involves a sample of many individuals. In this chapter, we build on what you have learned so far and consider hypothesis testing with a sample of more than one individual. For example, a social psychologist is interested in the potential effect of perceptions of people's personality on perceptions of their physical attractiveness. The researcher's theory predicts that, if a you are told that a person has positive personality qualities (such as kindness, warmth, a sense of humor, and intelligence), you will rate that person as more attractive than if no mention had been made of the person's personality qualities. From extensive previous research (in which no mention was made of personality qualities), the researcher has established the population mean and standard deviation of the attractiveness rating of a photo of a particular person. The researcher then recruits a sample of 64 individuals to rate the attractiveness of the person in the photograph. However, prior to rating the person, each individual is told that the person whose photograph they are going to rate has many positive personality qualities. In this chapter, you will learn how to test hypotheses in situations such as those presented in this example, situations in which the population has a *known mean and standard deviation* and in which a sample has *more than one individual.* Mainly, this requires examining in some detail a new kind of distribution, called a "distribution of means." (We will return to this example later in the chapter.)

The Distribution of Means

Hypothesis testing in the usual research situation, where you are studying a sample of many individuals, is exactly the same as you learned in Chapter 4—with an important exception. When you have more than one person in your sample, there is a special problem with Step ❷, determining the characteristics of the comparison distribution. In each of our examples so far, the comparison distribution has been a distribution of *individual scores* (such as the population of ages when individual babies start walking). A distribution of individual scores has been the correct comparison distribution because we have used examples with a sample of *one individual.* That is, there has been consistency between the type of sample score we have been dealing with (a score from *one individual*) and the comparison distribution (a distribution of *individual scores*).

Now, consider the situation when you have a sample of, say, 64 individuals (as in the attractiveness rating example). You now have a *group of 64 scores* (an attractiveness rating from each of the 64 people in the study). As you will recall from Chapter 2, the mean is a very useful representative value of a group of scores. Thus, the score you care about when there is more than one individual in your sample is the *mean of the group of scores.* In this example, you would focus on the mean of the 64 individuals' scores. If you were to compare the mean of this sample of 64 individuals' scores to a distribution of a population of individual scores, this would be a mismatch—like comparing apples to oranges. *Instead, when you are interested in the mean of a sample of 64 scores, you need a comparison distribution that is a distribution of means of samples of 64 scores.* We call such a comparison distribution a **distribution of means.** So, the scores in a distribution of means are *means,* not scores of individuals.

A distribution of means is a distribution of the means of each of lots and lots of samples of the same size, with each sample randomly taken from the same population of individuals. (Statisticians also call this distribution of means a *sampling distribution of the mean.* In this book, however, we use the term *distribution of means* to keep it clear that we are talking about populations of *means,* not samples or some kind of distribution of samples.)

The distribution of means is the correct comparison distribution when there is more than one person in a sample. Thus, in most research situations, determining the characteristics of a distribution of means is necessary for Step ❷ of the hypothesis-testing procedure, determining the characteristics of the comparison distribution.

distribution of means distribution of means of samples of a given size from a population (also called a *sampling distribution of the mean*); comparison distribution when testing hypotheses involving a single sample of more than one individual.

Building a Distribution of Means

To help you understand the idea of a distribution of means, we consider how you could build up such a distribution from an ordinary population distribution of individual scores. Suppose our population of individual scores was of the grade levels of the

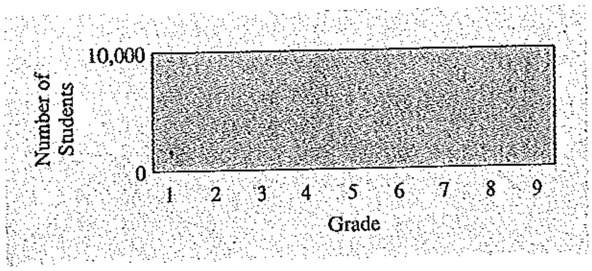

Figure 5–1 Distribution of grade levels among 90,000 schoolchildren (fictional data).

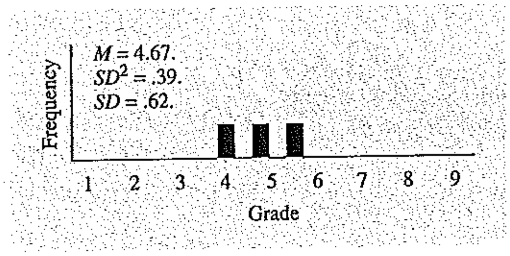

Figure 5–2 Distribution of the means of three randomly taken samples of two schoolchildren's grade levels each from a population of grade levels of 90,000 schoolchildren (fictional data).

90,000 elementary and junior-high schoolchildren in a particular region. Suppose further (to keep the example simple) that there are exactly 10,000 children at each grade level, from first through ninth grade. This population distribution would be rectangular, with a mean of 5, a variance of 6.67, and a standard deviation of 2.58 (see Figure 5–1).

Next, suppose that you wrote each child's grade level on a table tennis ball and put all 90,000 balls into a giant tub. The tub would have 10,000 balls with a 1 on them, 10,000 with a 2 on them, and so forth. You stir up the balls in the tub and then take two of them out. You have taken a random sample of two balls. Suppose one ball has a 2 on it and the other has a 9 on it. The mean grade level of this sample of two children's grade levels is 5.5, the average of 2 and 9. Now you put the balls back, mix up all the balls, and select two balls again. Maybe this time you get two 4s, making the mean of your second sample 4. Then you try again; this time you get a 2 and a 7, making your mean 4.5. So far you have three means: 5.5, 4, and 4.5.

Each of these three numbers is a mean of a sample of grade levels of two school children. And these three means can be thought of as a small distribution in its own right. The mean of this little distribution of means is 4.67 (the sum of 5.5, 4, and 4.5, divided by 3). The variance of this distribution of means is .39 (the variance of 5.5, 4, and 4.5). The standard deviation of this distribution of means is .62 (the square root of .39). A histogram of this distribution of three means is shown in Figure 5–2.

Suppose you continued selecting samples of two balls and taking the mean of the numbers on each pair of balls. The histogram of means would continue to grow. Figure 5–3 shows examples of distributions of means varying from a sample with just 50 means, up to a sample with 1,000 means (with each mean being of a sample of two randomly drawn balls). (We actually made the histograms shown in Figure 5–3 using a computer to make the random selections instead of using 90,000 table tennis balls and a giant tub.)

As you can imagine, the method we just described is not a practical way of determining the characteristics of a distribution of means. Fortunately, however, you can figure out the characteristics of a distribution of means directly, using some simple rules, without taking even one sample. The only information you need is (a) the characteristics of the distribution of the population of individuals and (b) the number of scores in each sample. (Don't worry for now about how you could know the characteristics of the population of individuals.) The laborious method of building up a distribution of means in the way we have just considered and the concise method you will learn shortly give the same result. We have had you think of the process in terms of the painstaking method only because it helps you understand the idea of a distribution of means.

TIP FOR SUCCESS

Before moving on to later chapters, be sure you fully understand the idea of a distribution of means (and why it is the correct comparison distribution when a sample contains more than one individual). You may need to go through this chapter a couple of times to achieve full understanding of this crucial concept.

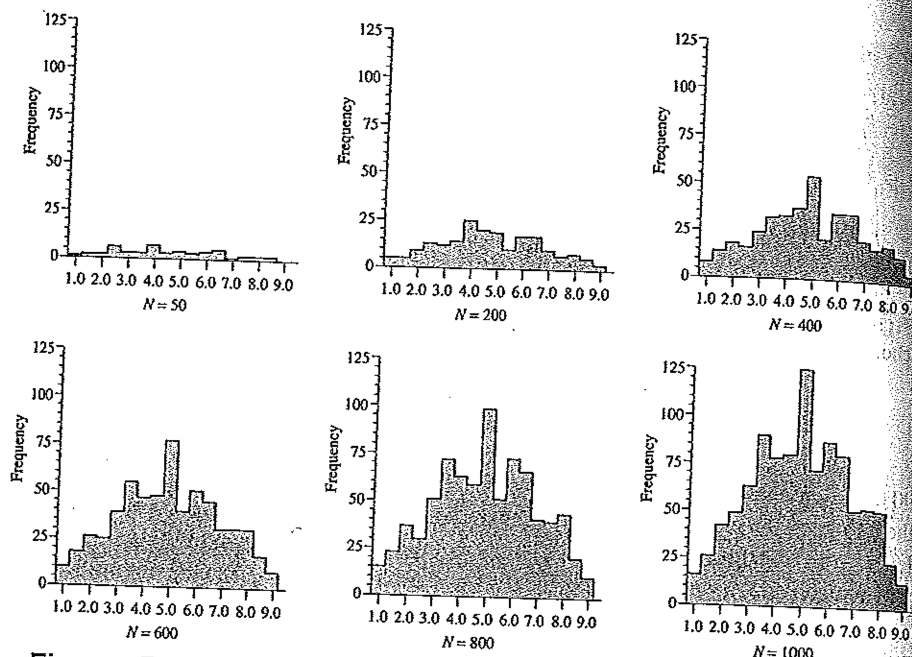

Figure 5–3 Histograms of means of two grade levels randomly selected from a large group of students with equal numbers of grades 1 through 9. Histograms are shown for 50 such means, 200 such means, 400 such means, 600 such means, 800 such means, and 1,000 such means. Notice that the histograms become increasingly like a normal curve as the number of means increases.

Determining the Characteristics of a Distribution of Means

Recall that Step ❷ of hypothesis testing involves determining the characteristics of the comparison distribution. The three key characteristics of the comparison distribution that you need to determine are:

1. Its mean.
2. Its spread (which you can measure using the variance and standard deviation).
3. Its shape.

Notice three things about the distribution of means we built in our example, as shown in Figure 5–3:

1. The mean of the distribution of means is about the same as the mean of the original population of individuals (both are 5).
2. The spread of the distribution of means is less than the spread of the distribution of the population of individuals.
3. The shape of the distribution of means is approximately normal.

The first two observations, regarding the mean and the spread, are true for all distributions of means. The third, regarding the shape, is true for most distributions of means. These three observations, in fact, illustrate three basic rules you can use to find the mean, the spread (that is, variance and standard deviation), and the shape of any distribution of means without having to write on plastic balls and take endless samples.

Now let's look at the three rules more closely. The first is for the **mean of a distribution of means.**

mean of a distribution of means the mean of a distribution of means of samples of a given size from a population; the same as the mean of the population of individuals.

Rule 1: The mean of a distribution of means is the same as the mean of the population of individuals. Stated as a formula,

$$\mu_M = \mu \tag{5-1}$$

The mean of a distribution of means is equal to the mean of the population of individuals.

μ_M is the mean of the distribution of means (it uses a Greek letter because the distribution of means is also a kind of population). μ is the mean of the population of individuals.

Each sample is based on randomly selected individuals from the population of individuals. Thus, the mean of a sample will sometimes be higher and sometimes lower than the mean of the whole population of individuals. However, because the selection process is random and we are taking a very large number of samples, eventually the high means and the low means perfectly balance each other out.

In Figure 5–3, as the number of sample means in the distributions of means increases, the mean of the distribution of means becomes more similar to the mean of the population of individuals, which in this example was 5. It can be proven mathematically that, if you took an infinite number of samples, the mean of the distribution of means of these samples would come out to be exactly the same as the mean of the distribution of individuals.

The second rule is about spread. Rule 2a is for the **variance of a distribution of means.**

Rule 2a: The variance of a distribution of means is the variance of the population of individuals divided by the number of individuals in each sample. A distribution of means will be less spread out than the distribution of individuals from which the samples are taken. If you are taking a sample of two scores, it is less likely that *both* scores will be extreme. Further, for a particular random sample to have an extreme mean, the two extreme scores would both have to be extreme in the same direction (both very high or both very low). Thus, having more than a single score in each sample has a moderating effect on the mean of such samples. In any one sample, the extremes tend to be balanced out by a middle score or by an extreme in the opposite direction. This makes each sample mean tend toward the middle and away from extreme values. With fewer extreme means, the variance of the means is less than the variance of the population of individuals.

Consider again our example. There were plenty of 1s and 9s in the population, making a fair amount of spread. That is, about a ninth of the time, if you were taking samples of single scores, you would get a 1 and about a ninth of the time you would get a 9. If you are taking samples of two at a time, you would get a sample with a mean of 1 (that is, in which *both* balls were 1s) or a mean of 9 (both balls 9s) much less often. Getting two balls that average out to a middle value such as 5 is much more likely. (This is because several combinations could give this result—1 and 9, 2 and 8, 3 and 7, 4 and 6, or two 5s).

The more individuals in each sample, the less spread out will be the means of the samples. This is because, the more scores in each sample, the rarer it will be for extremes in any particular sample not to be balanced out by middle scores or extremes in the other direction. In terms of the table tennis balls in our example, we rarely got a mean of 1 when taking samples of two balls at a time. If we were taking three balls at a time, getting a sample with a mean of 1 (all three balls would have to be 1s) is even less likely. Getting middle values for the means becomes even more likely.

Using samples of two balls at a time, the variance of the distribution of means came out to about 3.34. This is half of the variance of the population of individuals, which was 6.67. If we had built up a distribution of means using samples of three

μ_M mean of a distribution of means.

variance of a distribution of means variance of the population divided by the number of scores in each sample.

balls each, the variance of the distribution of means would have been 2.22. This is one-third of the variance of our population of individuals. Had we randomly selected five balls for each sample, the variance of the distribution of means would have been one-fifth of the variance of the population of individuals.

These examples follow a general rule—our Rule 2a for the distribution of means: the variance of a distribution of means is the variance of the population of individuals divided by the number of individuals in each of the samples. This rule holds in all situations and can be proven mathematically.

Here is Rule 2a stated as a formula:

> The variance of a distribution of means is the variance of the population of individuals divided by the number of individuals in each sample.

$$\sigma_M^2 = \frac{\sigma^2}{N} \qquad (5\text{-}2)$$

σ_M^2 is the variance of the distribution of means (it uses a Greek letter because the distribution of means is also a kind of population). σ^2 is the variance of the population of individuals, and N is the number of individuals in each sample.

TIP FOR SUCCESS

When you figure the variance of a distribution of means (σ_M^2), be sure to divide the *population variance* (σ^2) by the number of individuals in each sample. In many of the examples, you are told the population standard deviation (σ), which you will first have to square to find the population variance (σ^2); then you can use Formula 5-2 to find the variance of the distribution of means (σ_M^2).

In our example, the variance of the population of individual children's grade levels was 6.67, and there were two children's grade levels in each sample. Thus,

$$\sigma_M^2 = \frac{\sigma^2}{N} = \frac{6.67}{2} = 3.34$$

To use a different example, suppose a population had a variance of 400 and you wanted to know the variance of a distribution of means of 25 individuals each:

$$\sigma_M^2 = \frac{\sigma^2}{N} = \frac{400}{25} = 16$$

The second rule also tells us about the **standard deviation of a distribution of means.**

> The standard deviation of a distribution of means is the square root of the variance of the distribution of means and also the square root of the result of dividing the variance of the population of individuals by the number of individuals in each sample.

Rule 2b: The standard deviation of a distribution of means is the square root of the variance of the distribution of means. Stated as a formula,

$$\sigma_M = \sqrt{\sigma_M^2} = \sqrt{\frac{\sigma^2}{N}} \qquad (5\text{-}3)$$

σ_M is the standard deviation of the distribution of means.[1]

The standard deviation of the distribution of means also has a special name of its own, the **standard error of the mean (*SEM*)**, or the **standard error (*SE*)**, for short. (Thus, σ_M also stands for the standard error.) It has this name because it tells you how much the means of samples are typically "in error" as estimates of the mean of the population of individuals. That is, it tells you how much the various means in the distribution of means deviate from the mean of the population. We have more to say about the standard error later in the chapter.

Finally, the third rule for finding the characteristics of a distribution of means focuses on its shape.

Rule 3: The shape of a distribution of means is approximately normal if either (a) each sample is of 30 or more individuals or (b) the distribution of the population of individuals is normal. Whatever the shape of the distribution of the population of individuals, the distribution of means tends to be unimodal and symmetrical. In the grade-level example, the population distribution was rectangular.

σ_M^2 variance of a distribution of means.

standard deviation of a distribution of means square root of the variance of a distribution of means; also called *standard error of the mean (SEM)* and *standard error (SE)*.

σ_M standard deviation of a distribution of means.

standard error of the mean (*SEM*) same as *standard deviation of a distribution of means;* also called *standard error (SE).*

standard error (*SE*) same as *standard deviation of a distribution of means;* also called *standard error of the mean (SEM).*

CHAPTER 7

Introduction to *t* Tests

Single Sample and Dependent Means

Chapter Outline

At this point, you may think you know all about hypothesis testing. Here's a surprise: what you know will not help you much as a researcher. Why? The procedures for testing hypotheses described up to this point were, of course, absolutely necessary for what you will now learn. However, these procedures involved comparing a group of scores to a *known population*. In real research practice, you often compare two or more groups of scores to each other, without any direct information about populations. For example, you may have two scores for each person in a group of people, such as scores on an anxiety test before and after psychotherapy or number of familiar versus unfamiliar words recalled in a memory experiment. Or you might have one score per person for two groups of people, such as an experimental group and a control group in a study of the effect of sleep loss on problem solving, or comparing the self-esteem test scores of a group of 10-year-old girls to a group of 10-year-old boys.

These kinds of research situations are among the most common in psychology, where usually the only information available is from samples. Nothing is known about the populations that the samples are supposed to come from. In particular, the researcher does not know the variance of the populations involved, which is a crucial ingredient in Step ❷ of the hypothesis-testing process (determining the characteristics of the comparison distribution).

In this chapter, we first look at the solution to the problem of not knowing the population variance by focusing on a special situation: comparing the mean of a single sample to a population with a known mean but an unknown variance. Then, after describing how to handle this problem of not knowing the population variance, we go on to consider the situation in which there is no known population at all—the situation in which all we have are two scores for each of a number of people.

The hypothesis-testing procedures you learn in this chapter, those in which the population variance is unknown, are examples of ***t* tests.** The *t* test is sometimes called "Student's *t*" because its main principles were originally developed by William S. Gosset, who published his research articles anonymously using the name "Student" (see Box 7–1).

The *t* Test for a Single Sample

Let's begin with an example. Suppose your college newspaper reports an informal survey showing that students at your college study an average of 17 hours per week. However, you think that the students in your dormitory study much more than that. You randomly pick 16 students from your dormitory and ask them how much they study each day. (We will assume that they are all honest and accurate.) Your result is that these 16 students study an average of 21 hours per week. Should you conclude that students in your dormitory study more than the college average? Or should you conclude that your results are close enough to the college average that the small difference of 4 hours might well be due to your having picked, purely by chance, 16 of the more studious residents in your dormitory?

In this example you have scores for a sample of individuals and you want to compare the mean of this sample to a population for which you know the mean but not the variance. Hypothesis testing in this situation is called a ***t* test for a single sample.** (It is also called a *one-sample t test.*) The *t* test for a single sample works basically the same way as the Z test you learned in Chapter 5. In the studies we considered in that chapter, you had scores for a sample of individuals (such as a group of 64 students rating the attractiveness of a person in a photograph after being told that the person has positive personality qualities) and you wanted to compare the mean of this sample to a population (in this case, a population of students not told about the person's personality qualities). However, in the studies we considered in Chapter 5, you knew both the mean and variance of the general population to which you were going to compare your sample. In the situations we are now going to consider, everything is the same, but you don't know the population variance. This presents two important new wrinkles affecting the details of how you carry out two of the steps of the hypothesis-testing process.

The first important new wrinkle is in Step ❷. Because the population variance is not known, you have to estimate it. So the first new wrinkle we consider is how to estimate an unknown population variance. The other important new wrinkle affects Steps ❷ and ❸. When the population variance has to be estimated, the shape of the comparison

***t* test** hypothesis-testing procedure in which the population variance is unknown; it compares *t* scores from a sample to a comparison distribution called a *t* distribution.

***t* test for a single sample** hypothesis-testing procedure in which a sample mean is being compared to a known population mean and the population variance is unknown.

BOX 7–1 William S. Gosset, Alias "Student": Not a Mathematician, But a Practical Man

The Granger Collection

William S. Gosset graduated from Oxford University in 1899 with degrees in mathematics and chemistry. It happened that in the same year the Guinness brewers in Dublin, Ireland, were seeking a few young scientists to take a first-ever scientific look at beer making. Gosset took one of these jobs and soon had immersed himself in barley, hops, and vats of brew.

The problem was how to make beer of a consistently high quality. Scientists such as Gosset wanted to make the quality of beer less variable, and they were especially interested in finding the cause of bad batches. A proper scientist would say, "Conduct experiments!" But a business such as a brewery could not afford to waste money on experiments involving large numbers of vats, some of which any brewer worth his hops knew would fail. So Gosset was forced to contemplate the probability of, say, a certain strain of barley producing terrible beer when the experiment could consist of only a few batches of each strain. Adding to the problem was that he had no idea of the variability of a given strain of barley—perhaps some fields planted with the same strain grew better barley. (Does this sound familiar? Poor Gosset, like today's psychologists, had no idea of his population's variance.)

Gosset was up to the task, although at the time only he knew that. To his colleagues at the brewery, he was a professor of mathematics and not a proper brewer at all. To his statistical colleagues, mainly at the Biometric Laboratory at University College in London, he was a mere brewer and not a proper mathematician.

So Gosset discovered the *t* distribution and invented the *t* test—simplicity itself (compared to most of statistics)—for situations when samples are small and the variability of the larger population is unknown. However, the Guinness brewery did not allow its scientists to publish papers, because one Guinness scientist had revealed brewery secrets. To this day, most statisticians call the *t* distribution "Student's *t*" because Gosset wrote under the anonymous name "Student." A few of his fellow statisticians knew who "Student" was, but apparently meetings with others involved the secrecy worthy of a spy novel. The brewery learned of his scientific fame only at his death, when colleagues wanted to honor him.

In spite of his great achievements, Gosset often wrote in letters that his own work provided "only a rough idea of the thing" or so-and-so "really worked out the complete mathematics." He was remembered as a thoughtful, kind, humble man, sensitive to others' feelings. Gosset's friendliness and generosity with his time and ideas also resulted in many students and younger colleagues making major breakthroughs based on his help.

To learn more about William Gosset, go to *http://www-history.mcs.st-andrews.ac.uk/Biographies/Gosset.html.*

Sources: Peters (1987); Salsburg (2001); Stigler (1986); Tankard (1984).

---

distribution is not quite a normal curve; so the second new wrinkle we consider is the shape of the comparison distribution (for Step ❷) and how to use a special table to find the cutoff (Step ❸) on what is a slightly differently shaped distribution.

Let's return to the amount of studying example. Step ❶ of the hypothesis-testing procedure is to restate the problem as hypotheses about populations. There are two populations:

Population 1: The kind of students who live in your dormitory.
Population 2: The kind of students in general at your college.

The research hypothesis is that Population 1 students study more than Population 2 students; the null hypothesis is that Population 1 students do not study more than Population 2 students. So far, the problem is no different from those in Chapter 5.

Step ❷ is to determine the characteristics of the comparison distribution. In this example, its mean will be 17, what the survey found for students at your college generally (Population 2).

The next part of Step ❷ is finding the variance of the distribution of means. Now you face a problem. Up to now in this book, you have always known the variance of the population of individuals. Using that variance, you then figured the variance of the distribution of means. However, in the present example, the variance of the number of hours studied for students at your college (the Population 2 students) was not reported in the newspaper article. So you email the paper. Unfortunately, the reporter did not figure the variance, and the original survey results are no longer available. What to do?

Basic Principle of the *t* Test: Estimating the Population Variance from the Sample Scores

If you do not know the variance of the population of individuals, you can *estimate* it from what you do know—the scores of the people in your sample.

In the logic of hypothesis testing, the group of people you study is considered to be a random sample from a particular population. The variance of this sample ought to reflect the variance of that population. If the scores in the population have *a lot* of variation, then the scores in a sample randomly selected from that population should also have *a lot* of variation. If the population has *very little* variation, the scores in a sample from that population should also have *very little* variation. Thus, it should be possible to use the variation among the scores in the sample to make an informed guess about the spread of the scores in the population. That is, you could figure the variance of the sample's scores, and that should be *similar* to the variance of the scores in the population. (See Figure 7–1.)

There is, however, one small hitch. The variance of a sample will generally be slightly smaller than the variance of the population from which it is taken. For this reason, the variance of the sample is a **biased estimate** of the population variance.[1] It is a *biased estimate* because it consistently *underestimates* the actual variance of the population. (For example, if a population has a variance of 180, a typical sample

biased estimate estimate of a population parameter that is likely systematically to overestimate or underestimate the true value of the population parameter. For example, SD^2 would be a biased estimate of the population variance (it would systematically underestimate it).

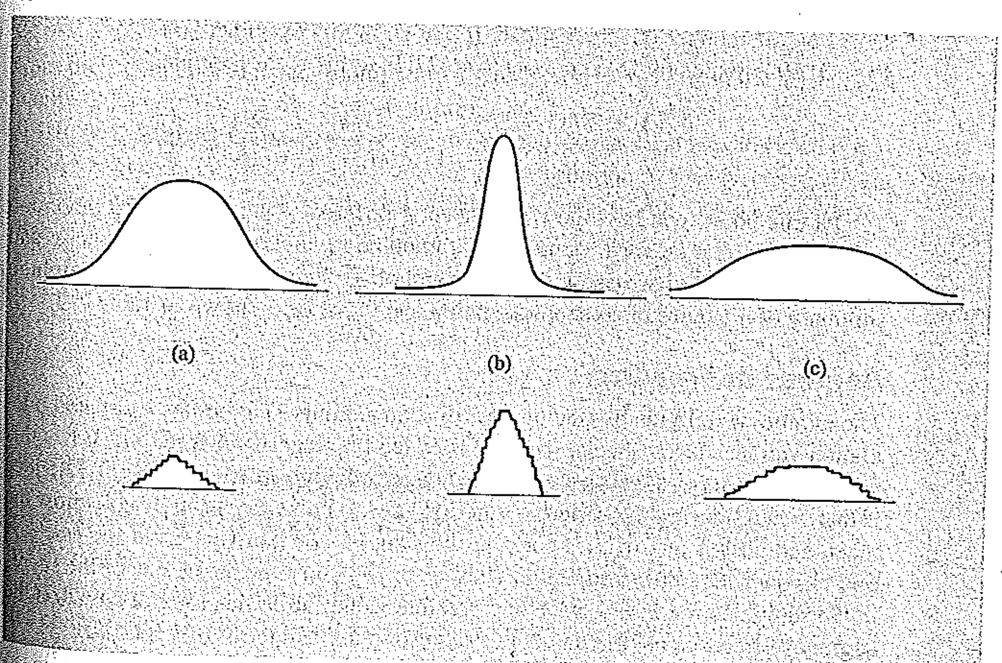

Figure 7–1 The variation in samples (as in each of the lower distributions) is similar to the variations in the populations they are taken from (each of the upper distributions).

unbiased estimate of the population variance (S^2) estimate of the population variance, based on sample scores, which has been corrected so that it is equally likely to overestimate or underestimate the true population variance; the correction used is dividing the sum of squared deviations by the sample size minus 1, instead of the usual procedure of dividing by the sample size directly.

degrees of freedom (df) number of scores free to vary when estimating a population parameter; usually part of a formula for making that estimate—for example, in the formula for estimating the population variance from a single sample, the degrees of freedom is the number of scores minus 1.

of 20 scores might have a variance of only 171.) If we used a biased estimate of the population variance in our research studies, our results would not be accurate. Therefore, we need to identify an *unbiased estimate* of the population variance.

Fortunately, you can figure an **unbiased estimate of the population variance** by slightly changing the ordinary variance formula. The ordinary variance formula is the sum of the squared deviation scores divided by the number of scores. The changed formula still starts with the sum of the squared deviation scores, but divides this by the number of scores *minus 1.* Dividing by a slightly smaller number makes the result slightly larger. Dividing by the number of scores minus 1 makes the variance you get just enough larger to make it an *unbiased estimate* of the population variance. (This unbiased estimate is our best estimate of the population variance. However, it is still an *estimate,* so it is unlikely to be exactly the same as the true population variance. But we can be certain that our unbiased estimate of the population variance is equally likely to be too high as it is to be too low. This is what makes the estimate *unbiased.*)

The symbol we will use for the unbiased estimate of the population variance is S^2. The formula is the usual variance formula, but now dividing by $N - 1$:

The estimated population variance is the sum of the squared deviation scores divided by the number of scores minus 1.

$$S^2 = \frac{\sum(X - M)^2}{N - 1} = \frac{SS}{N - 1} \qquad (7\text{-}1)$$

The estimated population standard deviation is the square root of the estimated population variance.

$$S = \sqrt{S^2} \qquad (7\text{-}2)$$

Let's return again to the example of hours spent studying and figure the estimated population variance from the sample's 16 scores. First, you figure the sum of squared deviation scores. (Subtract the mean from each of the scores, square those deviation scores, and add them.) Presume in our example that this comes out to 694 ($SS = 694$). To get the estimated population variance, you divide this sum of squared deviation scores by the number of scores minus 1; that is, in this example, you divide 694 by $16 - 1$; 694 divided by 15 comes out to 46.27. In terms of the formula,

$$S^2 = \frac{\sum(X - M)^2}{N - 1} = \frac{SS}{N - 1} = \frac{694}{16 - 1} = \frac{694}{15} = 46.27$$

At this point, you have now seen several different types of standard deviation and variance (that is, for a sample, for a population, and unbiased estimates); and each of these types has used a different symbol. To help you keep them straight, a summary of the types of standard deviation and variance is shown in Table 7–1.

Table 7-1 Summary of Different Types of Standard Deviation and Variance

Statistical Term	Symbol
Sample standard deviation	SD
Population standard deviation	σ
Estimated population standard deviation	S
Sample variance	SD^2
Population variance	σ^2
Estimated population variance	S^2

Degrees of Freedom

The number you divide by (the number of scores minus 1) to get the estimated population variance has a special name. It is called the **degrees of freedom.** It has this name because it is the number of scores in a sample that are "free to vary." The idea is that, when figuring the variance, you first have to know the mean. If you know the mean and all but one of the scores in the sample, you can figure out the one you don't know with a little arithmetic. Thus, once you know the mean, one of the scores in the sample is not free to have any possible value. So in this kind of situation the degrees of freedom are the number of scores minus 1. In terms of a formula,

The degrees of freedom are the number of scores in the sample minus 1.

$$df = N - 1 \qquad (7\text{-}3)$$

df is the degrees of freedom.

In our example, $df = 16 - 1 = 15$. (In some situations you learn about in later chapters, the degrees of freedom are figured a bit differently. This is because in those situations, the number of scores free to vary is different. For all the situations you learn about in this chapter, $df = N - 1$.)

The formula for the estimated population variance is often written using *df* instead of $N - 1$:

$$S^2 = \frac{\sum (X - M)^2}{df} = \frac{SS}{df} \quad (7\text{–}4)$$

The estimated population variance is the sum of squared deviations divided by the degrees of freedom.

The Standard Deviation of the Distribution of Means

Once you have figured the estimated population variance, you can figure the standard deviation of the comparison distribution using the same procedures you learned in Chapter 5. Just as before, when you have a sample of more than one, the comparison distribution is a *distribution of means,* and the variance of a distribution of means is the variance of the population of individuals divided by the sample size. You have just estimated the variance of the population. Thus, you can estimate the variance of the distribution of means by dividing the estimated population variance by the sample size. The standard deviation of the distribution of means is the square root of its variance. Stated as formulas,

$$S_M^2 = \frac{S^2}{N} \quad (7\text{–}5)$$

The variance of the distribution of means based on an estimated population variance is the estimated population variance divided by the number of scores in the sample.

$$S_M = \sqrt{S_M^2} \quad (7\text{–}6)$$

The standard deviation of the distribution of means based on an estimated population variance is the square root of the variance of the distribution of means based on an estimated population variance.

Note that, with an estimated population variance, the symbols for the variance and standard deviation of the distribution of means use S instead of σ.

In our example, the sample size was 16 and we worked out the estimated population variance to be 46.27. The variance of the distribution of means, based on that estimate, will be 2.89. That is, 46.27 divided by 16 equals 2.89. The standard deviation is 1.70, the square root of 2.89. In terms of the formulas,

$$S_M^2 = \frac{S^2}{N} = \frac{46.27}{16} = 2.89$$

$$S_M = \sqrt{S_M^2} = \sqrt{2.89} = 1.70$$

TIP FOR SUCCESS

Be sure that you fully understand the difference between S^2 and S_M^2. These terms look quite similar, but they are quite different. S^2 is the estimated variance of the population of individuals. S_M^2 is the estimated variance of the distribution of means (based on the estimated variance of the population of individuals, S^2).

The Shape of the Comparison Distribution When Using an Estimated Population Variance: The *t* Distribution

In Chapter 5 you learned that when the population distribution follows a normal curve, the shape of the distribution of means will also be a normal curve. However, this changes when you do hypothesis testing with an estimated population variance. When you are using an estimated population variance, you have less true information and more room for error. The mathematical effect is that there are likely to be slightly more extreme means than in an exact normal curve. Further, the smaller your

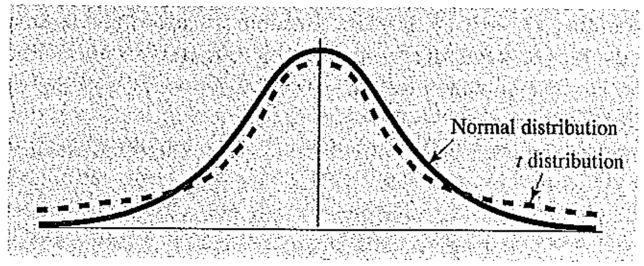

Figure 7–2 A *t* distribution (dashed blue line) compared to the normal curve (solid black line).

sample size, the bigger this tendency. This is because, with a smaller sample size, your estimate of the population variance is based on less information.

The result of all this is that, when doing hypothesis testing using an estimated variance, your comparison distribution will not be a normal curve. Instead, the comparison distribution will be a slightly different curve called a ***t* distribution.**

Actually, there is a whole family of *t* distributions. They vary in shape according to the degrees of freedom you used to estimate the population variance. However, for any particular degrees of freedom, there is only one *t* distribution.

Generally, *t* distributions look to the eye like a normal curve—bell-shaped, symmetrical, and unimodal. A *t* distribution differs subtly in having heavier tails (that is, slightly more scores at the extremes). Figure 7–2 shows the shape of a *t* distribution compared to a normal curve.

This slight difference in shape affects how extreme a score you need to reject the null hypothesis. As always, to reject the null hypothesis, your sample mean has to be in an extreme section of the comparison distribution of means, such as the top 5%. However, if the comparison distribution has more of its means in the tails than a normal curve would have, then the point where the top 5% begins has to be farther out on this comparison distribution. The result is that it takes a slightly more extreme sample mean to get a significant result when using a *t* distribution than when using a normal curve.

Just how much the *t* distribution differs from the normal curve depends on the degrees of freedom, the amount of information used in estimating the population variance. The *t* distribution differs most from the normal curve when the degrees of freedom are low (because your estimate of the population variance is based on a very small sample). For example, using the normal curve, you may recall that 1.64 is the cutoff for a one-tailed test at the .05 level. On a *t* distribution with 7 degrees of freedom (that is, with a sample size of 8), the cutoff is 1.895 for a one-tailed test at the .05 level. If your estimate is based on a larger sample, say a sample of 25 (so that $df = 24$), the cutoff is 1.711, a cutoff much closer to that for the normal curve. If your sample size is infinite, the *t* distribution is the same as the normal curve. (Of course, if your sample size were infinite, it would include the entire population!) But even with sample sizes of 30 or more, the *t* distribution is nearly identical to the normal curve.

Shortly, you will learn how to find the cutoff using a *t* distribution, but let's first return briefly to the example of how much students in your dorm study each week. You finally have everything you need for Step ❷ about the characteristics of the comparison distribution. We have already seen that the distribution of means in this example has a mean of 17 hours and a standard deviation of 1.70. You can now add that the shape of the comparison distribution will be a *t* distribution with 15 degrees of freedom.[2]

***t* distribution** mathematically defined curve that is the comparison distribution used in a *t* test.

The Cutoff Sample Score for Rejecting the Null Hypothesis: Using the *t* Table

Step ❸ of hypothesis testing is determining the cutoff for rejecting the null hypothesis. There is a different *t* distribution for any particular degrees of freedom. However, to avoid taking up pages and pages with tables for each possible *t* distribution, you use a simplified table that gives only the crucial cutoff points. We have included such a **t table** in the Appendix (Table A–2). Just as with the normal curve table, the *t* table shows only positive *t* scores. If you have a one-tailed test, you need to decide whether your cutoff score is a positive *t* score or a negative *t* score. If your one-tailed test is testing whether the mean of Population 1 is greater than the mean of Population 2, the cutoff *t* score is positive. However, if your one-tailed test is testing whether the mean of Population 1 is less than the mean of Population 2, the cutoff *t* score is negative.

In the hours-studied example, you have a one-tailed test. (You want to know whether students in your dorm study *more* than students in general at your college study.) You will probably want to use the 5% significance level, because the cost of a Type I error (mistakenly rejecting the null hypothesis) is not great. You have 16 participants, making 15 degrees of freedom for your estimate of the population variance.

Table 7–2 shows a portion of the *t* table from Table A–2 in the Appendix. Find the column for the .05 significance level for one-tailed tests and move down to the row for 15 degrees of freedom. The crucial cutoff is 1.753. In this example, you are testing whether students in your dormitory (Population 1) study *more* than students in general at your college (Population 2). In other words, you are testing whether

t table table of cutoff scores on the *t* distribution for various degrees of freedom, significance levels, and one- and two-tailed tests.

Table 7-2 Cutoff Scores for *t* Distributions with 1 Through 17 Degrees of Freedom (Highlighting Cutoff for Hours-Studied Example)

	One-Tailed Tests			Two-Tailed Tests		
df	.10	.05	.01	.10	.05	.01
1	3.078	6.314	31.821	6.314	12.706	63.657
2	1.886	2.920	6.965	2.920	4.303	9.925
3	1.638	2.353	4.541	2.353	3.182	5.841
4	1.533	2.132	3.747	2.132	2.776	4.604
5	1.476	2.015	3.365	2.015	2.571	4.032
6	1.440	1.943	3.143	1.943	2.447	3.708
7	1.415	1.895	2.998	1.895	2.365	3.500
8	1.397	1.860	2.897	1.860	2.306	3.356
9	1.383	1.833	2.822	1.833	2.262	3.250
10	1.372	1.813	2.764	1.813	2.228	3.170
11	1.364	1.796	2.718	1.796	2.201	3.106
12	1.356	1.783	2.681	1.783	2.179	3.055
13	1.350	1.771	2.651	1.771	2.161	3.013
14	1.345	1.762	2.625	1.762	2.145	2.977
15	1.341	**1.753**	2.603	1.753	2.132	2.947
16	1.337	1.746	2.584	1.746	2.120	2.921
17	1.334	1.740	2.567	1.740	2.110	2.898

students in your dormitory have a higher t score than students in general. This means that the cutoff t score is positive. Thus, you will reject the null hypothesis if your sample's mean is 1.753 or more standard deviations above the mean on the comparison distribution. (If you were using a known variance, you would have found your cutoff from a normal curve table. The Z score to reject the null hypothesis based on the normal curve would have been 1.645.)

One other point about using the t table: In the full t table in the Appendix, there are rows for each degree of freedom from 1 through 30, then for 35, 40, 45, and so on up to 100. Suppose your study has degrees of freedom between two of these higher values. To be safe, you should use the nearest degrees of freedom to yours given on the table that is less than yours. For example, in a study with 43 degrees of freedom, you would use the cutoff for $df = 40$.

The Sample Mean's Score on the Comparison Distribution: The *t* Score

Step ❹ of hypothesis testing is figuring your sample mean's score on the comparison distribution. In Chapter 5, this meant finding the Z score on the comparison distribution—the number of standard deviations your sample's mean is from the mean on the distribution. You do exactly the same thing when your comparison distribution is a t distribution. The only difference is that, instead of calling this a Z score, because it is from a t distribution, you call it a **t score**. In terms of a formula,

The t score is your sample's mean minus the population mean, divided by the standard deviation of the distribution of means.

$$t = \frac{M - \mu}{S_M} \qquad \textbf{(7–7)}$$

In the example, your sample's mean of 21 is 4 hours from the mean of the distribution of means, which amounts to 2.35 standard deviations from the mean (4 hours divided by the standard deviation of 1.70 hours).[3] That is, the t score in the example is 2.35. In terms of the formula,

$$t = \frac{M - \mu}{S_M} = \frac{21 - 17}{1.70} = \frac{4}{1.70} = 2.35$$

Deciding Whether to Reject the Null Hypothesis

Step ❺ of hypothesis testing is deciding whether to reject the null hypothesis. This step is exactly the same with a t test, as it was in the hypothesis-testing situations discussed in previous chapters. In the example, the cutoff t score was 1.753 and the actual t score for your sample was 2.35. Conclusion: reject the null hypothesis. The research hypothesis is supported that students in your dorm study more than students in the college overall.

Figure 7–3 shows the various distributions for this example.

Summary of Hypothesis Testing When the Population Variance Is Not Known

Table 7–3 compares the hypothesis-testing procedure we just considered (for a t test for a single sample) with the hypothesis-testing procedure for a Z test from Chapter 5. That is, we are comparing the current situation in which you know the population's mean but not its variance to the Chapter 5 situation, where you knew the population's mean *and* variance.

t score on a t distribution, number of standard deviations from the mean (like a Z score, but on a t distribution).

CHAPTER 8

The *t* Test for Independent Means

Chapter Outline

In the previous chapter, you learned how to use the *t* test for dependent means to compare two sets of scores from a *single group of people* (such as the same men measured on communication quality before and after premarital counseling).

In this chapter, you learn how to compare two sets of scores, one from each of *two entirely separate groups of people.* This is a very common situation in psychology research. For example, a study may compare the scores from individuals in an experimental group and individuals in a control group (or from a group of men and a group of women). This is a *t* test situation because you don't know the population variances (so they must be estimated). The scores of the two groups are independent of each other; so the test you learn in this chapter is called a **t test for independent means.**

***t* test for independent means** hypothesis-testing procedure in which there are two separate groups of people tested and in which the population variance is not known.

Let's consider an example. A team of researchers is interested in the effect on physical health of writing about thoughts and feelings associated with traumatic life events. This kind of writing is called expressive writing. Suppose the researchers recruit undergraduate students to take part in a study and randomly assign them to be in an expressive writing group or a control group. Students in the expressive writing group are instructed to write four 20-minute essays over four consecutive days about their most traumatic life experiences. Students in the control group write four 20-minute essays over four consecutive days describing their plans for that day. One month later, the researchers ask the students to rate their overall level of physical health (on a scale from 0 = *very poor health* to 100 = *perfect health*). Since the expressive writing and the control group contain different students, a *t* test for independent means is the appropriate test of the effect of expressive writing on physical health. We will return to this example later in the chapter. But first, you will learn about the logic of the *t* test for independent means, which involves learning about a new kind of distribution (called the *distribution of differences between means*).

The Distribution of Differences Between Means

In the previous chapter, you learned the logic and figuring for the *t* test for dependent means. In that chapter, the same group of people each had two scores; that is, you had a pair of scores for each person. This allowed you to figure a difference score for each person. You then carried out the hypothesis-testing procedure using these difference scores. The comparison distribution you used for this hypothesis testing was a *distribution of means of difference scores.*

In the situation you face in this chapter, the scores in one group are for different people than the scores in the other group. So you don't have any pairs of scores, as you did when the same group of people each had two scores. Thus, it wouldn't make sense to create difference scores, and you can't use difference scores for the hypothesis-testing procedure in this chapter. Instead, when the scores in one group are for different people than the scores in the other group, what you can compare is the *mean* of one group to the *mean* of the other group.

So the *t* test for independent means focuses on the *difference between the means* of the two groups. The hypothesis-testing procedure, however, for the most part works just like the hypothesis-testing procedures you have already learned. Since the focus is now on the difference between means, the comparison distribution is a **distribution of differences between means.**

A distribution of differences between means is, in a sense, two steps removed from the populations of individuals: First, there is a distribution of means from each population of individuals; second, there is a distribution of differences between pairs of means, one of each pair from each of these distributions of means.

Think of this distribution of differences between means as being built up as follows: (a) randomly select one mean from the distribution of means for the first group's population, (b) randomly select one mean from the distribution of means for the second group's population, and (c) subtract. (That is, take the mean from the first distribution of means and subtract the mean from the second distribution of means.) This gives a difference score between the two selected means. Then repeat the process. This creates a second difference score, a difference between the two newly selected means. Repeating this process a large number of times creates a distribution of differences between means. You would never actually create a distribution of differences between means using this lengthy method. But it shows clearly what makes up the distribution.

TIP FOR SUCCESS

The comparison distributions for the *t* test for dependent means and the *t* test for independent means have similar names: a distribution of means of difference scores, and a distribution of differences between means, respectively. Thus, it can be easy to confuse these comparison distributions. To remember which is which, think of the logic of each *t* test. The *t* test for dependent means involves *difference scores*. So, its comparison distribution is a distribution of means of *difference scores*. The *t* test for independent means involves *differences between means*. Thus, its comparison distribution is a distribution of *differences between means*.

distribution of differences between means distribution of differences between means of pairs of samples such that, for each pair of means, one is from one population and the other is from a second population; the comparison distribution in a *t* test for independent means.

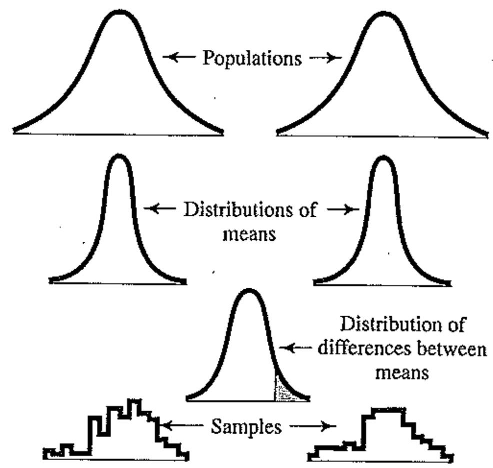

Figure 8–1 Diagram of the logic of a distribution of differences between means.

The Logic

Figure 8–1 shows the entire logical construction for a distribution of differences between means. At the top are the two population distributions. We do not know the characteristics of these population distributions, but we do know that if the null hypothesis is true, the two population means are the same. That is, the null hypothesis is that $\mu_1 = \mu_2$. We also can estimate the variance of these populations based on the sample information (these estimated variances will be S_1^2 and S_2^2).

Below each population distribution is the distribution of means for that population. Using the estimated population variance and knowing the size of each sample, you can figure the variance of each distribution of means in the usual way. (It is the estimated variance of its parent population divided by the size of the sample from that population that is being studied.)

Below these two distributions of means, and built from them, is the crucial *distribution of differences between means.* This distribution's variance is ultimately based on estimated population variances. Thus, we can think of it as a *t* distribution. The goal of a *t* test for independent means is to decide whether the difference between the means of your two actual samples is a more extreme difference than the cutoff difference on this distribution of differences between means. The two actual samples are shown (as histograms) at the bottom.

Remember, this whole procedure is really a kind of complicated castle in the air. It exists only in our minds to help us make decisions based on the results of an actual experiment. The only concrete reality in all of this is the actual scores in the two samples. You estimate the population variances from these sample scores. The variances of the two distributions of means are based entirely on these estimated population variances (and the sample sizes). And, as you will see shortly, the characteristics of the distribution of differences between means are based on these two distributions of means.

Still, the procedure is a powerful one. It has the power of mathematics and logic behind it. It helps you develop general knowledge based on the specifics of a particular study.

With this overview of the basic logic, we now turn to six key details: (1) the mean of the distribution of differences between means, (2) the estimated population variance, (3) the variance of the two distributions of means, (4) the variance and standard deviation of the distribution of differences between means, (5) the shape of the distribution of differences between means, and (6) the *t* score for the difference between the two means being compared.

Mean of the Distribution of Differences Between Means

In a t test for independent means, you are considering two populations: for example, one population from which an experimental group is taken and one population from which a control group is taken. In practice, you don't know the mean of either population. You do know that if the null hypothesis is true, these two populations have equal means. Also, if these two populations have equal means, the two distributions of means have equal means. (This is because each distribution of means has the same mean as its parent population of individuals.) Finally, if you take random samples from two distributions with equal means, the differences between the means of these random samples, in the long run, balance out to 0. The result of all this is the following: whatever the specifics of the study, you know that, if the null hypothesis is true, the distribution of differences between means has a mean of 0.

Estimating the Population Variance

In Chapter 7, you learned to estimate the population variance from the scores in your sample. It is the sum of squared deviation scores divided by the degrees of freedom (the number in the sample minus 1). To do a t test for independent means, it has to be reasonable to assume that the populations the two samples come from have the same variance (which, in statistical terms, is called *homogeneity of variance*). (If the null hypothesis is true, they also have the same mean. However, whether or not the null hypothesis is true, you must be able to assume that the two populations have the same variance.) Therefore, when you estimate the variance from the scores in either sample, you are getting two separate estimates of what should be the same number. In practice, the two estimates will almost never be exactly identical. Since they are both supposed to be estimating the same thing, the best solution is to average the two estimates to get the best single overall estimate. This is called the **pooled estimate of the population variance (S^2_{Pooled})**.

pooled estimate of the population variance (S^2_{Pooled}) in a t test for independent means, weighted average of the estimates of the population variance from two samples (each estimate weighted by the proportion consisting of its sample's degrees of freedom divided by the total degrees of freedom for both samples).

weighted average average in which the scores being averaged do not have equal influence on the total, as in figuring the pooled variance estimate in a t test for independent means.

In making this average, however, you also have to take into account that the two samples may not be the same size. If one sample is larger than the other, the estimate it provides is likely to be more accurate (because it is based on more information). If both samples are exactly the same size, you could just take an ordinary average of the two estimates. On the other hand, when they are not the same size, you need to make some adjustment in the averaging to give more weight to the larger sample. That is, you need a **weighted average,** an average weighted by the amount of information each sample provides.

Also, to be precise, the amount of information each sample provides is not its number of scores, but its degrees of freedom (its number of scores minus 1). Thus, your weighted average needs to be based on the degrees of freedom each sample provides. To find the weighted average, you figure out what proportion of the total degrees of freedom each sample contributes and multiply that proportion by the population variance estimate from that sample. Finally, you add up the two results, and that is your weighted, pooled estimate. In terms of a formula,

$$S^2_{\text{Pooled}} = \frac{df_1}{df_{\text{Total}}}(S^2_1) + \frac{df_2}{df_{\text{Total}}}(S^2_2) \qquad \textbf{(8–1)}$$

The pooled estimate of the population variance is the degrees of freedom in the first sample divided by the total degrees of freedom (from both samples), multiplied by the population estimate based on the first sample, plus the degrees of freedom in the second sample divided by the total degrees of freedom multiplied by the population variance estimate based on the second sample.

In this formula, S^2_{Pooled} is the pooled estimate of the population variance. df_1 is the degrees of freedom in the sample from Population 1, and df_2 is the degrees of freedom in the sample from Population 2. (Remember, each sample's df is its number of scores minus 1.) df_{Total} is the total degrees of freedom ($df_{\text{Total}} = df_1 + df_2$). S^2_1 is the estimate of the population variance based on the scores in Population 1's sample; S_2^2 is the estimate based on the scores in Population 2's sample.

Consider a study in which the population variance estimate based on an experimental group of 11 participants is 60, and the population variance estimate based on a control group of 31 participants is 80. The estimate from the experimental group is based on 10 degrees of freedom (11 participants minus 1), and the estimate from the control group is based on 30 degrees of freedom (31 minus 1). The total information on which the estimate is based is the total degrees of freedom—in this example, 40 (that is, 10 + 30). Thus, the experimental group provides one-quarter of the information (10/40 = 1/4), and the control group provides three-quarters of the information (30/40 = 3/4).

You then multiply the estimate from the experimental group by 1/4, making 15 (that is, 60 × 1/4 = 15), and you multiply the estimate from the control group by 3/4, making 60 (that is, 80 × 3/4 = 60). Adding the two gives an overall estimate of 15 plus 60, which is 75. Using the formula,

$$S^2_{\text{Pooled}} = \frac{df_1}{df_{\text{Total}}}(S_1^2) + \frac{df_2}{df_{\text{Total}}}(S_2^2) = \frac{10}{40}(60) + \frac{30}{40}(80)$$

$$= \frac{1}{4}(60) + \frac{3}{4}(80) = 15 + 60 = 75.$$

TIP FOR SUCCESS

You know you have made a mistake in figuring S^2_{Pooled} if it does not come out between the two estimates of the population variance. (You also know you have made a mistake if it does not come out closer to the estimate from the larger sample.)

Notice that this procedure does not give the same result as ordinary averaging (without weighting).

Ordinary averaging would give an estimate of 70 (that is, [60 + 80]/2 = 70). Your weighted, pooled estimate of the population variance of 75 is closer to the estimate based on the control group alone than to the estimate based on the experimental group alone. This is as it should be, because the control group estimate in this example was based on more information.

Figuring the Variance of Each of the Two Distributions of Means

The pooled estimate of the population variance is the best estimate for both populations. (Remember, to do a t test for independent means, you have to be able to assume that the two populations have the same variance.) However, even though the two populations have the same variance, if the samples are not the same size, the distributions of means taken from them do not have the same variance. That is because the variance of a distribution of means is the population variance divided by the sample size. In terms of formulas,

The variance of the distribution of means for the first population (based on an estimated population variance) is the pooled estimate of the population variance divided by the number of participants in the sample from the first population.

$$S^2_{M_1} = \frac{S^2_{\text{Pooled}}}{N_1} \quad (8\text{–}2)$$

$$S^2_{M_2} = \frac{S^2_{\text{Pooled}}}{N_2} \quad \textbf{(8–3)}$$

> The variance of the distribution of means for the second population (based on an estimated population variance) is the pooled estimate of the population variance divided by the number of participants in the sample from the second population.

Consider again the study with 11 in the experimental group and 31 in the control group. We figured the pooled estimate of the population variance to be 75. For the experimental group, the variance of the distribution of means would be 75/11, which is 6.82. For the control group, the variance would be 75/31, which is 2.42. Using the formulas,

$$S^2_{M_1} = \frac{S^2_{\text{Pooled}}}{N_1} = \frac{75}{11} = 6.82$$

$$S^2_{M_2} = \frac{S^2_{\text{Pooled}}}{N_2} = \frac{75}{31} = 2.42.$$

TIP FOR SUCCESS

Remember that when figuring estimated variances, you divide by the degrees of freedom. But when figuring the variance of a distribution of means, which does not involve any additional estimation, you divide by the actual number in the sample.

The Variance and Standard Deviation of the Distribution of Differences Between Means

The **variance of the distribution of differences between means** ($S^2_{\text{Difference}}$) is the variance of Population 1's distribution of means plus the variance of Population 2's distribution of means. (This is because, in a difference between two numbers, the variation in each contributes to the overall variation in their difference. It is like subtracting a moving number from a moving target.) Stated as a formula,

$$S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2} \quad \textbf{(8–4)}$$

> The variance of the distribution of differences between means is the variance of the distribution of means for the first population (based on an estimated population variance) plus the variance of the distribution of means for the second population (based on an estimated population variance).

The **standard deviation of the distribution of differences between means** ($S_{\text{Difference}}$) is the square root of the variance:

$$S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}} \quad \textbf{(8–5)}$$

> The standard deviation of the distribution of differences between means is the square root of the variance of the distribution of differences between means.

In the example we have been considering, the variance of the distribution of means for the experimental group was 6.82, and the variance of the distribution of means for the control group was 2.42; the variance of the distribution of the difference between means is thus 6.82 plus 2.42, which is 9.24. This makes the standard deviation of this distribution the square root of 9.24, which is 3.04. In terms of the formulas,

$$S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2} = 6.82 + 2.42 = 9.24$$

$$S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}} = \sqrt{9.24} = 3.04.$$

variance of a distribution of differences between means ($S^2_{\text{Difference}}$) one of the numbers figured as part of a *t* test for independent means; it equals the sum of the variances of the distributions of means associated with each of the two samples.

standard deviation of the distribution of differences between means ($S_{\text{Difference}}$) in a *t* test for independent means, square root of the variance of the distribution of differences between means.

Steps to Find the Standard Deviation of the Distribution of Differences Between Means

❶ **Figure the estimated population variances based on each sample.** That is, figure one estimate for each population using the formula $S^2 = SS/(N - 1)$.

❹ Figure the pooled estimate of the population variance:

$$S^2_{\text{Pooled}} = \frac{df_1}{df_{\text{Total}}}(S_1^2) + \frac{df_2}{df_{\text{Total}}}(S_2^2)$$

$$(df_1 = N_1 - 1 \text{ and } df_2 = N_2 - 1;\ df_{\text{Total}} = df_1 + df_2)$$

❺ Figure the variance of each distribution of means: $S^2_{M_1} = S^2_{\text{Pooled}}/N_1$ and $S^2_{M_2} = S^2_{\text{Pooled}}/N_2$.

❻ Figure the variance of the distribution of differences between means: $S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2}$.

❼ Figure the standard deviation of the distribution of differences between means: $S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}}$.

The Shape of the Distribution of Differences Between Means

The distribution of differences between means is based on estimated population variances. Thus, the distribution of differences between means (the comparison distribution) is a *t* distribution. The variance of this distribution is figured based on population variance estimates from two samples. Therefore, the degrees of freedom for this *t* distribution are the sum of the degrees of freedom of the two samples. In terms of a formula,

The total degrees of freedom for a *t* test for independent means is the degrees of freedom in the first sample plus the degrees of freedom in the second sample.

$$df_{\text{Total}} = df_1 + df_2 \qquad \textbf{(8–6)}$$

In the example we have been considering with an experimental group of 11 and a control group of 31, we saw earlier that the total degrees of freedom is 40 (that is, $11 - 1 = 10$; $31 - 1 = 30$; and $10 + 30 = 40$). To find the *t* score needed for significance, you look up the cutoff point in the *t* table in the row with 40 degrees of freedom. Suppose you are conducting a one-tailed test using the .05 significance level. The *t* table in the Appendix (Table A–2) shows a cutoff of 1.684 for 40 degrees of freedom. That is, for a result to be significant, the difference between the means has to be at least 1.684 standard deviations above the mean difference of 0 on the distribution of differences between means.

The t Score for the Difference Between the Two Actual Means

Here is how you figure the *t* score for Step ❹ of the hypothesis testing: First, figure the difference between your two samples' means. (That is, subtract one from the other). Then, figure out where this difference is on the distribution of differences between means. You do this by dividing your difference by the standard deviation of this distribution. In terms of a formula,

The *t* score is the difference between the two sample means divided by the standard deviation of the distribution of differences between means.

$$t = \frac{M_1 - M_2}{S_{\text{Difference}}} \qquad \textbf{(8–7)}$$

For our example, suppose the mean of the first sample is 198 and the mean of the second sample is 190. The difference between these two means is 8 (that is, $198 - 190 = 8$). Earlier we figured the standard deviation of the distribution of differences between means in this example to be 3.04. That would make a *t* score of 2.63 (that is, $8/3.04 = 2.63$). In other words, in this example the difference between the two means is 2.63 standard deviations above the mean of the distribution of differences between means. In terms of the formula,

$$t = \frac{M_1 - M_2}{S_{\text{Difference}}} = \frac{198 - 190}{3.04} = \frac{8}{3.04} = 2.63$$