Introduction to the *t* Test

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 so far 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 a score on a test of attitudes toward the courts before and after having gone through a law suit. 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 a new method of training teachers, or comparing the self-esteem test scores of a group of 10-year-old girls to a group of 10-year-old boys.

TIP FOR SUCCESS

In order to understand this chapter, you will need to know well the logic and procedures for hypothesis testing and the distribution of means.

These kinds of research situations are very common in behavioral and social science research, where usually the only information available is from the 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 you 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 what are called ***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 articles anonymously using the name "Student" (see Box 1).

The *t* Test for a Single Sample

Let's begin with the following situation. You carry out a study that gives you scores for a sample of people and you want to compare the mean of the scores in this sample to a population for which the mean is known but the variance is unknown. 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. Consider studies in which you have scores for a sample of individuals (such as a group of 64 fifth-graders who had taken a standard test with special instructions) and you wanted to compare the mean of this sample to a population (such as fifth-graders in general). In these studies, you know 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 both Steps ❷ and ❸. When the population variance has to be estimated, the shape of the comparison 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.

***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.

An Example

Suppose your university's newspaper reports an informal survey showing that students at your school 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 general in your dormitory study more than the university average? Or should you conclude that your results are close enough to the university 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 of your dormitory?

Step ❶ of the hypothesis-testing process 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 university.

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.

BOX 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 a degree 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 researchers, 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 meetings with others involved 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).

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 university 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 text, 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 university (the Population 2 students) was not reported in the newspaper article. So you email the newspaper. Unfortunately, the reporter did not calculate 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 samples from that population should also have *very little* variation. Thus, it should be possible to use the variation among the scores in a sample to make an informed guess about the variation 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 1).

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).

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

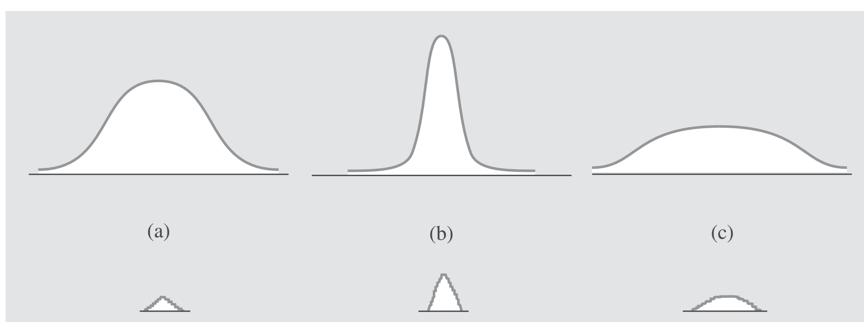


Figure 1 The variation in samples (lower distributions) is similar to the variation in the populations from which they are taken (upper distributions).

[1]A sample's variance is slightly smaller than the population's because it is based on deviations from the sample's mean. A sample's mean is the optimal balance point for its scores. Thus, deviations of a sample's scores from its mean will be smaller than deviations from any other number. The mean of a sample generally is not exactly the same as the mean of the population it comes from. Thus, deviations of a sample's scores from its mean will generally be smaller than deviations of that sample's scores from the true population mean.

population. (For example, if a population has a variance of 180, a typical sample 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 sum 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 resulting variance 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 $N - 1$ (instead of by N):

$$S^2 = \frac{\Sigma(X - M)^2}{N - 1} \qquad (1)$$

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

Let's return to our 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. (That is, you 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. 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,

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

unbiased estimate of the population variance 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 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.

df Degrees of freedom.

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** (and is abbreviated as *df*). 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 is the number of scores minus 1. In terms of a formula,

$$df = N - 1 \qquad (2)$$

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

df is the degrees of freedom.

In our example, $df = 16 - 1 = 15$. 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 $N - 1$.

$$S^2 = \frac{\Sigma(X - M)^2}{df} \qquad (3)$$

The estimated population variance is the sum of squared deviation scores 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. As always, 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,

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.

$$S_M^2 = \frac{S^2}{N} \tag{4}$$

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.

$$S_M = \sqrt{S_M^2} \tag{5}$$

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

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. Although S^2 and S_M^2 look similar, 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

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 using 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. Furthermore, the smaller your 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.

TIP FOR SUCCESS

Be careful. To find the variance of a distribution of means, you always divide the population variance by the sample size. This is true whether the population's variance is known or estimated. In our example, you divided the population variance, which you had estimated, by 16 (the sample size). It is only when making the estimate of the population variance that you divide by the sample size minus 1. That is, the degrees of freedom are used only when estimating the variance of the population of individuals.

The result of all of 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.**

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

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, completely symmetrical, and unimodal. A *t* distribution differs subtly in having higher tails (that is, slightly more scores at the extremes). Figure 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

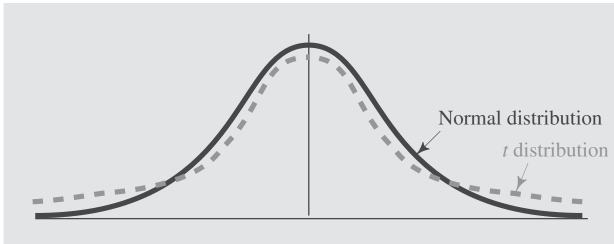


Figure 2 A *t* distribution (dashed blue line) compared to the normal curve (solid black line).

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 the 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 dormitory 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]

[2]Statisticians make a subtle distinction in this situation between the comparison distribution and the distribution of means. (We avoid this distinction to simplify your learning of what is already fairly difficult.) The general procedure of hypothesis testing with means of samples can be described as figuring a *Z* score for your sample's mean, and then comparing this *Z* score to a cutoff *Z* score from the normal curve table. We described this process as using the distribution of means as your comparison distribution. Statisticians would say that actually you are comparing the *Z* score you figured for your sample mean to a distribution of *Z* scores (which is simply an ordinary normal curve). Similarly, for a *t* test, you will see shortly that we figure what is called a *t* score in the same way as a *Z* score, but using a standard deviation of the distribution of means based on an estimated population variance. You then compare the *t* score you figure for your sample mean to a cutoff *t* score from a *t* distribution table. According to the formal statistical logic, you can think of this process as involving a comparison distribution of *t* scores for means, not of means themselves.

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

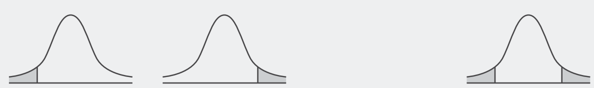
Step ❸ of hypothesis testing is determining the cutoff for rejecting the null hypothesis. There is a different *t* distribution for any particular degrees of freedom. However, to avoid taking up pages and pages with tables for each different *t* distribution, you use a simplified table that gives only the crucial cutoff points. Such a ***t* table** is Table 2 of the appendix "Tables." 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 dormitory study *more* than students in general at your university.) 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 people, making 15 degrees of freedom for your estimate of the population variance.

Table 1 shows a portion of the *t* table. 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 university (Population 2). In other words, you are testing whether students in your dormitory have a higher *t* score than students in general. This means that the cutoff *t* score is positive.

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

Table 1 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	2.97
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

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 needed to reject the null hypothesis based on the normal curve would have been 1.64.)

One other point about using the *t* table: In the full *t* table, 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 that is given in 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 the Z test, 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 comparison 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,

$$t = \frac{M - \text{Population } M}{S_M} \qquad (6)$$

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

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). That is, the *t* score in the example is 2.35. In terms of the formula,

$$t = \frac{M - \text{Population } M}{S_M} = \frac{21 - 17}{1.70} = \frac{4}{1.70} = 2.35.$$

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. You compare the cutoff score from Step ❸ with the sample's score on the comparison distribution from Step ❹. In the example, the cutoff *t* score was 1.753 and the actual *t* score for our sample was 2.35. Conclusion: reject the null hypothesis. The research hypothesis is supported that students in your dormitory study more than students in the university overall. Figure 3 shows the various distributions for this example.

Summary of Hypothesis Testing When the Population Variance Is Not Known

Table 2 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. That is, we are comparing the current situation in which you know the population's mean but not its variance to a situation where you know the population's mean *and* variance. Table 3 summarizes the steps of hypothesis testing for the *t* test for a single sample.

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

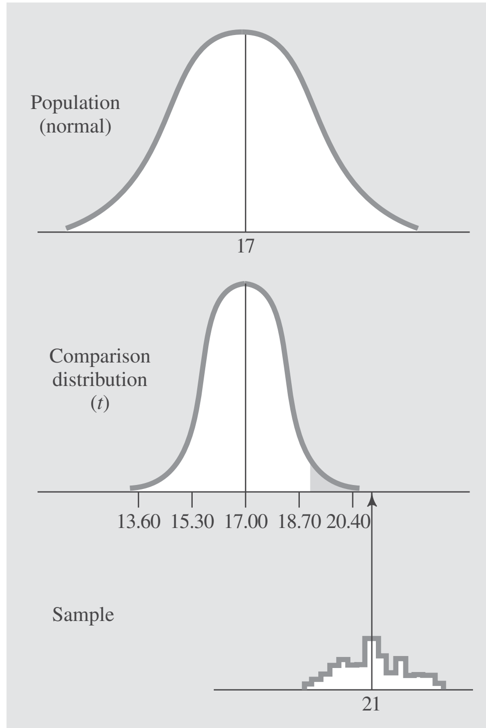


Figure 3 Distributions for the hours studied example.

Table 2 Hypothesis Testing with a Single Sample Mean When Population Variance Is Unknown (*t* Test for a Single Sample) Compared to When Population Variance Is Known (*Z* Test)

Steps in Hypothesis Testing	Difference from When Population Variance Is Known
❶ Restate the question as a research hypothesis and a null hypothesis about the populations.	No difference in method.
❷ Determine the characteristics of the comparison distribution:	
Population mean	No difference in method.
Population variance	*Estimate* from the sample.
Standard deviation of the distribution of means	No difference in method (but based on *estimated* population variance).
Shape of the comparison distribution	Use the *t* distribution with $df = N - 1$.
❸ Determine the significance cutoff.	Use the *t* table.
❹ Determine your sample's score on the comparison distribution.	No difference in method (but called a *t* score).
❺ Decide whether to reject the null hypothesis.	No difference in method.

Table 3 Steps for a *t* Test for a Single Sample

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.**

❷ **Determine the characteristics of the comparison distribution.**

a. The mean is the same as the known population mean.

b. The standard deviation is figured as follows:

Ⓐ Figure the estimated population variance: $S^2 = [\Sigma(X - M)^2]/df$

Ⓑ Figure the variance of the distribution of means: $S_M^2 = S^2/N$

Ⓒ Figure the standard deviation of the distribution of means: $S_M = \sqrt{S_M^2}$

c. The shape is a *t* distribution with $N - 1$ degrees of freedom.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.**

a. Decide the significance level and whether to use a one-tailed or a two-tailed test.

b. Look up the appropriate cutoff in a *t* table.

❹ **Determine your sample's score on the comparison distribution:** $t = (M - \text{Population } M)/S_M$

❺ **Decide whether to reject the null hypothesis:** Compare the scores from Steps ❸ and ❹

How are you doing?

1. In what sense is a sample's variance a biased estimate of the variance of the population from which the sample is taken? That is, in what way does a sample's variance typically differ from the population's?
2. What is the difference between the usual formula for figuring the variance and the formula for estimating a population's variance from the scores in a sample (that is, the formula for an unbiased estimate of the population variance)?
3. (a) What are degrees of freedom? (b) How do you figure the degrees of freedom in a *t* test for a single sample? (c) What do they have to do with estimating the population variance? (d) What do they have to do with the *t* distribution?
4. (a) How does a *t* distribution differ from a normal curve? (b) How do degrees of freedom affect this? (c) What is the effect of the difference on hypothesis testing?
5. List three differences in how you do hypothesis testing for a *t* test for a single sample versus for the *Z* test.
6. A population has a mean of 23. A sample of four is given an experimental procedure and has scores of 20, 22, 22, and 20. Test the hypothesis that the procedure produces a lower score. Use the .05 significance level. (a) Use the steps of hypothesis testing and (b) make a sketch of the distributions involved.

Answers

1. The sample's variance will in general be slightly smaller than the variance of the population from which the sample is taken.

2. In the usual formula you divide by the number of participants (N); in the formula for estimating a population's variance from the scores in a sample, you divide by the number of participants in the sample minus 1 ($N - 1$).

3. (a) The degrees of freedom are the number of scores free to vary. (b) The degrees of freedom in a t test for a single sample are the number of scores in the sample minus 1. (c) In estimating the population variance, the formula is the sum of squared deviations divided by the degrees of freedom. (d) t distributions differ slightly from each other according to the degrees of freedom.
4. (a) A t distribution differs from a normal curve in that it has heavier tails; that is, it has more scores at the extremes. (b) The more degrees of freedom, the closer the shape (including the tails) is to a normal curve. (c) The cutoffs for significance are more extreme for a t distribution than for a normal curve.
5. In the t test you (a) estimate the population variance from the sample (it is not known in advance); (b) you look up the cutoff on a t table in which you also have to take into account the degrees of freedom (you don't use a normal curve table); and (c) your sample's score on the comparison distribution, which is a t distribution (not a normal curve), is called a t score (not a Z score).
6. (a) Steps of hypothesis testing:

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are two populations:

Population 1: People who are given the experimental procedure.
Population 2: The general population.

The research hypothesis is that Population 1 will score lower than Population 2. The null hypothesis is that Population 1 will not score lower than Population 2.

❷ **Determine the characteristics of the comparison distribution.**

a. The mean of the distribution of means is 23.
b. The standard deviation is figured as follows:

Ⓐ Figure the estimated population variance. You first need to figure the sample mean, which is $(20 + 22 + 22 + 20)/4 = 84/4 = 21$. The estimated population variance is $S^2 = [\Sigma (X - M)^2]/df = [(20 - 21)^2 + (22 - 21)^2 + (22 - 21)^2 + (20 - 21)^2]/(4 - 1) = (-1^2 + 1^2 + 1^2 + -1^2)/3 = (1 + 1 + 1 + 1)/3 = 4/3 = 1.33$.

Ⓑ Figure the variance of the distribution of means: $S_M^2 = S^2/N = 1.33/4 = .33$.

Ⓒ Figure the standard deviation of the distribution of means: $S_M = \sqrt{S_M^2} = \sqrt{.33} = .57$.

c. The shape of the comparison distribution will be a t distribution with $df = 3$.

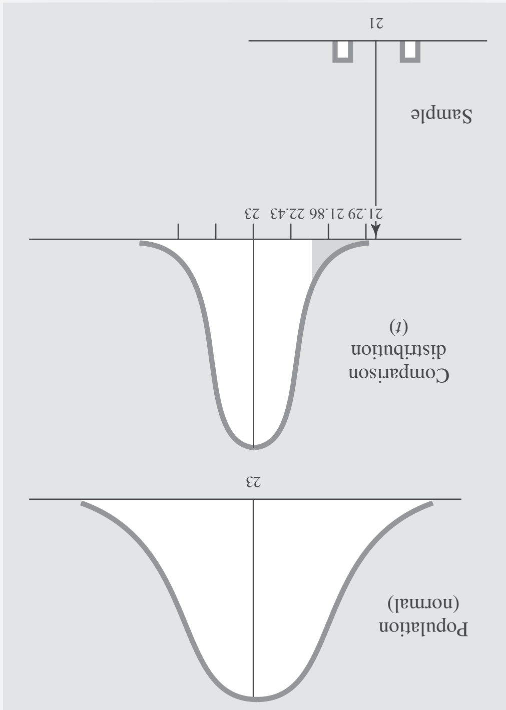
❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** From Table A–2, the cutoff for a one-tailed t test at the .05 level for $df = 3$ is -2.353. The cutoff t score is negative, since the research hypothesis is that the procedure produces a *lower* score.

❹ **Determine your sample's score on the comparison distribution.** $t = (M - \text{Population } M)/S_M = (21 - 23)/.57 = -2/.57 = -3.51$.

❺ **Decide whether to reject the null hypothesis.** The t of -3.51 is more extreme than the cutoff t of -2.353. Therefore, reject the null hypothesis; the research hypothesis is supported.

(b) Sketches of distributions are shown in Figure 4.

Figure 4 Distributions for answer to "How are you doing?" question 6b.



The *t* Test for Dependent Means

The situation you just learned about (the *t* test for a single sample) is for when you know the population mean but not its variance and you have a single sample of scores. It turns out that in most research you do not even know the population's *mean;* plus, in most research situations, you usually have not one set but *two* sets of scores. These two things, not knowing the population mean and having two sets of scores, almost always go together. This situation is very common.

The rest of this chapter focuses specifically on the important research situation in which you have two scores from each person in your sample. This kind of research situation is called a **repeated-measures design** (also known as a *within-subjects design*). A common example is when you measure the same people before and after some social or psychological intervention. This specific kind of repeated-measures situation is called a *before–after design.* For example, a researcher might measure the quality of men's communication before and after receiving premarital counseling.

The hypothesis-testing procedure for the situation in which each person is measured twice (that is, the situation in which we have a repeated-measures design) is called a ***t* test for dependent means.** It has the name *dependent means* because the mean for each group of scores (for example, a group of before-scores and a group of after-scores) are dependent on each other in that they are both from the same people.

repeated-measures design Research strategy in which each person is tested more than once; same as *within-subjects design.*

***t* test for dependent means** Hypothesis-testing procedure in which there are two scores for each person and the population variance is not known; it determines the significance of a hypothesis that is being tested using difference or change scores from a single group of people.

The *t* test for dependent means is also called a *paired-sample t test,* a *t test for correlated means,* a *t test for matched samples,* and a *t test for matched pairs.* Each of these names comes from the same idea that in this kind of *t* test you are comparing two sets of scores that are related to each other in a direct way, such as each person being tested before and after some procedure.

You do a *t* test for dependent means in exactly the same way as a *t* test for a single sample, except that (a) you use something called *difference scores* and (b) you assume that the population mean (of the difference scores) is 0. We will now consider each of these two new aspects.

Difference Scores

With a repeated-measures design, your sample includes two scores for each person instead of just one. The way you handle this is to make the two scores per person into one score per person! You do this magic by creating **difference scores:** For each person, you subtract one score from the other. If the difference is before versus after, difference scores are also called *change scores.*

Consider the example of the quality of men's communication before and after receiving premarital counseling. The researcher subtracts the communication quality score before the counseling from the communication quality score after the counseling. This gives an after-minus-before difference score for each man. When the two scores are a before-score and an after-score, you usually take the after-score minus the before-score to indicate the *change.*

Once you have the difference score for each person in the study, you do the rest of the hypothesis testing with difference scores. That is, you treat the study as if there were a single sample of scores (scores that in this situation happen to be difference scores).[3]

Population of Difference Scores with a Mean of 0

So far in the research situations we have considered in this text, you have always known the mean of the population to which you compared your sample's mean. For example, in the university dormitory survey of hours studied, you knew the population mean was 17 hours. However, now we are using difference scores, and we usually don't know the mean of the population of difference scores.

Here is the solution. Ordinarily, the null hypothesis in a repeated-measures design is that on the average there is *no difference* between the two groups of scores. For example, the null hypothesis in a study of the quality of men's communication before and after receiving premarital counseling is that on the average there is no difference between communication quality before and after the counseling. What does *no difference* mean? Saying there is on average no difference is the same as saying that the mean of the population of the difference scores is 0. Therefore, when working with difference scores, you are comparing the population of difference scores that your sample of difference scores comes from to a population of difference scores with a mean of 0. In other words, with a *t* test for dependent means, what we call Population 2 will ordinarily have a mean of 0 (that is, a population of difference scores that has a mean of 0).

difference score Difference between a person's score on one testing and the same person's score on another testing; often an after score minus a before score, in which case it is also called a *change score.*

[3]You can also use a *t* test for dependent means when you have scores from pairs of research participants. You consider each pair as if it were one person and figure a difference score for each pair. For example, suppose you have 30 married couples and want to test whether husbands are consistently older than wives. You could figure for each couple a difference score of husband's age minus wife's age. The rest of the figuring would then be exactly the same as for an ordinary *t* test for dependent means. When the *t* test for dependent means is used in this way, it is sometimes called a *t test for matched pairs.*

Example of a *t* Test for Dependent Means

Olthoff (1989) tested the communication quality of engaged couples 3 months before and again 3 months after marriage. One group studied was 19 couples who received ordinary (very minimal) premarital counseling from the ministers who were going to marry them. (To keep the example simple, we focus on just this one group, and only on the husbands in the group. Scores for wives were similar, though somewhat more varied, making it a more complicated example for learning the *t* test procedure.)

The scores for the 19 husbands are listed in the "Before" and "After" columns in Table 4, followed by all the *t* test figuring. The crucial column for starting the analysis is the difference scores. For example, the first husband, whose communication quality was 126 before marriage and 115 after, had a difference of −11. (We figured after minus before, so that an increase is positive and a decrease, as for this husband,

Table 4 *t* Test for Communication Quality Scores Before and After Marriage for 19 Husbands Who Received Ordinary Premarital Counseling

Husband	Communication Quality		Difference (After − Before)	Deviation (Difference − *M*)	Squared Deviation
	Before	*After*			
A	126	115	−11	1.05	1.10
B	133	125	−8	4.05	16.40
C	126	96	−30	−17.95	322.20
D	115	115	0	12.05	145.20
E	108	119	11	23.05	531.30
F	109	82	−27	−14.95	233.50
G	124	93	−31	−18.95	359.10
H	98	109	11	23.05	531.30
I	95	72	−23	−10.95	119.90
J	120	104	−16	−3.95	15.60
K	118	107	−11	1.05	1.10
L	126	118	−8	4.05	16.40
M	121	102	−19	−6.95	48.30
N	116	115	−1	11.05	122.10
O	94	83	−11	1.05	1.10
P	105	87	−18	−5.95	35.40
Q	123	121	−2	10.05	101.00
R	125	100	−25	−12.95	167.70
S	128	118	−10	2.05	4.20
Σ:	2,210	1,981	−229		2,772.90

For difference scores:

$M = -229/19 = -12.05.$

Population $M = 0$ (assumed as a no-change baseline of comparison).

$S^2 = [\Sigma(X - M)^2]/df = 2{,}772.90/(19 - 1) = 154.05.$

$S_M^2 = S^2/N = 154.05/19 = 8.11.$

$S_M = \sqrt{S_M^2} = \sqrt{8.11} = 2.85.$

t with $df = 18$ needed for 5% level, two-tailed $= \pm 2.101.$

$t = (M - \text{Population } M)/S_M = (-12.05 - 0)/2.85 = -4.23.$

Decision: Reject the null hypothesis.

Source: Data from Olthoff (1989).

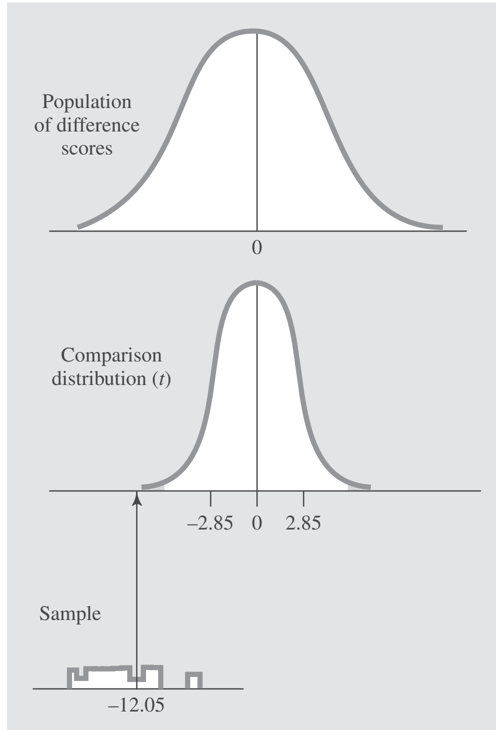


Figure 5 Distributions for the Olthoff (1989) example of a *t* test for dependent means.

is negative.) The mean of the difference scores is −12.05. That is, on the average, these 19 husbands' communication quality decreased by about 12 points.

Is this decrease significant? In other words, how likely is it that this sample of change scores is a random sample from a population of change scores whose mean is 0? Let's carry out the hypothesis-testing procedure. (The distributions involved are shown in Figure 5.)

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are two populations:

Population 1: Husbands who receive ordinary premarital counseling.
Population 2: Husbands whose communication quality does not change from before to after marriage.

TIP FOR SUCCESS
Population 2 is the population for whom the null hypothesis is true.

The research hypothesis is that Population 1's mean difference score (communication quality after marriage minus communication quality before marriage) is different from Population 2's mean difference score (of zero). That is, the research hypothesis is that husbands who receive ordinary premarital counseling, like the husbands Olthoff (1989) studied, *do* change in communication quality from before to after marriage. The null hypothesis is that the populations are the same. That is, the null hypothesis is that the husbands who receive ordinary premarital counseling do *not* change in their communication quality from before to after marriage.

Notice that you have no actual information about Population 2 husbands. The husbands in the study are a sample of Population 1 husbands. For the purposes of hypothesis testing, you set up Population 2 as a kind of straw man comparison group. That is, for the purpose of the analysis, you set up a comparison group of husbands who, if measured before and after marriage, would on average show no difference.

❷ **Determine the characteristics of the comparison distribution.** If the null hypothesis is true, the mean of the population of difference scores is 0. The variance of the population of difference scores can be estimated from the sample of difference scores. As shown in Table 4, the sum of squared deviations of the difference scores from the mean of the difference scores is 2,772.90. With 19 husbands in the study, there are 18 degrees of freedom. Dividing the sum of squared deviation scores by the degrees of freedom gives an estimated population variance of difference scores of 154.05.

The distribution of means of difference scores has a mean of 0, the same as the mean of the population of difference scores. The variance of the distribution of means of difference scores is the estimated population variance of difference scores (154.05) divided by the sample size (19), which gives 8.11. The standard deviation of the distribution of means of difference scores is 2.85, the square root of 8.11.

Because Olthoff (1989) was using an estimated population variance, the comparison distribution is a *t* distribution. The estimate of the population variance of difference scores is based on 18 degrees of freedom, so this comparison distribution is a *t* distribution for 18 degrees of freedom.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** Olthoff (1989) used a two-tailed test to allow for either an increase or a decrease in communication quality. Using the .05 significance level and 18 degrees of freedom, Table 2 of the appendix "Tables" shows that to reject the null hypothesis you need a *t* score of $+2.101$ or above, or a *t* score of -2.101 or below.

❹ **Determine your sample's score on the comparison distribution.** Olthoff's (1989) sample had a mean difference score of -12.05. That is, the mean was 12.05 points below the mean of 0 on the distribution of means of difference scores. The standard deviation of the distribution of means of difference scores is 2.85. Thus, the mean of the difference scores of -12.05 is 4.23 standard deviations below the mean of the distribution of means of difference scores. So Olthoff's sample of difference scores has a *t* score of -4.23.

❺ **Decide whether to reject the null hypothesis.** The *t* of -4.23 for the sample of difference scores is more extreme than the needed *t* of ± 2.101. Thus, Olthoff (1989) could reject the null hypothesis. This suggests that the husbands are from a population in which husbands' communication quality is different after marriage from what it was before (it got lower).

Olthoff's (1989) actual study was more complex. You may be interested to know that he found that the wives also showed this decrease in communication quality after marriage. But a group of similar engaged couples who were given special communication skills training by their ministers (much more than the usual short sessions) had no significant decline in marital communication quality after marriage. In fact, there is a great deal of research showing that on the average marital happiness declines steeply over time (VanLaningham, Johnson, & Amato, 2001). And many studies have now shown the value of a full course of premarital communications training. For example, a representative survey of 3,344 adults in the United States showed that those who had attended a premarital communication program had significantly greater marital satisfaction, had less marital conflict, and were 31% less likely to divorce (Stanley, Amato, Johnson, & Markman, 2006). Furthermore, benefits were greatest for those with a college education!

TIP FOR SUCCESS

You now have to deal with some rather complex terms, such as the *standard deviation of the distribution of means of difference scores.* Although these terms are complex, there is good logic behind them. The best way to understand such terms is to break them down into manageable pieces. For example, you will notice that these new terms are the same as the terms for the *t* test for a single sample, with the added phrase "of difference scores." This phrase has been added because all of the figuring for the *t* test for dependent means uses difference scores.

TIP FOR SUCCESS

Step ❷ of hypothesis testing for the *t* test for dependent means is a little trickier than previously. This can make it easy to lose track of the purpose of this step. Step ❷ of hypothesis testing determines the characteristics of the comparison distribution. In the case of the *t* test for dependent means, this comparison distribution is a distribution of means of difference scores. The key characteristics of this distribution are its mean (which is assumed to equal 0), its standard deviation (which is estimated as S_M), and its shape (a *t* distribution with degrees of freedom equal to the sample size minus 1).

Table 5 Steps for a *t* Test for Dependent Means

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.**

❷ **Determine the characteristics of the comparison distribution.**

a. Make each person's two scores into a difference score. Do all the remaining steps using these difference scores.

b. Figure the mean of the difference scores.

c. Assume a mean of the distribution of means of difference scores of 0.

d. The standard deviation of the distribution of means of difference scores is figured as follows:

Ⓐ Figure the estimated population variance of difference scores: $S^2 = [\Sigma(X - M)^2]/df$

Ⓑ Figure the variance of the distribution of means of difference scores: $S_M^2 = S^2/N$

Ⓒ Figure the standard deviation of the distribution of means of difference scores: $S_M = \sqrt{S_M^2}$

e. The shape is a *t* distribution with $N - 1$ degrees of freedom.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.**

a. Decide the significance level and whether to use a one-tailed or a two-tailed test.

b. Look up the appropriate cutoff in a *t* table.

❹ **Determine your sample's score on the comparison distribution:** $t = (M - \text{Population } M)/S_M$

❺ **Decide whether to reject the null hypothesis:** Compare the scores from Steps ❸ and ❹.

Summary of Steps for a *t* Test for Dependent Means

Table 5 summarizes the steps for a *t* test for dependent means.[4]

A Second Example of a *t* Test for Dependent Means

Here is another example. A team of researchers examined the brain systems involved in human romantic love (Aron, Fisher, Mashek, Strong, & Brown, 2005). One issue was whether romantic love engages a part of the brain called the caudate (a brain structure that is engaged when people win money in an experimental task, are given cocaine, and other such "rewards"). Thus, the researchers recruited people who had very recently fallen "madly in love." (For example, to be in the study participants had to think about their partner at least 80% of their waking hours.) Participants brought a picture of their beloved with them, plus a picture of a familiar, neutral person of the same age and sex as their beloved. Participants then went in to the functional magnetic resonance imaging (fMRI) machine and their brain was scanned while they looked at the two pictures—30 seconds at the neutral person's picture, 30 seconds at their beloved, 30 seconds at the neutral person, and so forth.

[4]The usual steps of carrying out a *t* test for dependent means can be somewhat combined into computational formulas for *S* and *t* based on difference scores (using the symbol *D*). For purposes of learning the ideas, we strongly recommend that you use the regular procedures as we have discussed them in this chapter when doing the practice problems. In a real research situation, the figuring is usually done by computer (see the "Using SPSS" section at the end of this chapter). However, if you ever have to do a *t* test for dependent means for an actual research study by hand (without a computer), you may find these formulas useful.

D in the formulas below is for difference score:

$$S = \sqrt{\frac{\Sigma D^2 - [(\Sigma D)^2/N]}{N - 1}} \tag{7}$$

$$t = \frac{(\Sigma D)/N}{S/\sqrt{N}} \tag{8}$$

Table 6 *t* Test for a Study of Romantic Love and Brain Activation in Part of the Caudate

Student	Brain Activation: Viewing Beloved's Photo	Brain Activation: Viewing Neutral Photo	Difference (Beloved − Neutral)	Deviation (Difference − *M*)	Squared Deviation
1	1487.8	1487.3	.5	−.700	.490
2	1329.4	1328.1	1.3	.100	.010
3	1407.9	1405.9	2.0	.800	.640
4	1236.1	1234.6	1.5	.300	.090
5	1299.8	1298.2	1.6	.400	.160
6	1447.2	1445.2	2.0	.800	.640
7	1354.1	1354.2	−.1	−1.300	1.690
8	1204.6	1203.7	.9	−.300	.090
9	1322.3	1321.4	.9	−.300	.090
10	1388.5	1387.1	1.4	.200	.040
Σ:	13477.7	13465.7	12.0		3.940

For difference scores:

$M = 12.0/10 = 1.200.$

Population $M = 0$ (assumed as a no-change baseline of comparison).

$S^2 = [\Sigma(X - M)^2]/df = 3.940/(10 - 1) = 3.940/9 = .438.$

$S_M^2 = S^2/N = .438/10 = .044.$

$S_M = \sqrt{S_M^2} = \sqrt{.044} = .210.$

t with $df = 9$ needed for 5% level, one-tailed $= 1.833.$

$t = (M - \text{Population } M)/S_M = (1.200 - 0)/.210 = 5.71.$

Decision: Reject the null hypothesis.

Source: Data based on Aron et al. (2005).

Table 6 shows average brain activations (mean fMRI scanner values) in the caudate area of interest during the two kinds of pictures. (We have simplified the example for teaching purposes, including using only 10 participants when the actual study had 17.) It also shows the figuring of the difference scores and all the other figuring for the *t* test for dependent means. Figure 6 shows the distributions involved. Here are the steps of hypothesis testing:

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are two populations:

Population 1: Individuals like those tested in this study.
Population 2: Individuals whose brain activation in the caudate area of interest is the same whether looking at a picture of their beloved or a picture of a familiar, neutral person.

The research hypothesis is that Population 1's mean difference score (brain activation when viewing the beloved's picture minus brain activation when viewing the neutral person's picture) is greater than Population 2's mean difference score (of no difference). That is, the research hypothesis is that brain activation in the caudate area of interest is greater when viewing the beloved person's picture than when viewing the neutral person's picture. The null hypothesis is that Population 1's mean difference score is not greater than Population 2's. That is, the null hypothesis is that brain activation in the caudate area of interest is not greater when viewing the beloved person's picture than when viewing the neutral person's picture.

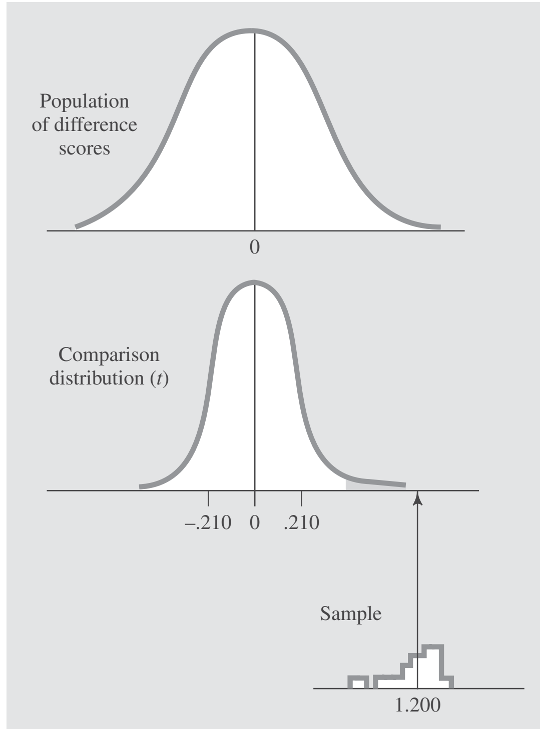


Figure 6 Distributions for the example of romantic love and brain activation in part of the caudate.

❷ **Determine the characteristics of the comparison distribution.**

a. Make each person's two scores into a difference score. This is shown in the column labeled "Difference" in Table 6. You do all the remaining steps using these difference scores.
b. Figure the mean of the difference scores. The sum of the difference scores (12.0) divided by the number of difference scores (10) gives a mean of the difference scores of 1.200. So, $M = 1.200$.
c. Assume a mean of the distribution of means of difference scores of 0: Population $M = 0$.
d. The standard deviation of the distribution of means of difference scores is figured as follows:
   - Ⓐ Figure the estimated population variance of difference scores: $S^2 = [\Sigma (X - M)^2]/df = 3.940/(10 - 1) = .438$.
   - Ⓑ Figure the variance of the distribution of means of difference scores: $S_M^2 = S^2/N = .438/10 = .044$.
   - Ⓒ Figure the standard deviation of the distribution of means of difference scores: $S_M = \sqrt{S_M^2} = \sqrt{.044} = .210$.
e. The shape is a t distribution with $df = N - 1$. Therefore, the comparison distribution is a t distribution for 9 degrees of freedom. It is a t distribution because we figured its variance based on an estimated population variance. It has 9 degrees of freedom because there were 9 degrees of freedom in the estimate of the population variance.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.**

a. We will use the standard .05 significance level. This is a one-tailed test because the researchers were interested only in a specific direction of difference.

b. Using the .05 significance level with 9 degrees of freedom, Table 2 of the appendix "Tables" shows a cutoff *t* of 1.833. In Table 6, the difference score is figured as brain activation when viewing the beloved's picture minus brain activation when viewing the neutral person's picture. Thus, the research hypothesis predicts a positive difference score, which means that our cutoff is $+1.833$.

❹ **Determine your sample's score on the comparison distribution.** $t = (M - \text{Population } M)/S_M = (1.200 - 0)/.210 = 5.71$. The sample's mean difference of 1.200 is 5.71 standard deviations (of .251 each) above the mean of 0 on the distribution of means of difference scores.

❺ **Decide whether to reject the null hypothesis.** The sample's *t* score of 5.71 is more extreme than the cutoff *t* of 1.833. You can reject the null hypothesis. Brain activation in the caudate area of interest is greater when viewing a beloved's picture than when viewing a neutral person's picture. The results of this study are not limited to North Americans. In 2007 the study was replicated, with virtually identical results, in Beijing with Chinese students who were intensely in love (Xu et al., 2007).

Review and Comparison of *Z* Test, *t* Test for a Single Sample, and *t* Test for Dependent Means

In this chapter you have learned about the *t* test for a single sample and the *t* test for dependent means. Table 7 provides a review and comparison of the *Z* test, the *t* test for a single sample, and the *t* test for dependent means.

Table 7 Review of the *Z* Test, the *t* Test for a Single Sample, and the *t* Test for Dependent Means

	Type of Test		
Features	***Z* Test**	***t* Test for a Single Sample**	***t* Test for Dependent Means**
Population variance is known	Yes	No	No
Population mean is known	Yes	Yes	No
Number of scores for each participant	1	1	2
Shape of comparison distribution	*Z* distribution	*t* distribution	*t* distribution
Formula for degrees of freedom	Not applicable	$df = N - 1$	$df = N - 1$
Formula	$Z = (M - \text{Population } M)/\text{Population } SD_M$	$t = (M - \text{Population } M)/S_M$	$t = (M - \text{Population } M)/S_M$

TIP FOR SUCCESS

We recommend that you spend some time carefully going through Table 7. Test your understanding of the tests by covering up portions of the table and trying to recall the hidden information.

How are you doing?

1. Describe the situation in which you would use a *t* test for dependent means.
2. When doing a *t* test for dependent means, what do you do with the two scores you have for each participant?
3. In a *t* test for dependent means, (a) what is usually considered to be the mean of the "known" population (Population 2). (b) Why?
4. Five individuals are tested before and after an experimental procedure; their scores are given in the following table. Test the hypothesis that there is no

change, using the .05 significance level. (a) Use the steps of hypothesis testing and (b) sketch the distributions involved.

Person	Before	After
1	20	30
2	30	50
3	20	10
4	40	30
5	30	40

5. What about the research situation makes the difference in whether you should carry out a *Z* test or a *t* test for a single sample?

6. What about the research situation makes the difference in whether you should carry out a *t* test for a single sample or a *t* test for dependent means?

Answers

1. A *t* test for dependent means is used when you are doing hypothesis testing and you have two scores for each participant (such as a before-score and an after-score) and the population variance is unknown.

2. When doing a *t* test for dependent means, subtract one score from the other to create a difference (or change) score for each person. The *t* test is then done with these difference (or change) scores.

3. (a) The mean of the "known" population (Population 2) is 0. (b) You are comparing your sample to a situation in which there is no difference—a population of difference scores in which the average difference is 0.

4. Steps of hypothesis testing (all figuring is shown in Table 8):

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are two populations:

Population 1: People like those tested before and after the experimental procedure.
Population 2: People whose scores are the same before and after the experimental procedure.

The research hypothesis is that Population 1's mean change scores (after minus before) are different from Population 2's. The null hypothesis is that Population 1's change is the same as Population 2's.

❷ **Determine the characteristics of the comparison distribution.** The mean of the distribution of means of difference scores (the comparison distribution) is 0. The standard deviation of the distribution of means of difference scores is 6. It is a *t* distribution with 4 degrees of freedom.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** For a two-tailed test at the .05 level, the cutoff sample *t* scores are 2.776 and −2.776.

❹ **Determine your sample's score on the comparison distribution.** $t = (4 - 0)/6 = .67$.

❺ **Decide whether to reject the null hypothesis.** The sample's *t* score of .67 is not more extreme than the cutoff *t* of ± 2.776. Therefore, do *not* reject the null hypothesis.

The distributions are shown in Figure 7.

Table 8 Figuring for Answer to "How are you doing?" Question 4

	Score		Difference	Deviation	Squared
Person	Before	After	(After – Before)	(Difference − M)	Deviation
1	20	30	10	6	36
2	30	50	20	16	256
3	20	10	−10	−14	196
4	40	30	−10	−14	196
5	30	40	10	6	36
Σ:	140	160	20		720

For difference scores:

$M = 20/5 = 4.$

Population $M = 0.$

$S^2 = [\Sigma(X - M)^2]/df = 720/(5 - 1) = 720/4 = 180.$

$S_M^2 = S^2/N = 180/5 = 36.$

$S_M = \sqrt{S_M^2} = \sqrt{36} = 6.$

t for $df = 4$ needed for 5% significance level, two-tailed $= \pm 2.776.$

$t = (M - \text{Population } M)/S_M = (4 - 0)/6 = .67$

Decision: Do not reject the null hypothesis.

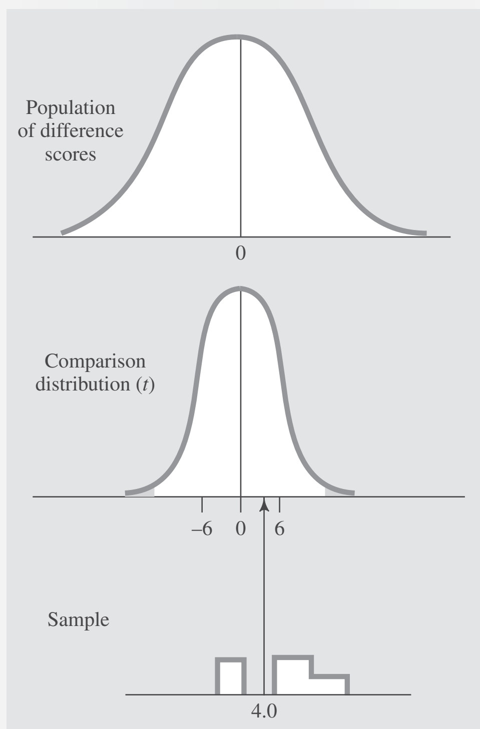


Figure 7 Distributions for answer to "How are you doing?" question 4.

5. As shown in Table 7, whether the population variance is known determines whether you should carry out a *Z* test or a *t* test for a single sample. You use a *Z* test when the population variance is known and you use the *t* test for a single sample when it is not known.
6. As shown in Table 7, when the population variance is not known, you do a *t* test. It is a *t* test for a single sample if you have one score for each participant; you use the *t* test for dependent means when you do not know the population mean and there are two scores for each participant.

Assumptions of the *t* Test for a Single Sample and *t* Test for Dependent Means

As we have seen, when you are using an estimated population variance, the comparison distribution is a *t* distribution. However, the comparison distribution (a distribution of means) will be exactly a *t* distribution only if the distribution of the population of *individuals* follows a normal curve. Otherwise, the comparison distribution will follow some other (usually unknown) shape.

Thus, strictly speaking, a normal population is a requirement within the logic and mathematics of the *t* test. A requirement like this for a hypothesis-testing procedure is called an **assumption.** That is, a normal population distribution is one assumption of the *t* test. You can think of an assumption as a requirement that you must meet for the results of a hypothesis-testing procedure to be accurate. In this case, the effect of this assumption of a normal population is that if the population distribution is not normal, the comparison distribution will be some unknown shape other than a *t* distribution—and thus the cutoffs on the *t* table will be incorrect.

Unfortunately, when you do a *t* test, you don't know whether the population is normal. This is because when doing a *t* test, usually all you have are the scores in your sample. Fortunately, however, distributions in the behavioral and social sciences quite often approximate a normal curve. (This applies to distributions of difference scores too.) Also, statisticians have found that in practice, you get reasonably accurate results with *t* tests even when the population is rather far from normal. The only very common situation in which using a *t* test for dependent means is likely to give a seriously distorted result is when you are using a one-tailed test and the population is highly skewed (that is, it is very asymmetrical, with a much longer tail on one side than the other). Thus, you need to be cautious about your conclusions when doing a one-tailed test if the sample of difference scores is highly skewed, suggesting that the population it comes from is also highly skewed.

assumption A condition, such as a population's having a normal distribution, required for carrying out a particular hypothesis-testing procedure; a part of the mathematical foundation for the accuracy of the tables used in determining cutoff values.

Effect Size and Power for the *t* Test for Dependent Means[5]

Effect Size

The estimated effect size for a study using a *t* test for dependent means is the mean of the difference scores divided by the estimated standard deviation of the population of difference scores. In terms of a formula,

The estimated effect size for a study using a *t* test for dependent means is the mean of the difference scores divided by the estimated standard deviation of the population of difference scores.

$$\text{Estimated Effect Size} = M/S \qquad (9)$$

[5]Note that effect size and power for the *t* test for a single sample are determined in the same way as for the *t* test for dependent means.

M is the mean of the difference scores and *S* is the estimated standard deviation of the population of individual difference scores.

The conventions for effect size may already be familiar to you: A small effect size is .20, a medium effect size is .50, and a large effect size is .80.

Consider our first example of a *t* test for dependent means, the study of husbands' change in communication quality. In that study, the mean of the difference scores was −12.05. The estimated population standard deviation of the difference scores would be 12.41. That is, we figured the estimated variance of the difference scores (S^2) to be 154.05 and the square root of 154.05 is 12.41. Therefore, the effect size is −12.05 divided by 12.41, which is −.97. This is a very large effect size. (The negative sign for the effect size means that the large effect was a decrease in communication scores.)

Power

Power for a *t* test for dependent means can be determined using a power software package, an Internet power calculator, or a power table. Table 9 gives the approximate power at the .05 significance level for small, medium, and large effect sizes and one-tailed and two-tailed tests.

Suppose a researcher plans a study using the .05 significance level, two-tailed, with 20 participants. Based on previous related research, the researcher predicts an effect size of about .50 (a medium effect size). The table shows the study would have a power of .59. This means that, if the research hypothesis is true and has a medium effect size, there is a 59% chance that this study will come out significant.

TIP FOR SUCCESS

Power can be expressed as a probability (such as .71) or as a percentage (such as 71%). Power is expressed as a probability in Table 9.

The power table (Table 9) is also useful when you are reading about a nonsignificant result in a published study. Suppose that a study using a *t* test for dependent means has a nonsignificant result. The study tested significance at the .05 level, was two-tailed, and had 10 participants. Should you conclude that there is in fact no difference at all between the populations? Probably not. Even assuming a medium effect size, Table 9 shows that there is only a 32% chance of getting a significant result in this study.

Consider another study that was not significant. This study also used the .05 significance level, one-tailed. This study had 100 research participants. Table 9 tells you that there would be a 63% chance of the study's coming out significant if there were even a true small effect size in the population. If there were a medium effect size in the population, the table indicates that there is almost a 100% chance that this study would have come out significant. In this study with 100 participants, you could conclude from the results of this study that in the population there is probably at most a small difference.

To keep Table 9 simple, we have given power figures for only a few different numbers of participants (10, 20, 30, 40, 50, and 100). This should be adequate for the kinds of rough evaluations you need to make when evaluating results of research articles.[6]

[6]Cohen (1988, pp. 239) provides more detailed tables in terms of numbers of participants, levels of effect size, and significance levels. If you use his tables, note that the *d* referred to is actually based on a *t* test for independent means. To use these tables for a *t* test for dependent means, first multiply your effect size by 1.4. For example, if your effect size is .30, for purposes of using Cohen's tables, you would consider it to be .42 (that is, $.30 \times 1.4 = .42$). The only other difference from our table is that Cohen describes the significance level by the letter *a* (for "alpha level"), with a subscript of either 1 or 2, referring to a one-tailed or two-tailed test. For example, a table that refers to "$a_1 = .05$" at the top means that this is the table for $p < .05$, one-tailed.

Table 9 Approximate Power for Studies Using the *t* Test for Dependent Means for Testing Hypothesis at the .05 Significance Level

Difference Scores in Sample (*N*)	Effect Size		
	Small (.20)	Medium (.50)	Large (.80)
One-tailed test			
10	.15	.46	.78
20	.22	.71	.96
30	.29	.86	*
40	.35	.93	*
50	.40	.97	*
100	.63	*	*
Two-tailed test			
10	.09	.32	.66
20	.14	.59	.93
30	.19	.77	.99
40	.24	.88	*
50	.29	.94	*
100	.55	*	*

*Power is nearly 1.

Planning Sample Size

Table 10 gives the approximate number of participants needed to have 80% power for a planned study. (Eighty percent is a common figure used by researchers for the minimum power to make a study worth doing.) The table gives the number of participants needed based on predicted small, medium, and large effect sizes, using one- and two-tailed tests, for the .05 significance levels. Suppose you plan a study in which you expect a large effect size and you use the .05 significance level, two-tailed. The table shows that you would need only 14 participants to have 80% power. On the other hand, a study using the same significance level, also two-tailed, but in which you expect only a small effect size, would need 196 participants for 80% power.[7]

The Power of Studies Using the *t* Test for Dependent Means

Studies using difference scores (that is, studies using a repeated-measures design) often have much larger effect sizes for the same amount of expected difference between means than have other kinds of research designs. That is, testing each of a group of participants twice (once under one condition and once under a different condition) usually produces a high-power type of study. In particular, this kind of study gives more power than dividing the participants up into two groups and testing each group once (one group tested under one condition and the other tested under the other condition). In fact, studies using difference scores usually have even more power than those in which you have twice as many participants, but tested each only once.

[7]More detailed tables, giving needed numbers of participants for levels of power other than 80% (and also for effect sizes other than .20, .50, and .80 and for other significance levels), are provided in Cohen (1988, pp. 54–55). However, see footnote 6 in this chapter about using Cohen's tables for a *t* test for dependent means.

Table 10 Approximate Number of Research Participants Needed for 80% Power for the *t* Test for Dependent Means in Testing Hypotheses at the .05 Significance Level

	Effect Size		
	Small (.20)	**Medium (.50)**	**Large (.80)**
One-tailed	156	26	12
Two-tailed	196	33	14

Why do repeated-measures designs have so much power? The reason is that the standard deviation of difference scores is usually quite low. (The standard deviation of difference scores is what you divide by to get the effect size when using difference scores.) This produces a larger effect size, which increases the power. In a repeated measures design, the only variation is in the difference scores. Variation among participants on each testing's scores are not part of the variation involved in the analysis. This is because difference scores are all comparing participants to themselves. The effect of all this is that studies using difference scores often have quite large effect sizes (and thus high power) even with a small number of people in the study.

However, although it has advantages from the point of view of power, the kind of repeated-measures study discussed in this chapter (testing a group of people before and after an experimental procedure, without any kind of control group that does not go through the procedure) often has disadvantages from the point of view of the meaning of the results. For example, consider a study where people are tested before and after some experimental procedure. Even if such a study produces a significant difference, it leaves many alternative explanations for that difference. For example, the research participants might have improved during that period anyway, or perhaps other events happened in between, or the participants not getting benefits may have dropped out. It is even possible that the initial test itself caused changes. The limitations of this kind of research are discussed in detail in research methods textbooks.

How are you doing?

1. (a) What is an assumption in hypothesis testing? (b) Describe a specific assumption for a *t* test for dependent means. (c) What is the effect of violating this assumption? (d) When is the *t* test for dependent means likely to give a very distorted result?
2. (a) Write the formula for estimated effect size in a *t* test for dependent means situation, and (b) describe each of its terms.
3. You are planning a study in which you predict an effect size of .50. You plan to test significance using a *t* test for dependent means, one-tailed, with an alpha of .05. (a) What is the power of this study if you carry it out with 20 participants? (b) How many participants would you need to have 80% power?
4. (a) Why do repeated-measures designs have so much power? (b) What is the main disadvantage of the kind of repeated-measures study discussed in this chapter?

Answers

1. (a) An assumption is a requirement that you must meet for the results of the hypothesis-testing procedure to be accurate.
(b) The population of individuals' difference scores is assumed to be a normal distribution.

(c) The significance level cutoff from the *t* table is not accurate.
(d) The *t* test for dependent means is likely to give a very distorted result when doing a one-tailed test and the population distribution is highly skewed.

2. (a) Estimated Effect Size $= M/S$
(b) *M* is the mean of the difference scores. *S* is the estimated standard deviation of the population of individual difference scores.

3. (a) Power of this study is .71. (b) Number of participant for 80% power: 26.

4. (a) Repeated-measures designs have so much power because the standard deviation of the difference scores is usually quite low (which makes the effect size relatively high, thereby increasing power).
(b) The main disadvantage of the kind of repeated-measures study discussed in this chapter is that even if a study produces a significant difference, there are alternative explanations for that difference.

Single-Sample *t* Tests and Dependent Means *t* Tests in Research Articles

Research articles usually describe *t* tests in a fairly standard format that includes the degrees of freedom, the *t* score, and the significance level. For example, $t(24) = 2.80$, $p < .05$ tells you that the researcher used a *t* test with 24 degrees of freedom, found a *t* score of 2.80, and the result was significant at the .05 level. Whether a one- or two-tailed test was used may also be noted. (If not, assume that it was two-tailed.) Usually the means, and sometimes the standard deviations, are given for each testing. Rarely does an article report the standard deviation of the difference scores.

Had our student in the dormitory example reported the results in a research article, she would have written something like this: "The sample from my dormitory studied a mean of 21 hours $(SD = 6.80)$. Based on a *t* test for a single sample, this was significantly different from the known mean of 17 for the university as a whole, $t(15) = 2.35$, $p < .05$, one-tailed."

As we noted earlier, behavioral and social scientists only occasionally use the *t* test for a single sample. We introduced it mainly as a stepping stone to the more widely used *t* test for dependent means. Olthoff (1989) might have reported the results of the *t* test for dependent means in his study of husbands' communication quality as follows: "There was a significant decline in communication quality, dropping from a mean of 116.32 before marriage to a mean of 104.26 after marriage, $t(18) = 2.76$, $p < .05$, two-tailed."

As another example, Rashotte and Webster (2005) carried out a study about people's general expectations about the abilities of men and women. In the study, the researchers showed 174 college students photos of women and men (referred to as the female and male targets, respectively). The students rated the person in each photo in terms of that person's general abilities (for example, in terms of the person's intelligence, abstract abilities, capability at most tasks, etc.). For each participant, these ratings were combined to create a measure of the perceived status of the female targets and the male targets. The researchers then compared the status ratings given for the female targets and male targets. Since each participant in the study rated *both* the female and the male targets, the researchers compared the status ratings assigned to the female and male targets using a *t* test for dependent means. Table 11 shows the results. The row titled "Whole sample $(N = 174)$" gives the result of the *t* test for all 174 participants and shows that the status rating assigned to the male targets was significantly higher

Table 11 Status Scale: Mean (and *SE*) General Expectations for Female and Male Targets

Respondents	Mean Score (*SE*) Female Target	Male Target	M − F Target Difference	*t* (1-Tailed *p*)
Whole sample ($N = 174$)	5.60 (.06)	5.85 (.07)	.25	3.46 (<.001)
Female respondents ($N = 111$)	5.62 (.07)	5.84 (.081)	.22	2.62 (<.05)
Male respondents ($N = 63$)	5.57 (.10)	5.86 (.11)	.29	2.26 (<.05)

Source: Rashotte, L.S., & Webster, M., Jr. (2005). Gender status beliefs. *Social Science Research, 34,* 618–633.

than the rating assigned to the female targets $(t = 3.46, p < .001)$. As shown in the table, the researchers also conducted two additional *t* tests to see if this effect was the same among the female participants and the male participants. The results showed that both the female and the male participants assigned higher ratings to the male targets. As with this study, the results of dependent means *t* tests are sometimes shown in a table. Asterisks will often be used to indicate the level of significance, with a note at the bottom of the table listing the corresponding significance level (although this wasn't the case in Table 11). And sometimes the *t* value itself is not given, just the asterisks indicating the significance level.

Often the results of a *t* test for dependent means will be given in the text and not in a table. For example, Song and Schwarz (2009) asked 20 university students to rate the harm of food additives that were either easy to pronounce (e.g., Magnalroxate) or hard to pronounce (e.g., Hnegripitrom). (In case you are wondering, these are not real food additives!) Each student rated all of the food additives using a scale from $1 =$ *very safe* to $7 =$ *very harmful*. Here is how Song and Schwarz reported the results of the *t* test for dependent means: "As predicted, participants . . . rated substances with hard-to-pronounce names $(M = 4.12, SD = 0.78)$ as more harmful than substances with easy-to-pronounce names $(M = 3.7, SD = 0.74), t(19) = 2.41, p < .03$" (p. 136).

Learning Aids

1. You use the standard five steps of hypothesis testing even when you don't know the population variance. However, in this situation you have to estimate the population variance from the scores in the sample, using a formula that divides the sum of squared deviation scores by the degrees of freedom $(df = N - 1)$.
2. When the population variance is estimated, the comparison distribution of means is a *t* distribution (with cutoffs given in a *t* table). A *t* distribution has slightly heavier tails than a normal curve (just how much heavier depends on how few the degrees of freedom are). Also, in this situation, a sample's number of standard deviations from the mean of the comparison distribution is called a *t* score.
3. You use a *t* test for a single sample when a sample's mean is being compared to a known population mean and the population variance is unknown.
4. You use a *t* test for dependent means in studies where each participant has two scores, such as a before-score and an after-score, or a score in each of two experimental conditions. In this *t* test, you first figure a difference score for each participant, then go through the usual five steps of hypothesis testing with the

modifications described in summary points 1 and 2 and making Population 2 a population of difference scores with a mean of 0 (no difference).

5. An assumption of the *t* test is that the population distribution is a normal curve. However, even when it is not, the *t* test is usually fairly accurate.
6. The effect size of a study using a *t* test for dependent means is the mean of the difference scores divided by the standard deviation of the difference scores. Power and needed sample size for 80% power can be looked up using power software packages, an Internet power calculator, or special tables.
7. The power of studies using difference scores is usually much higher than that of studies using other designs with the same number of participants. However, research using a single group tested before and after some intervening event, without a control group, allows for alternative explanations of any observed changes.
8. *t* tests are reported in research articles using a standard format. For example, "$t(24) = 2.80$, $p < .05$."

Key Terms

t tests
t test for a single sample
biased estimate
unbiased estimate of the population variance (S^2)
degrees of freedom (*df*)
t distribution
t table
t score
repeated-measures design
t test for dependent means
difference scores)
assumption

t Test for a Single Sample

Eight participants are tested after being given an experimental procedure. Their scores are 14, 8, 6, 5, 13, 10, 10, and 6. The population (of people not given this procedure) is normally distributed with a mean of 6. Using the .05 level, two-tailed, does the experimental procedure make a difference? (a) Use the five steps of hypothesis testing and (b) sketch the distributions involved.

Answer

(a) Steps of hypothesis testing:

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are two populations:

Population 1: People who are given the experimental procedure.
Population 2: The general population.

The research hypothesis is that Population 1 will score differently than Population 2. The null hypothesis is that Population 1 will score the same as Population 2.

❷ **Determine the characteristics of the comparison distribution.** The mean of the distribution of means is 6 (the known population mean). To figure the estimated population variance, you first need to figure the sample mean, which is $(14 + 8 + 6 + 5 + 13 + 10 + 10 + 6)/8 = 72/8 = 9$. The estimated population variance is $S^2 = [\Sigma (X - M)^2]/df = 78/7 = 11.14$; the variance of the distribution of means is $S_M^2 = S^2/N = 11.14/8 = 1.39$. The standard

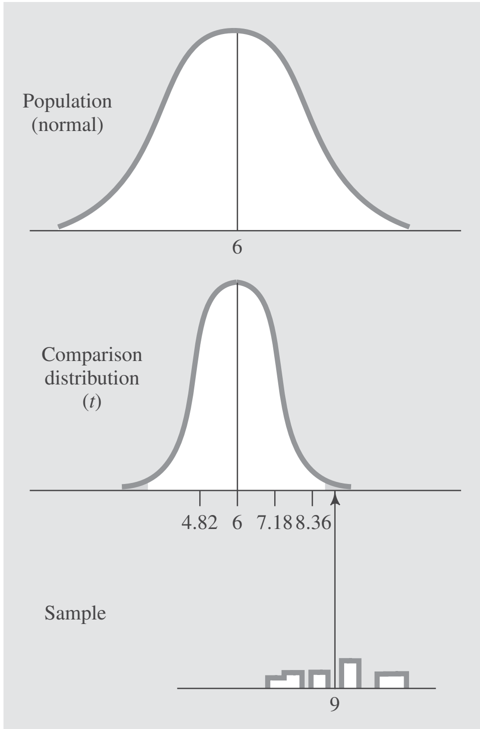


Figure 8 Distributions for answer to Example Worked-Out Problem for *t* test for a single sample.

deviation of the distribution of means is $S_M = \sqrt{S_M^2} = \sqrt{1.39} = 1.18$. Its shape will be a *t* distribution for $df = 7$.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** From Table 2 of the appendix "Tables," the cutoffs for a two-tailed *t* test at the .05 level for $df = 7$ are 2.365 and -2.365.

❹ **Determine your sample's score on the comparison distribution.** $t = (M - \text{Population } M)/S_M = (9 - 6)/1.18 = 3/1.18 = 2.54$.

❺ **Decide whether to reject the null hypothesis.** The *t* of 2.54 is more extreme than the needed *t* of 2.365. Therefore, reject the null hypothesis; the research hypothesis is supported. The experimental procedure does make a difference.

(b) Sketches of distributions are shown in Figure 8.

t Test for Dependent Means

A researcher tests 10 individuals before and after an experimental procedure. The results are shown in the table to the right.

Test the hypothesis that there is an increase in scores, using the .05 significance level. (a) Use the five steps of hypothesis testing and (b) sketch the distributions involved.

Participant	Before	After
1	10.4	10.8
2	12.6	12.1
3	11.2	12.1
4	10.9	11.4
5	14.3	13.9
6	13.2	13.5
7	9.7	10.9
8	11.5	11.5
9	10.8	10.4
10	13.1	12.5

Answer

(a) Table 12 shows the results, including the figuring of difference scores and all the other figuring for the *t* test for dependent means. Here are the steps of hypothesis testing:

Table 12 Figuring for Answer to Example Worked-Out Problem for *t* Test for Dependent Means

Participant	Score		Difference (After − Before)	Deviation (Difference − *M*)	Squared Deviation
	Before	*After*			
1	10.4	10.8	.4	.260	.068
2	12.6	12.1	−.5	−.640	.410
3	11.2	12.1	.9	.760	.578
4	10.9	11.4	.5	.360	.130
5	14.3	13.9	−.4	−.540	.292
6	13.2	13.5	.3	.160	.026
7	9.7	10.9	1.2	1.060	1.124
8	11.5	11.5	0.0	−.140	.020
9	10.8	10.4	−.4	−.540	.292
10	13.1	12.5	−.6	−.740	.548
Σ	117.7	119.1	1.4		3.488

For difference scores:
$M = 1.4/10 = .140.$
Population $M = 0.$
$S^2 = [\Sigma(X - M)^2]/df = 3.488/(10 - 1) = 3.488/9 = .388.$
$S_M^2 = S^2/N = .388/10 = .039.$
$S_M = \sqrt{S_M^2} = \sqrt{.039} = .197.$
t for $df = 9$ needed for 5% significance level, one-tailed $= 1.833.$
$t = (M - \text{Population } M)/S_M = (.140 - 0)/.197 = .71.$
Decision: Do not reject the null hypothesis.

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are two populations:

Population 1: People like those who are given the experimental procedure.
Population 2: People who show no change from before to after.

The research hypothesis is that Population 1's mean difference score (figured using "after" scores minus "before" scores) is greater than Population 2's. The null hypothesis is that Population 1's mean difference (after minus before) is not greater than Population 2's.

❷ **Determine the characteristics of the comparison distribution.** Its population mean is 0 difference. The estimated population variance of difference scores is shown in Table 12 to be .388. The standard deviation of the distribution of means of difference scores, S_M, is .197. Therefore, the comparison distribution has a mean of 0 and a standard deviation of .197. It will be a t distribution for $df = 9$.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** For a one-tailed test at the .05 level with $df = 9$, the cutoff is 1.833. (The cutoff is positive because the research hypothesis is that Population 1's mean difference score will be *greater* than Population 2's.)

❹ **Determine your sample's score on the comparison distribution.** The sample's mean difference of .140 is .71 standard deviations (of .197 each) on the distribution of means above that distribution's mean of 0. That is, $t = (M - \text{Population } M)/S_M = (.140 - 0)/.197 = .71.$

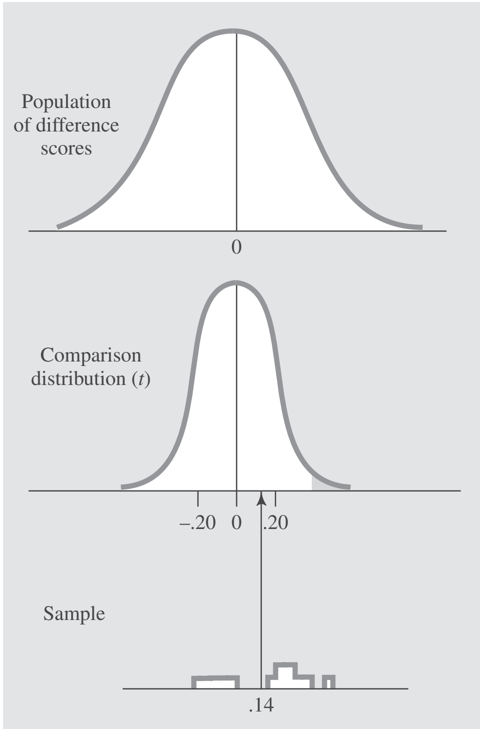


Figure 9 Distributions for answer to Example Worked-Out Problem for *t* test for dependent means.

❺ **Decide whether to reject the null hypothesis.** The sample's *t* of .71 is less extreme than the needed *t* of 1.833. Thus, you cannot reject the null hypothesis. The study is inconclusive.

(b) Sketches of distributions are shown in Figure 9.

Outline for Writing Essays for a *t* Test for a Single Sample

1. Describe the core logic of hypothesis testing in this situation. Be sure to mention that the *t* test for a single sample is used for hypothesis testing when you have scores for a sample of individuals and you want to compare the mean of this sample to a population for which the mean is known but the variance is unknown. Be sure to explain the meaning of the research hypothesis and the null hypothesis in this situation.
2. Outline the logic of estimating the population variance from the sample scores. Explain the idea of biased and unbiased estimates of the population variance, and describe the formula for estimating the population variance and why it is different from the ordinary variance formula.
3. Describe the comparison distribution (the *t* distribution) that is used with a *t* test for a single sample, noting how it is different from a normal curve and why. Explain why a *t* distribution (as opposed to the normal curve) is used as the comparison distribution.
4. Describe the logic and process for determining the cutoff sample score(s) on the comparison distribution at which the null hypothesis should be rejected.
5. Describe why and how you figure the *t* score of the sample mean on the comparison distribution.

6. Explain how and why the scores from Steps ❸ and ❹ of the hypothesis-testing process are compared. Explain the meaning of the result of this comparison with regard to the specific research and null hypotheses being tested.

Outline for Writing Essays for a *t* Test for Dependent Means

1. Describe the core logic of hypothesis testing in this situation. Be sure to mention that the *t* test for dependent means is used for hypothesis testing when you have two scores from each person in your sample. Be sure to explain the meaning of the research hypothesis and the null hypothesis in this situation. Explain the logic and procedure for creating difference scores.
2. Explain why you use 0 as the mean for the comparison distribution.
3. Outline the logic of estimating the population variance of difference scores from the sample scores. Explain the idea of biased and unbiased estimates of the population variance, and describe the formula for estimating the population variance. Describe how to figure the standard deviation of the distribution of means of difference scores.
4. Describe the comparison distribution (the *t* distribution) that is used with a *t* test for dependent means. Explain why a *t* distribution (as opposed to the normal curve) is used as the comparison distribution.
5. Describe the logic and process for determining the cutoff sample score(s) on the comparison distribution at which the null hypothesis should be rejected.
6. Describe why and how you figure the *t* score of the sample mean on the comparison distribution.
7. Explain how and why the scores from Steps ❸ and ❹ of the hypothesis-testing process are compared. Explain the meaning of the result of this comparison with regard to the specific research and null hypotheses being tested.

These problems involve figuring. Most real-life statistics problems are done on a computer with special statistical software. Even if you have such software, do these problems by hand to ingrain the method in your mind. To learn how to use a computer to solve statistics problems like those in this chapter, refer to the "Using SPSS" section at the end of this chapter.

All data are fictional unless an actual citation is given.

Set I (for answers, see the end of this chapter)

1. In each of the following studies, a single sample's mean is being compared to a population with a known mean but an unknown variance. For each study, decide whether the result is significant. (Be sure to show all of your calculations.)

Study	Sample Size (N)	Population Mean	Estimated Population Variance (S^2)	Sample Mean (M)	Tails	Significance Level
(a)	64	12.40	9.00	11.00	1 (low predicted)	.05
(b)	49	1,006.35	317.91	1,009.72	2	.01
(c)	400	52.00	7.02	52.41	1 (high predicted)	.01

2. Suppose a candidate running for sheriff claims that she will reduce the average speed of emergency response to less than 30 minutes, which is thought to be the average response time with the current sheriff. There are no past records, so the actual standard deviation of such response times cannot be determined. Thanks to this campaign, she is elected sheriff, and careful records are now kept. The response times for the first month are 26, 30, 28, 29, 25, 28, 32, 35, 24, and 23 minutes.

   Using the .05 significance level, did she keep her promise? (a) Go through the five steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who has never taken a course in statistics.
3. A researcher tests five individuals who have seen paid political ads about a particular issue. These individuals take a multiple-choice test about the issue in which people in general (who know nothing about the issue) usually get 40 questions correct. The number correct for these five individuals was 48, 41, 40, 51, and 50.

   Using the .05 level of significance, two-tailed, do people who see the ads do better on this test? (a) Use the steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who is familiar with the *Z* test but is unfamiliar with *t* tests.
4. For each of the following studies using difference scores, test the significance using a *t* test for dependent means. Also, figure the estimated effect size for each study.

Study	Number of Difference Scores in Sample	Mean of Difference Scores in Sample	Estimated Population Variance of Difference Scores	Tails	Significance Level
(a)	20	1.7	8.29	1 (high predicted)	.05
(b)	164	2.3	414.53	2	.05
(c)	15	−2.2	4.00	1 (low predicted)	.01

5. A program to decrease littering was carried out in four cities in California's Central Valley starting in August 2009. The amount of litter in the streets (average pounds of litter collected per block per day) was measured during the July before the program was started and then the next July, after the program had been in effect for a year. The results were as follows:

City	July 2009	July 2010
Fresno	9	2
Merced	10	4
Bakersfield	8	9
Stockton	9	1

   Using the .01 level of significance, was there a significant decrease in the amount of litter? (a) Use the five steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who understands mean, standard deviation, and variance but knows nothing else about statistics.
6. A researcher assesses the level of a particular hormone in the blood in five patients before and after they begin taking a hormone treatment program. Results for the five are as follows:

Patient	Before	After
A	.20	.18
B	.16	.16
C	.24	.20
D	.22	.19
E	.17	.16

Using the .05 level of significance, was there a significant change in the level of this hormone? (a) Use the steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who understands the *t* test for a single sample but is unfamiliar with the *t* test for dependent means.

7. Figure the estimated effect size and indicate whether it is approximately small, medium, or large, for each of the following studies:

Study	Mean Change	*S*
(a)	20	32
(b)	5	10
(c)	.1	.4
(d)	100	500

8. What is the power of each of the following studies, using a *t* test for dependent means (based on the .05 significance level)?

Study	Effect Size	*N*	Tails
(a)	Small	20	One
(b)	Medium	20	One
(c)	Medium	30	One
(d)	Medium	30	Two
(e)	Large	30	Two

9. About how many participants are needed for 80% power in each of the following planned studies that will use a *t* test for dependent means with $p < .05$?

Study	Predicted Effect Size	Tails
(a)	Medium	Two
(b)	Large	One
(c)	Small	One

10. Weller and Weller (1997) conducted a study of the tendency for the menstrual cycles of women who live together (such as sisters) to become synchronized. For their statistical analysis, they compared scores on a measure of synchronization of pairs of sisters living together versus the degree of synchronization that would be expected by chance (lower scores mean more synchronization). Their key

results (reported in a table not reproduced here) were synchrony scores of 6.32 for the 30 roommate sister pairs in their sample compared to an expected synchrony score of 7.76; they then reported a *t* score of 2.27 for this difference. This result is statistically significant at $p < .05$. Explain this result to a person who is familiar with hypothesis testing with a known population variance but not with the *t* test for a single sample.

11. A researcher conducts a study of perceptual illusions under two different lighting conditions. Twenty participants were each tested under both of the two different conditions. The experimenter reported: "The mean number of effective illusions was 6.72 under the bright conditions and 6.85 under the dimly lit conditions, a difference that was not significant, $t(19) = 1.62$." Explain this result to a person who has never had a course in statistics. Be sure to use sketches of the distributions in your answer.

Set II

12. In each of the following studies, a single sample's mean is being compared to a population with a known mean but an unknown variance. For each study, decide whether the result is significant.

Study	Sample Size (*N*)	Population Mean	Estimated Population Standard Deviation (*S*)	Sample Mean (*M*)	Tails	Significance Level
(a)	16	100.31	2.00	100.98	1 (high predicted)	.05
(b)	16	.47	4.00	.00	2	.05
(c)	16	68.90	9.00	34.00	1 (low predicted)	.01

13. Evolutionary theories often emphasize that humans have adapted to their physical environment. One such theory hypothesizes that people should spontaneously follow a 24-hour cycle of sleeping and waking—even if they are not exposed to the usual pattern of sunlight. To test this notion, eight paid volunteers were placed (individually) in a room in which there was no light from the outside and no clocks or other indications of time. They could turn the lights on and off as they wished. After a month in the room, each individual tended to develop a steady cycle. Their cycles at the end of the study were as follows: 25, 27, 25, 23, 24, 25, 26, and 25.

    Using the .05 level of significance, what should we conclude about the theory that 24 hours is the natural cycle? (That is, does the average cycle length under these conditions differ significantly from 24 hours?) (a) Use the steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who has never taken a course in statistics.

14. In a particular country, it is known that college seniors report falling in love an average of 2.20 times during their college years. A sample of five seniors, originally from that country but who have spent their entire college career in the United States, were asked how many times they had fallen in love during their college years. Their numbers were 2, 3, 5, 5, and 2. Using the .05 significance level, do students like these who go to college in the United States fall in love more often than those from their country who go to college in their own country? (a) Use the steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who is familiar with the *Z* test but is not familiar with the *t* test for a single sample.

15. For each of the following studies using difference scores, test the significance using a *t* test for dependent means.

Study	Number of Difference Scores in Sample	Mean of Difference Scores in Sample	Estimated Population Variance of Difference Scores	Tails	Significance Level
(a)	10	3.8	50	1 (high predicted)	.05
(b)	100	3.8	50	1 (high predicted)	.05
(c)	100	1.9	50	1 (high predicted)	.05
(d)	100	1.9	50	2	.05
(e)	100	1.9	25	2	.05

16. Four individuals with high levels of cholesterol went on a special diet, avoiding high-cholesterol foods and taking special supplements. Their total cholesterol levels before and after the diet were as follows:

Participant	Before	After
J. K.	287	255
L. M. M.	305	269
A. K.	243	245
R. O. S.	309	247

Using the .05 level of significance, was there a significant change in cholesterol level? (a) Use the steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who has never taken a course in statistics.

17. Five people who were convicted of speeding were ordered by the court to attend a workshop. A special device put into their cars kept records of their speeds for 2 weeks before and after the workshop. The maximum speeds for each person during the 2 weeks before and the 2 weeks after the workshop follow.

Participant	Before	After
L. B.	65	58
J. K.	62	65
R. C.	60	56
R. T.	70	66
J. M.	68	60

Using the .05 significance level, should we conclude that people are likely to drive more slowly after such a workshop? (a) Use the steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who is familiar with hypothesis testing involving known populations but has never learned anything about *t* tests.

18. Five sophomores were given an English achievement test before and after receiving instruction in basic grammar. Their scores are shown below.

Student	Before	After
A	20	18
B	18	22
C	17	15
D	16	17
E	12	9

Is it reasonable to conclude that future students would show higher scores after instruction? Use the .05 significance level. (a) Use the steps of hypothesis testing. (b) Sketch the distributions involved. (c) Explain your answer to someone who understands mean, standard deviation, and variance but knows nothing else about statistics.

19. Figure the estimated effect size and indicate whether it is approximately small, medium, or large, for each of the following studies:

Study	Mean Change	*S*
(a)	8	30
(b)	8	10
(c)	16	30
(d)	16	10

20. What is the power of each of the following studies, using a *t* test for dependent means (based on the .05 significance level)?

Study	Effect Size	*N*	Tails
(a)	Small	50	Two
(b)	Medium	50	Two
(c)	Large	50	Two
(d)	Small	10	Two
(e)	Small	40	Two
(f)	Small	100	Two
(g)	Small	100	One

21. About how many participants are needed for 80% power in each of the following planned studies that will use a *t* test for dependent means with $p < .05$?

Study	Predicted Effect Size	Tails
(a)	Small	Two
(b)	Medium	One
(c)	Large	Two

22. A study compared union activity of employees in 10 plants during two different decades. The researchers reported "a significant increase in union activity, $t(9) = 3.28$, $p < .01$." Explain this result to a person who has never had a course in statistics. Be sure to use sketches of the distributions in your answer.

23. Baker and Moore (2008) surveyed 58 people when they first started using MySpace (the researchers called this "Time 0") and again 2 months later (referred to as "Time 1" by the researchers). At both time points, the participants completed a measure of social integration. Also, at Time 1, the researchers examined participants' MySpace profiles to see whether they had started blogging in the past 2 months. The researchers used *t* tests for dependent means (which they refer to as *paired sample t tests*) to test whether bloggers and nonbloggers differed in their level of social integration from Time 0 to Time 1. Here is how the researchers reported the results: "Paired samples *t* tests showed bloggers were significantly higher in social integration at Time 1 than Time 0, $t(30) = 3.19$, $p < .01$, ... with no difference for nonbloggers, $t(26) = -0.370$, $p > .05$." Explain these results to someone who is familiar with the *t* test for a single sample but not with the *t* test for dependent means.
24. Table 13 (reproduced from Table 4 of Larson, Dworkin, & Verma, 2001) shows ratings of various aspects of work and home life of 100 middle-class men in India who were fathers. Pick three rows of interest to you and explain the results to someone who is familiar with the mean, variance, and *Z* scores but knows nothing else about statistics.

Table 13 Comparison of Fathers' Mean Psychological States in the Job and Home Spheres ($N = 100$)

		Sphere		
Scale	**Range**	**Work**	**Home**	**Work vs. Home**
Important	0–9	5.98	5.06	6.86***
Attention	0–9	6.15	5.13	7.96***
Challenge	0–9	4.11	2.41	11.49***
Choice	0–9	4.28	4.74	−3.38***
Wish doing else	0–9	1.50	1.44	0.61
Hurried	0–3	1.80	1.39	3.21**
Social anxiety	0–3	0.81	0.64	3.17**
Affect	1–7	4.84	4.98	−2.64**
Social climate	1–7	5.64	5.95	4.17***

Note: Values for column 3 are *t* scores; $df = 99$ for all *t* tests.
**$p < .01$. ***$p < .001$.
Source: Larson, R., Dworkin, J., & Verma, S. (2001). Men's work and family lives in India: The daily organization of time and emotions. *Journal of Family Psychology, 15,* 206–224. Copyright © 2001 by the American Psychological Association.

The ✐ in the following steps indicates a mouse click. (We used SPSS version 17.0 for Windows to carry out these analyses. The steps and output may be slightly different for other versions of SPSS.)

t Test for a Single Sample

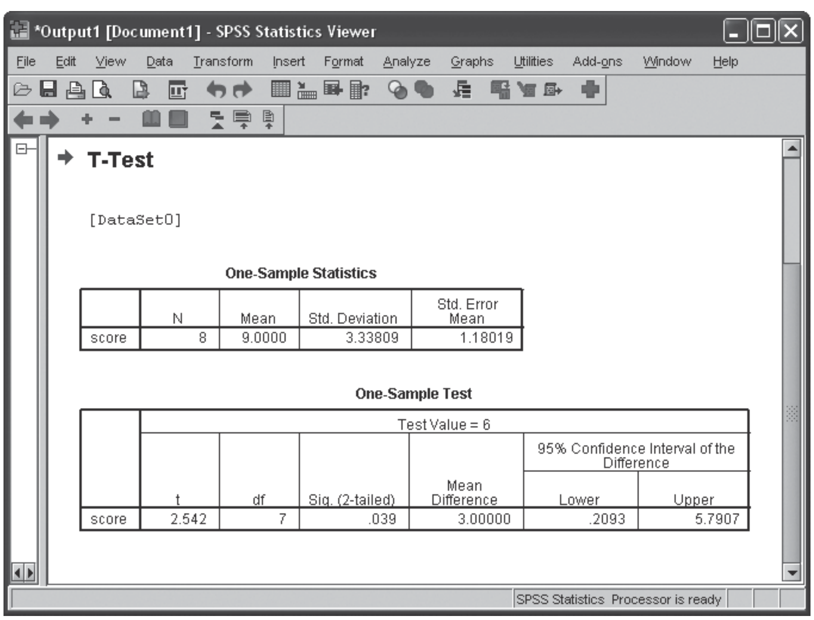
❶ Enter the scores from your distribution in one column of the data window.
❷ ✐ *Analyze.*
❸ ✐ *Compare means.*

❹ *One-sample T test* (this is the name SPSS uses for a *t* test for a single sample).
❺ on the variable for which you want to carry out the *t* test and then the arrow.
❻ Enter the population mean in the "Test Value" box.
❼ *OK*.

Practice the steps above by carrying out a single sample *t* test for the example worked-out problem for a *t* test for a single sample. In that example, eight participants were tested after an experimental procedure. Their scores were 14, 8, 6, 5, 13, 10, 10, and 6. The population of people not given this procedure is normally distributed with a mean of 6. Your SPSS output window should look like Figure 10. The first table provides information about the variable: the number of scores ("N"); the mean of the scores ("Mean"); the estimated population standard deviation, S ("Std. Deviation"); and the standard deviation of the distribution of means, S_M ("Std. Error Mean"). Check that the values in that table are consistent (allowing for rounding error) with the values in the "Example Worked-Out Problem" section. The second table in the SPSS output window gives the outcome of the *t* test. Compare the values of *t* and *df* in that table with the values given in the "Example Worked-Out Problem" section. The exact two-tailed significance level of the *t* test is given in the "Sig. 2-tailed" column. In this study, the researcher was using the .05 significance level. The significance level given by SPSS (.039) is more extreme than .05, which means that the researcher can reject the null hypothesis and the research hypothesis is supported.

t Test for Dependent Means

❶ Enter one set of scores (for example, the "before" scores) in the first column of the data window. Then enter the second set of scores (for example, the "after" scores) in the second column of the data window. (Be sure to enter the scores in the order they are listed.) Since each row in the SPSS data window represents a



One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
score	8	9.0000	3.33809	1.18019

One-Sample Test

	Test Value = 6					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
score	2.542	7	.039	3.00000	.2093	5.7907

Figure 10 Using SPSS to carry out a *t* test for a single sample for the Example Worked-Out Problem for a single sample *t* test.

separate person, it is important that you enter each person's scores in two separate columns.

❷ ✐ *Analyze.*

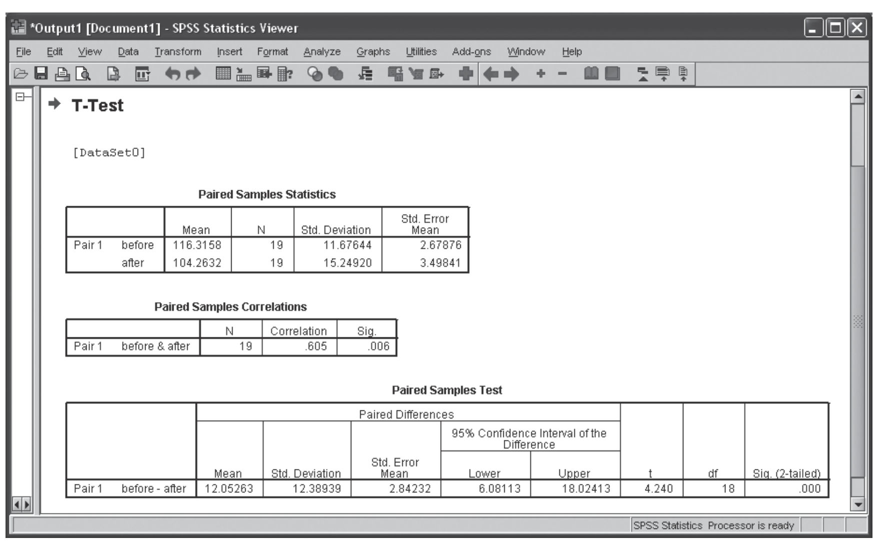
❸ ✐ *Compare means.*

❹ ✐ *Paired-Samples T Test* (this is the name SPSS uses for a *t* test for dependent means).

❺ ✐ on the first variable (this will highlight the variable). ✐ the arrow. ✐ on the second variable (this will highlight the variable). ✐ the arrow. The two variables will now appear as "Variable1" and "Variable2" for "Pair 1" in the "Paired Variables" box.

❻ ✐ *OK.*

Practice these steps by carrying out a *t* test for dependent means for Olthoff's (1989) study of communication quality of 19 men who received ordinary premarital counseling. The scores and figuring for that study are shown in Table 4. Your SPSS output window should look like Figure 11. The key information is contained in the third table (labeled "Paired Samples Test"). The final three columns of this table give the *t* score (4.240), the degrees of freedom (18), and the two-tailed significance level (.000 in this case) of the *t* test. The significance level is so small that even after rounding to three decimal places, it is less than .001. Since the significance level is more extreme than the .05 significance level we set for this study, you can reject the null hypothesis. By looking at the means for the "before" variable and the "after" variable in the first table (labeled "Paired Samples Statistics"), you can see that the husbands' communication quality was lower after marriage (a mean of 104.2632) than before marriage (a mean 116.3158). Don't worry that the *t* value figured in Table 4 was negative, whereas the *t* value in the SPSS output is positive. This happens because the difference score in Table 4 was figured as after minus before, but SPSS figured the difference scores as before minus after. Both ways of figuring the difference score are mathematically correct and the overall result is the same in each case.



*Output1 [Document1] - SPSS Statistics Viewer

T-Test

[DataSet0]

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	before	116.3158	19	11.67644	2.67876
	after	104.2632	19	15.24920	3.49841

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	before & after	19	.605	.006

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	before - after	12.05263	12.38939	2.84232	6.08113	18.02413	4.240	18	.000

SPSS Statistics Processor is ready

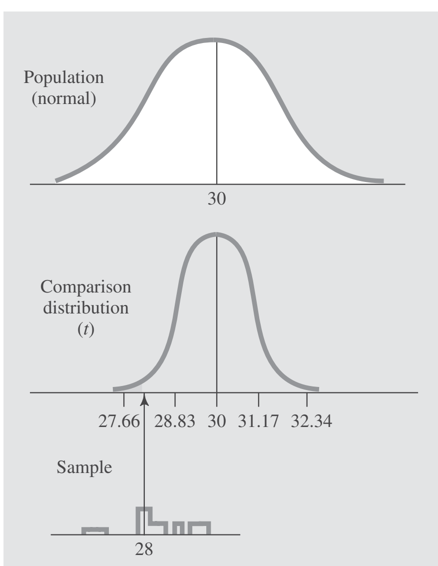
Figure 11 Using SPSS to carry out a *t* test for dependent means for Olthoff's (1989) study of communication quality among 19 men who received ordinary (very minimal) premarital counseling.

Answers to Set I Practice Problems

1. (a) t needed $(df = 63, p < .05$, one-tailed$) = -1.671$; $S_M^2 = S^2/N = 9/64 = .141$. $S_M = \sqrt{S_M^2} = \sqrt{.141} = .38$; $t = (M - \text{Population } M)/S_M = (11 - 12.40)/.38 = -3.68$; reject null hypothesis. (b) t needed $= 2.690$; $S_M = 2.55$; $t = 1.32$; do not reject null hypothesis. (c) t needed $= 2.364$; $S_M = .13$; $t = 3.15$; reject null hypothesis.
2. (a)
   ❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are two populations of interest:

   Population 1: Response times under the new sheriff.
   Population 2: Response times under the old sheriff.

   The research hypothesis is that response times under the new sheriff are lower than response times under the old sheriff. The null hypothesis is that the response times under the new sheriff are not less than under the old sheriff.
   ❷ **Determine the characteristics of the comparison distribution.** Population 2: shape = assumed normal; Population $M = 30$; The estimated population variance is $S^2 = [\Sigma(X - M)^2]/df = 124/(10 - 1) = 13.78$. Distribution of means: shape $= t\ (df = 9)$; mean of the distribution of means = 30; variance of the distribution of means is $S_M^2 = S^2/N = 13.78/10 = 1.378$; standard deviation of the distribution of means is $S_M = \sqrt{S_M^2} = \sqrt{1.378} = 1.17$.
   ❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** t needed $(df = 9, p < .05$, one-tailed$) = -1.833$.
   ❹ **Determine your sample's score on the comparison distribution.** $M = (\Sigma X)/N = 280/10 = 28$; $t = (M - \text{Population } M)/S_M = (28 - 30)/1.17 = -1.71$.
   ❺ **Decide whether to reject the null hypothesis.** -1.71 is not more extreme than -1.833; do not reject the null hypothesis.
   (b)



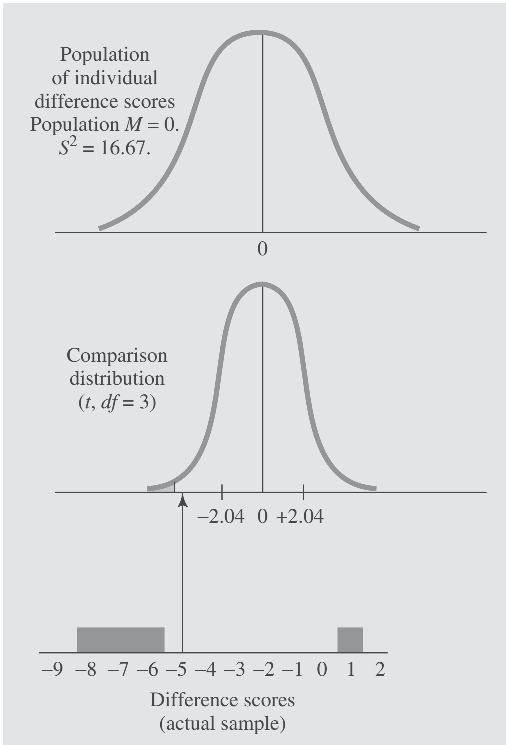
   (c) Similar to 5c (see next page), except instead of difference scores, actual scores are used here, and the expected population mean is the 30 minutes that the sheriff had promised to do better than the current sheriff.
3. (a) and (b) Hypothesis-testing steps and sketch similar to 2a and 2b above; t needed $= 2.776$; $t = (M - \text{Population } M)/S_M = (46 - 40)/2.3 = 2.61$; do not reject the null hypothesis. (c) Similar to 5c (see next page), except instead of difference scores, actual scores are used here, and the expected population mean is the 40 questions that people usually get correct.
4. (a) t needed $(df = 19, p < .05$, one-tailed$) = 1.729$; $S_M^2 = S^2/N = 8.29/20 = .415$; $S_M = \sqrt{S_M^2} = \sqrt{.415} = .64$; $t = (M - \text{Population } M)/S_M = (1.7 - 0)/.64 = 2.66$; reject null hypothesis. Estimated effect size $= M/S = 1.7/\sqrt{8.29} = .59$.
   (b) t needed $= -1.984, 1.984$; $S_M^2 = S^2/N = 414.53/164 = 2.53$; $S_M = \sqrt{S_M^2} = \sqrt{2.53} = 1.59$; $t = (2.3 - 0)/1.59 = 1.45$; do not reject the null hypothesis. Estimated effect size $= .10$.
   (c) t needed $= -2.625$; $S_M^2 = S^2/N = 4/15 = .27$; $S_M = \sqrt{S_M^2} = \sqrt{.27} = .52$; $t = -4.23$; reject null hypothesis. Estimated effect size $= -1.10$.
5. (a)
   ❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are two populations of interest:

   Population 1: Cities like those who participated in the anti-littering program.
   Population 2: Cities that do not change in the amount of litter over a 1-year period.

   The research hypothesis is that Population 1 has a greater mean decrease in litter than Population 2. The null hypothesis is that Population 1 doesn't have a greater mean decrease in litter than Population 2.
   ❷ **Determine the characteristics of the comparison distribution.** Population 2: shape = assumed normal; Population $M = 0$; The estimated population variance is $S^2 = [\Sigma(X - M)^2]/df = (50)/(4 - 1) = 16.67$. Distribution of means: shape $= t\ (df = 3)$, mean of the distribution of means of difference scores = 0; variance of the distribution of means of difference scores is $S_M^2 = S^2/N = 16.67/4 = 4.17$; standard deviation of the distribution of means of difference scores is $S_M = \sqrt{S_M^2} = \sqrt{4.17} = 2.04$.
   ❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** t needed $(df = 3, p < .01$, one-tailed$) = -4.541$. (The cutoff score is negative, because we will figure change scores as July 2010 litter minus July 2009 litter. Thus, negative change scores will be in the same direction as the hypothesized *decrease* in litter.)
   ❹ **Determine your sample's score on the comparison distribution.** Change scores $= -7, -6, 1, -8$; $M = -20/4 = -5$;

$t = (M - \text{Population } M)/S_M = (-5 - 0)/2.04 = -2.45$. ❺ **Decide whether to reject the null hypothesis.** -2.45 (from Step ❹) is not more extreme than the cutoff of (from Step ❸); do not reject the null hypothesis.

(b)



(c) In this situation, I am using a *t* test for dependent means to test whether the amount of litter in four cities changes from before to after the introduction of a program to decrease littering. The first thing I did was to simplify things by converting the numbers to change scores—postprogram (2010) litter minus preprogram (2009) litter for each city. Then I found the mean of these difference scores, which was -5. That is, there is an average decrease of 5 pounds of litter per block per day.

The next step was to see whether this result, found in these four cities, indicates some more general, real difference due to being in this program. The alternative is the possibility that this much change could have happened in four randomly selected cities just by chance, even if in general the program would have no real effect. That is, we imagine that the average change for cities in general that would do this program would actually be 0, and maybe this study just happened to pick four cities that would have decreased this much anyway.

I then considered just how much a group of four cities would have to change before I could conclude that they have changed too much to chalk it up to chance. This required figuring out the characteristics of this imagined population of cities in which, on the average, there is no change. An average of no change is the same as saying there is a mean of 0 change. Therefore, I used 0 as the mean of the comparison distribution. Since I didn't know the variance of this hypothetical distribution of cities that don't change, I estimated it from the information in the sample of four cities. If the sample cities were drawn at chance from the hypothetical distribution of no change, the variance of these cities should reflect the variance of this hypothetical distribution (which would be the distribution they come from). However, the variance figured from a particular group (sample) from a larger population will in general be slightly smaller than the true population's variance. In other words, it will be a biased estimate of the population variance. Thus, I had to modify the variance formula to take this into account: Instead of dividing the sum of the squared deviations by the number of scores (the usual method), I divided it instead by the "degrees of freedom." The degrees of freedom is the number of scores minus 1—in this case, $4 - 1 = 3$. (This adjustment exactly accounts for the tendency of the variance in the sample to underestimate the true population variance and results in an unbiased estimate of the population variance.) As shown in the calculations in the steps of hypothesis testing, this gave an estimated population variance (S^2) of 16.67.

I was interested not in individual cities but in a group of four. Thus, what I really needed to know was the characteristics of a distribution of means of samples of four taken from this hypothetical population of individual city difference scores. Such a distribution of means will have the same mean of 0. (This is because there is no reason to expect the means of such groups of four drawn randomly to be systematically higher or lower than the population's mean of 0.) However, such a distribution will have a much smaller variance than the variance of the population it comes from. (This is because the average of a group of four scores is a lot less likely to be extreme than any individual score.) Fortunately, it is known (and can be proved mathematically) that the variance of a distribution of means is the variance of the distribution of individuals, divided by the number of individuals in each sample. In our example, this works out to 16.67 divided by 4, which is 4.17. The standard deviation of this distribution is thus the square root of 4.17, or 2.04.

It also turns out that if we assume that the hypothetical population of individual cities' difference scores is normally distributed (and we have no reason to think otherwise), the distribution of means of samples from that distribution can be thought of as having a precise known shape, called a *t* distribution. (A *t* distribution is similar to a normal curve, but has slightly thicker tails). Thus, I looked in a table for a *t* distribution for the situation in which 3 degrees of freedom are used to estimate the population variance. The table shows that there is less than a 1% chance of getting a score that is -4.541 standard deviations from the mean of this distribution.

The mean difference score for the sample of four cities was -5, which is 2.45 (that is, $-5/2.04$) standard deviations below the mean of 0 change on this distribution of means of difference scores. This is not as extreme as -4.541. Thus, there is more than a 1% chance that these results could have come from a hypothetical distribution with no change. Therefore, the researcher would not rule out that possibility, and the experiment would be considered inconclusive.

6. (a), (b), and (c). Hypothesis-testing steps, sketch, and explanation similar to 5 above. *t* needed $= -2.776, 2.776$; $t = (M - \text{Population } M)/S_M = (-.02 - 0)/.007 = -2.86$; reject the null hypothesis.

7. (a) Estimated effect size $= M/S = 20/32 = .63$, medium effect size; (b) .50, medium; (c) .25, small; (d) .20, small.
8. From Table 9: (a) .22; (b) .71; (c) .86; (d) .77; (e) .99.
9. From Table 10: (a) 33; (b) 12; (c) 156.
10. Similar to 5c above except focusing on this study and the simpler situation involving just a single sample. Also, you do not need to explain the basic logic of hypothesis testing (only what is added when you have an unknown population variance).
11. Similar to 5b and 5c above, except your explanation should focus on this study and you should add material on mean, standard deviation, and variance.

Steps of Hypothesis Testing for Major Procedures

***t* test for a single sample**

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.**

❷ **Determine the characteristics of the comparison distribution.**
(a) The mean of the distribution of means is the same as the population mean. **(b)** Figure its standard deviation:

Ⓐ $S^2 = [\Sigma(X - M)^2]/df$

Ⓑ $S_M^2 = S^2/N$

Ⓒ $S_M = \sqrt{S_M^2}$

(c) t distribution, $df = N - 1$.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** Use t table.

❹ **Determine your sample's score on the comparison distribution.** $t = (M - \text{Population } M)/S_M$

❺ **Decide whether to reject the null hypothesis.** Compare scores from Steps ❸ and ❹.

***t* test for dependent means**

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.**

❷ **Determine the characteristics of the comparison distribution.**
(a) All based on difference scores. **(b)** Figure mean of the difference scores. **(c)** Mean of distribution of means of difference scores $= 0$. **(d)** Figure its standard deviation:

Ⓐ $S^2 = [\Sigma(X - M)^2]/df$

Ⓑ $S_M^2 = S^2/N$

Ⓒ $S_M = \sqrt{S_M^2}$

(e) t distribution, $df = N - 1$.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** Use t table.

❹ **Determine your sample's score on the comparison distribution.** $t = (M - \text{Population } M)/S_M$

❺ **Decide whether to reject the null hypothesis.** Compare scores from Steps ❸ and ❹.