The *t* Test for Independent Means

Chapter Outline

You should already know how to use the *t* test for dependent means to compare two sets of scores from a *single group of people* (such as the same men measured on communication before and after premarital counseling). In this chapter, you learn how to compare two sets of scores, one from each of *two entirely separate groups of people.* This is a very common situation in behavioral and social sciences research. For example, a study may compare the scores from individuals in an experimental group with the scores from individuals in a control group. (Another example is comparing scores from a group of men with scores from a group of women.) The scores of these two groups are independent of each other, so the test you will learn is called a ***t* test for independent means.**

TIP FOR SUCCESS
You should be thoroughly comfortable with the basic logic and procedures of the *t* test for a single sample before going on to the material in this chapter.

Let's consider an example. A team of researchers is interested in whether writing one's thoughts and feelings about traumatic life events can affect physical health. This kind of writing is called "expressive writing." Suppose the researchers recruit

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

From Chapter 9 of *Statistics for the Behavioral & Social Sciences: A Brief Course*, Fifth Edition. Arthur Aron, Elliot J. Coups, Elaine N. Aron.

undergraduate students to take part in a study and then randomly assign them to be in either an expressive writing group or a control group. Students in the expressive writing group are instructed to write four 20-minute essays, one over each of 4 consecutive days, about their most traumatic life experiences. Students in the control group also write four 20-minute essays over 4 consecutive days, but their essays are each about their plans for that day. One month later, the researchers ask the students to rate their overall level of physical health (on a scale from 0 = *very poor health* to 100 = *perfect health*). The expressive writing and the control group contain different students. Thus, the researchers need to use a *t* test for independent means to test the effect of expressive writing on physical health. We return to this example later in the chapter. But first, you will learn about the logic of the *t* test for independent means, which involves learning about a new kind of distribution (called the *distribution of differences between means*).

The Distribution of Differences between Means

In the situation you face in this chapter, you have two groups of people with one score for each person. You don't have any pairs of scores for each person. Thus, it wouldn't make sense in this new situation to create difference scores, and you can't use difference scores for the hypothesis-testing procedure in this chapter. Instead, when you have one score for each person with two different groups of people, what you can compare is the *mean* of one group to the *mean* of the other group.

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

A distribution of differences between means is, in a sense, two steps removed from the populations of individuals. First, there is a distribution of means from each population of individuals. Second, there is a distribution of differences between pairs of means, one of each pair of means from its particular distribution of means.

Think of this distribution of differences between means as being built up as follows: (1) randomly select one mean from the distribution of means for the first group's population, (2) randomly select one mean from the distribution of means for the second group's population, and (3) subtract. (That is, take the mean from the first distribution of means and subtract the mean from the second distribution of means.) This gives a difference score between the two selected means. Then repeat the process. This creates a second difference, a difference between the two newly selected means. Repeating this process a large number of times creates a distribution of differences between means. You would never actually create a distribution of differences between means using this lengthy method (there is a simpler mathematical rule that accomplishes the same thing). But it shows clearly what makes up the distribution.

TIP FOR SUCCESS

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

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

The Logic and a Visual Picture of It All

Figure 1 shows the entire logical construction involved in a distribution of differences between means. At the top are the two population distributions. We do not know the characteristics of these population distributions. But we do know that if the null hypothesis is true, the two population means are the same. That is, the null hypothesis is that Population M_1 = Population M_2. We also can estimate the variance of these populations based on the sample information (these estimated variances will be S_1^2 and S_2^2).

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

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

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

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

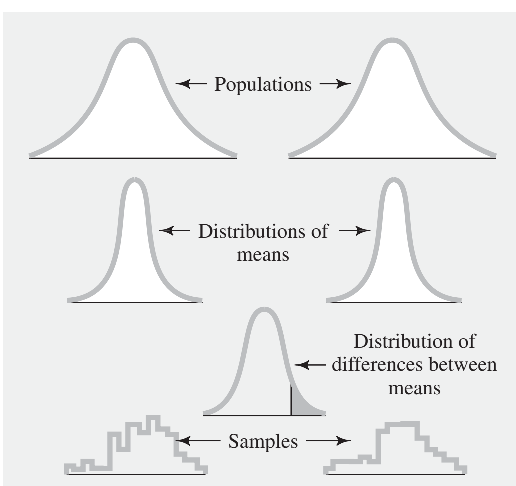


Figure 1 A diagram of the logic of a distribution of differences between means.

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

Mean of the Distribution of Differences between Means

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

How are you doing?

1. (a) When would you carry out a *t* test for independent means? (b) How is this different from the situation in which you would carry out a *t* test for dependent means?
2. (a) What is the comparison distribution in a *t* test for independent means? (b) How is this different from the comparison distribution in a *t* test for dependent means?
3. (a) In the context of the *t* test for independent means, explain the logic of going from scores in two samples to an estimate of the variance of this comparison distribution. (b) Illustrate your answer with sketches of the distributions involved. (c) Why is the mean of this distribution 0?

Answers

1. (a) You carry out a *t* test for independent means when you have done a study in which you have scores from two samples of different individuals and you do not know the population variance. (b) In a *t* test for dependent means you have two scores from each of several individuals, but there is only one group of people studied.

2. (a) The comparison distribution in a *t* test for independent means is a distribution of differences between means. (b) In a *t* test for dependent means, you have two sets of scores from a single group, such as before scores and after scores. This allows you to create difference scores and the hypothesis testing is carried out with these difference scores. Thus, the comparison distribution for a *t* test for dependent means is a distribution of means of difference scores.

3. (a) You estimate the population variance from each sample's scores. Based on each estimated population variance, you figure the variance of the distribution of means for each population. You can then use the variances of the distributions of means from the two populations to estimate the variance of the distribution of differences between means, which is the comparison distribution for the *t* test for independent means. (b) Your sketch should look like Figure 1. (c) The mean of the distribution will be zero because if the null hypothesis is true, the two populations have the same mean. So differences between means would on the average come out to zero.

Estimating the Population Variance

The population variance is the sum of squared deviation scores divided by the degrees of freedom (the number in the sample minus 1). To do a *t* test for independent means, it has to be reasonable to assume that the populations the two samples come from have the same variance. (If the null hypothesis is true, they also have the same mean. However, regardless of whether the null hypothesis is true, you must be able to assume that the two populations have the same variance.) Therefore, when you estimate the population variance from the scores in either sample, you are getting two separate estimates of what should be the same number. In practice, the two estimates will almost never be exactly identical. Since they are both supposed to be estimating the same thing, the best solution is to average the two estimates to get the best single overall estimate. This is called the **pooled estimate of the population variance (S^2_{Pooled}).**

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

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

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

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

The pooled estimate of the population variance is the first sample's proportion of the total degrees of freedom multiplied by the first sample's estimated variance, plus the second sample's proportion of the total degrees of freedom multiplied by the second sample's estimated population variance.

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

In this formula, S^2_{Pooled} is the pooled estimate of the population variance. df_1 is the degrees of freedom in the sample for Population 1, and df_2 is the degrees of freedom in the sample for Population 2. (Remember, each sample's df is its number of scores minus 1.) df_{Total} is the total degrees of freedom $(df_{\text{Total}} = df_1 + df_2)$. (Thus, df_1/df_{Total} is the first sample's proportion of the total degrees of freedom in the study.) S_1^2 is the estimate of the population variance based on the scores in the sample from Population 1; S_2^2 is the estimate based on the scores in the sample from Population 2.

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

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

$$S^2_{\text{Pooled}} = \frac{df_1}{df_{\text{Total}}}(S_1^2) + \frac{df_2}{df_{\text{Total}}}(S_2^2) = \frac{10}{40}(60) + \frac{30}{40}(80)$$
$$= \frac{1}{4}(60) + \frac{3}{4}(80) = 15 + 60 = 75$$

TIP FOR SUCCESS

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

Notice that this procedure does not give the same result as ordinary averaging (without weighting). Ordinary averaging would give an estimate of 70 (that is, $[60 + 80]/2 = 70$). Your weighted, pooled estimate of the population variance of 75 is closer to the estimate based on the control group alone than to the estimate based on the experimental group alone. This is as it should be, because the control group estimate in this example was based on more information.

Figuring the Variance of Each of the Two Distributions of Means

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

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

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

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

$$S^2_{M_2} = \frac{S^2_{\text{Pooled}}}{N_2} \qquad (3)$$

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

In terms of formulas,

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

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

TIP FOR SUCCESS

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

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

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

$$S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2} \qquad (4)$$

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

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

$$S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}} \qquad (5)$$

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

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

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

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

Steps to find the Standard Deviation of the Distribution of Differences between Means

Ⓐ **Figure the estimated population variances based on each sample.** That is, figure one estimate for each population using the formula $S^2 = [\Sigma(X - M)^2]/(N - 1)$.

Ⓑ **Figure the pooled estimate of the population variance:**

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

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

Ⓒ **Figure the variance of each distribution of means:**

$$S^2_{M_1} = S^2_{\text{Pooled}}/N_1 \text{ and } S^2_{M_2} = S^2_{\text{Pooled}}/N_2$$

Ⓓ **Figure the variance of the distribution of differences between means:**

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

Ⓔ **Figure the standard deviation of the distribution of differences between means:**

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

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

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

The Shape of the Distribution of Differences between Means

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

The total degrees of freedom is the degrees of freedom in the first sample plus the degrees of freedom in the second sample.

$$df_{\text{Total}} = df_1 + df_2 \tag{6}$$

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

The *t* Score for the Difference between the Two Actual Means

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

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

$$t = \frac{M_1 - M_2}{S_{\text{Difference}}} \tag{7}$$

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

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

How are you doing?

1. Write the formula for each of the following: (a) pooled estimate of the population variance, (b) variance of the distribution of means for the first population, (c) variance of the distribution of differences between means, and (d) *t* score in a *t* test for independent means. (e) Define all the symbols used in these formulas.
2. Explain (a) why a *t* test for independent means uses a single pooled estimate of the population variance, (b) why, and (c) how this estimate is "weighted."

3. For a particular study comparing means of two samples, the first sample has 21 participants and an estimated population variance of 100; the second sample has 31 participants and an estimated population variance of 200. (a) What is the standard deviation of the distribution of differences between means? (b) What is its mean? (c) What will be its shape? (d) Illustrate your answer with sketches of the distributions involved.

Answers

1. (a) Pooled estimate of the population variance: $S^2_{Pooled} = \frac{df_1}{df_{Total}}(S^2_1) + \frac{df_2}{df_{Total}}(S^2_2)$

(b) Variance of the distribution of means for the first population: $S^2_{M_1} = \frac{S^2_{Pooled}}{N_1}$

(c) Variance of the distribution of differences between means: $S^2_{Difference} = S^2_{M_1} + S^2_{M_2}$

(d) *t* score in a *t* test for independent means: $t = \frac{M_1 - M_2}{S_{Difference}}$

(e) S^2_{Pooled} is the pooled estimate of the population variance; df_1 and df_2 are the degrees of freedom in the samples from the first and second populations, respectively; df_{Total} is the total degrees of freedom (the sum of df_1 and df_2); S^2_1 and S^2_2 are the population variance estimates based on the samples from the first and second populations, respectively; $S^2_{M_1}$ is the variance of the distribution of means for the first population based on an estimated variance of the population of individuals; N_1 is the number of participants in the sample from the first population; $S^2_{Difference}$ is the variance of the distribution of differences between means based on estimated variances of the populations of individuals; *t* is the *t* score for a *t* test for independent means (the number of standard deviations from the mean on the distribution of differences between means); M_1 and M_2 are the means of the samples from the first and second populations, respectively; and $S_{Difference}$ is the standard deviation of the distribution of differences between means based on estimated variances of the populations of individuals.

2. (a) You assume that both populations have the same variance; thus the estimates from the two samples should be estimates of the same number. (b) We weight (give more influence to) an estimate from a larger sample because, being based on more information, it is likely to be more accurate. (c) The actual weighting is done by multiplying each sample's estimate by the degrees of freedom for that sample divided by the total degrees of freedom; you then sum these two products.

3. (a) Standard deviation of the distribution of differences between means: $S^2_{Pooled} = (20/50)(100) + (30/50)(200) = 40 + 120 = 160$;
$S^2_{M_1} = 160/21 = 7.62$; $S^2_{M_2} = 160/31 = 5.16$;
$S^2_{Difference} = 7.62 + 5.16 = 12.78$;
$S_{Difference} = \sqrt{12.78} = 3.57$.

(b) Mean: 0; (c) Shape: *t* distribution with $df = 50$; (d) Should look like Figure 1 with numbers written in (see Figure 2 for an example).

Hypothesis Testing with a *t* Test for Independent Means

Considering the five steps of hypothesis testing, there are three new wrinkles for a *t* test for independent means: (1) the comparison distribution is now a distribution of differences between means (this affects Step ❷); (2) the degrees of freedom for finding the cutoff on the *t* table is based on two samples (this affects Step ❸); and (3) your sample's score on the comparison distribution is based on the difference between your two means (this affects Step ❹).

Example of a *t* Test for Independent Means

Let's return to the expressive writing study example from the start of the chapter. Twenty students were recruited to take part in the study. The 10 students randomly assigned to the expressive writing group wrote about their thoughts and feelings associated with their most traumatic life events. The 10 students randomly assigned to the control group wrote about their plans for the day. One month later, all of the students rated their overall level of physical health on a scale from 0 = *very poor health* to 100 = *perfect health*. (Although based on actual studies, we have made up the details of this study to be an easier example to follow for learning. Mainly, actual studies usually have large samples. Real studies on this kind of topic also often use more objective measures of health, such as number of physician visits or days missed from school.)

The scores and figuring for the *t* test are shown in Table 1. Figure 2 shows the distributions involved. Let's go through the five 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: Students who do expressive writing.
Population 2: Students who write about a neutral topic (their plans for the day).

The researchers were interested in whether there was a positive or a negative health effect of expressive writing. Thus, the research hypothesis was that Population 1 students would rate their health differently from Population 2 students (a two-tailed test). The null hypothesis was that Population 1 students would rate their health the same as Population 2 students.

❷ **Determine the characteristics of the comparison distribution.** The comparison distribution is a distribution of differences between means. (a) Its mean is 0 (as it almost always is in a *t* test for independent means, because we are interested in whether there is more than 0 difference between the two populations). (b) Regarding its standard deviation,

Ⓐ **Figure the estimated population variances based on each sample.** As shown in Table 1, S_1^2 comes out to 94.44 and $S_2^2 = 111.33$.

Ⓑ **Figure the pooled estimate of the population variance.** As shown in Table 1, the figuring for S^2_{Pooled} gives a result of 102.89.

Ⓒ **Figure the variance of each distribution of means.** Dividing S^2_{Pooled} by the N in each sample, as shown in Table 1, gives $S^2_{M_1} = 10.29$ and $S^2_{M_2} = 10.29$.

Ⓓ **Figure the variance of the distribution of differences between means.** Adding up the variances of the two distributions of means, as shown in Table 1, comes out to $S^2_{\text{Difference}} = 20.58$.

Ⓔ **Figure the standard deviation of the distribution of differences between means.** $S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}} = \sqrt{20.58} = 4.54$.

TIP FOR SUCCESS

In this example, note that the value for $S^2_{M_1}$ is the same as the value for $S^2_{M_2}$. This is because there are the same number of students in the two groups (that is, N_1 was the same as N_2). When the number of individuals in the two groups is not the same, the values for $S^2_{M_1}$ and $S^2_{M_2}$ will be different.

Table 1 *t* Test for Independent Means for a Fictional Study of the Effect of Expressive Writing on Physical Health

Expressive Writing Group			Control Writing Group		
Score	Deviation from Mean (Score − M)	Squared Deviation from Mean	Score	Deviation from Mean (Score − M)	Squared Deviation from Mean
77	−2	4	87	19	361
88	9	81	77	9	81
77	−2	4	71	3	9
90	11	121	70	2	4
68	−11	121	63	−5	25
74	−5	25	50	−18	324
62	−17	289	58	−10	100
93	14	196	63	−5	25
82	3	9	76	8	64
79	0	0	65	−3	9
Σ: 790		850	680		1002

$M_1 = 79.00$; $S_1^2 = 850/9 = 94.44$; $M_2 = 68.00$; $S_2^2 = 1002/9 = 111.33$

$N_1 = 10$; $df_1 = N_1 - 1 = 9$; $N_2 = 10$; $df_2 = N_2 - 1 = 9$

$df_{\text{Total}} = df_1 + df_2 = 9 + 9 = 18$

$$S^2_{\text{Pooled}} = \frac{df_1}{df_{\text{Total}}}(S_1^2) + \frac{df_2}{df_{\text{Total}}}(S_2^2) = \frac{9}{18}(94.44) + \frac{9}{18}(111.33) = 47.22 + 55.67 = 102.89$$

$S^2_{M_1} = S^2_{\text{Pooled}}/N_1 = 102.89/10 = 10.29$

$S^2_{M_2} = S^2_{\text{Pooled}}/N_2 = 102.89/10 = 10.29$

$S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2} = 10.29 + 10.29 = 20.58$

$S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}} = \sqrt{20.58} = 4.54$

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

$t = (M_1 - M_2)/S_{\text{Difference}} = (79.00 - 68.00)/4.54 = 2.42$

Decision: Reject the null hypothesis.

The shape of this comparison distribution will be a *t* distribution with a total of 18 degrees of freedom.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** This requires a two-tailed test because the researchers were interested in an effect in either direction. As shown in Table 2 of the appendix "Tables," the cutoff *t* scores at the .05 level are 2.101 and −2.101.

❹ **Determine your sample's score on the comparison distribution.** The *t* score is the difference between the two sample means (79.00 − 68.00, which is 11.00), divided by the standard deviation of the distribution of differences between means (which is 4.54). This comes out to 2.42.

❺ **Decide whether to reject the null hypothesis.** The *t* score of 2.42 for the difference between the two actual means is larger than the cutoff *t* score of 2.101. You can reject the null hypothesis. The research hypothesis is supported: Students who do expressive writing report a higher level of health than students who write about a neutral topic.

The actual numbers in this study were fictional. However, the results are consistent with those from many actual studies that have shown beneficial effects of

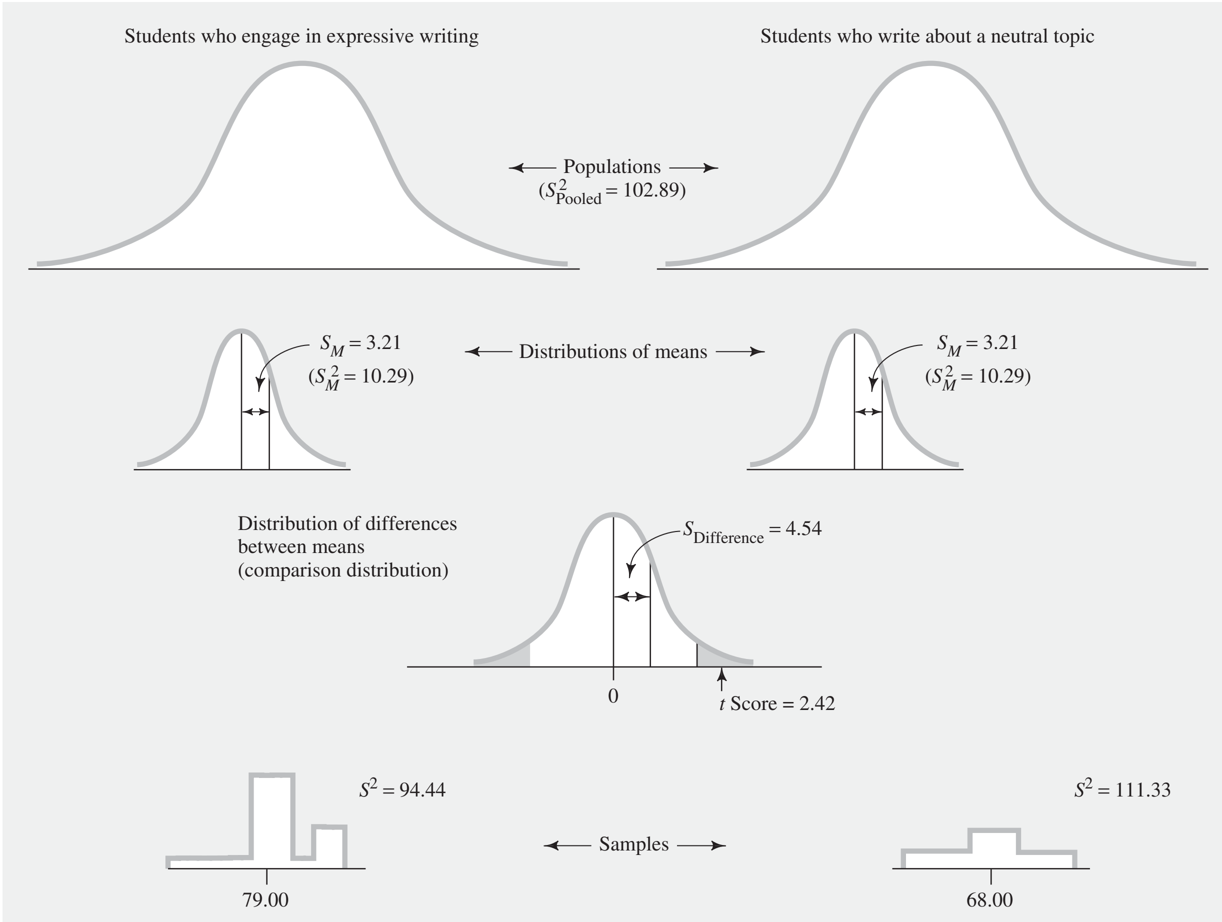


Figure 2 The distributions for a *t* test for independent means for the expressive writing example.

expressive writing on self-reported health outcomes, as well as on other outcomes, such as number of doctor visits and psychological well-being (e.g., Pennebaker & Beall, 1986; Warner et al., 2006; see also Frattaroli, 2006).

Summary of Steps for Conducting a *t* Test for Independent Means

Table 2 summarizes the steps for a *t* test for independent means.[1]

[1]The steps of figuring the standard deviation of the distribution of differences between means can be combined into a single overall computational formula:

$$S_{\text{Difference}} = \sqrt{\frac{(N_1 - 1)(S_1^2) + (N_2 - 1)(S_2^2)}{N_1 + N_2 - 1}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)} \quad (8)$$

As usual, we urge you to use the full set of steps and the regular, definitional formulas in your figuring when doing the practice problems in this text. Those steps help you learn the basic principles. However, this computational formula will be useful if statistics software is not available and you have to figure by hand a *t* test for independent means on scores from a real study with many participants in each group.

Table 2 Steps for a t Test for Independent 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. Its mean will be 0.

b. Figure its standard deviation.

Ⓐ **Figure the estimated population variances based on each sample.** For each population, $S^2 = [\Sigma(X - M)^2]/(N - 1)$.

Ⓑ **Figure the pooled estimate of the population variance:**

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

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

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

Ⓓ **Figure the variance of the distribution of differences between means:**

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

Ⓔ **Figure the standard deviation of the distribution of differences between means:**

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

c. The comparison distribution will be a t distribution with df_{Total} degrees of freedom.

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

a. Determine the degrees of freedom (df_{Total}), desired significance level, and tails in the test (one or two).

b. Look up the appropriate cutoff in a t table. If the exact df is not given, use the df below it.

❹ **Determine your sample's score on the comparison distribution:**

$t = (M_1 - M_2)/S_{\text{Difference}}$

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

A Second Example of a t Test for Independent Means

Suppose a researcher wants to study the effectiveness of a new job skills training program for people who have not been able to hold a job. Fourteen people who have not been able to hold a job agree to be in the study. The researcher randomly picks six of these volunteers to be an experimental group that will go through the special training program. The other eight volunteers are put in a control group that will go through an ordinary job skills training program. After finishing their training program (of whichever type), all 14 are placed in similar jobs.

A month later, each volunteer's employer is asked to rate how well the new employee is doing using a 9-point scale. The scores and figuring for the t test are shown in Table 3. The distributions involved are shown in Figure 3. Let's carry out the analysis, following the five 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 who could not hold a job who then participate in the special job skills program.
Population 2: Individuals who could not hold a job who then participate in an ordinary job skills program.

Table 3 Figuring for a *t* Test for Independent Means for the Job Skills Example

Experimental Group (Receiving Special Program)			Control Group (Receiving Ordinary Program)		
Score	Deviation from Mean	Squared Deviation from Mean	Score	Deviation from Mean	Squared Deviation from Mean
6	0	0	6	3	9
4	−2	4	1	−2	4
9	3	9	5	2	4
7	1	1	3	0	0
7	1	1	1	−2	4
3	−3	9	1	−2	4
Σ: 36	0	24	4	1	1
			3	0	0
			Σ: 24	0	26

$M_1 = 36/6$; $S_1^2 = 24/5 = 4.8$; $M_2 = 24/8 = 3$; $S_2^2 = 26/7 = 3.71$

$N_1 = 6$; $df_1 = 6 - 1 = 5$; $N_1 = 8$; $df_1 = 8 - 1 = 7$

$df_{\text{Total}} = df_1 + df_2 = 5 + 7 = 12$

$S^2_{\text{Pooled}} = \frac{df_1}{df_{\text{Total}}}(S_1^2) + \frac{df_2}{df_{\text{Total}}}(S_2^2) = \frac{5}{12}(4.8) + \frac{7}{12}(3.71) = 2 + 2.16 = 4.16$

$S^2_{M_1} = S^2_{\text{Pooled}}/N_1 = 4.16/6 = 0.69$

$S^2_{M_2} = S^2_{\text{Pooled}}/N_2 = 4.16/8 = 0.52$

$S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2} = 0.69 + 0.52 = 1.21$

$S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}} = \sqrt{1.21} = 1.10$

Needed *t* with $df = 12$, 5% level, two-tailed $= \pm 2.179$

$t = (M_1 - M_2)/S_{\text{Difference}} = (6 - 3)/1.10 = 3/1.10 = 2.73$

Decision: Reject the null hypothesis, the research hypothesis is supported.

It is possible for the special program to have either a positive or a negative effect compared to the ordinary program, and either result would be of interest. Thus, the research hypothesis is that the means of Population 1 and Population 2 are different. This is a nondirectional hypothesis. The null hypothesis is that the means of Population 1 and Population 2 are the same.

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

a. Its mean will be 0.

b. Figure its standard deviation. See Table 3 for the figuring for each step below.

Ⓐ **Figure the estimated population variances based on each sample.** $S_1^2 = 4.8$ and $S_2^2 = 3.71$.

Ⓑ **Figure the pooled estimate of the population variance.** $S^2_{\text{Pooled}} = 4.16$.

Ⓒ **Figure the variance of each distribution of means.** Dividing S^2_{Pooled} by the N in each sample gives $S^2_{M_1} = 0.69$ and $S^2_{M_2} = 0.52$.

Ⓓ **Figure the variance of the distribution of differences between means.** Adding up the variances of the two distribution of means comes out to $S^2_{\text{Difference}} = 1.21$.

Ⓔ **Figure the standard deviation of the distribution of differences between means.** $S_{\text{Difference}} = \sqrt{1.21} = 1.10$

c. It is a *t* distribution with $df_{\text{Total}} = 12$.

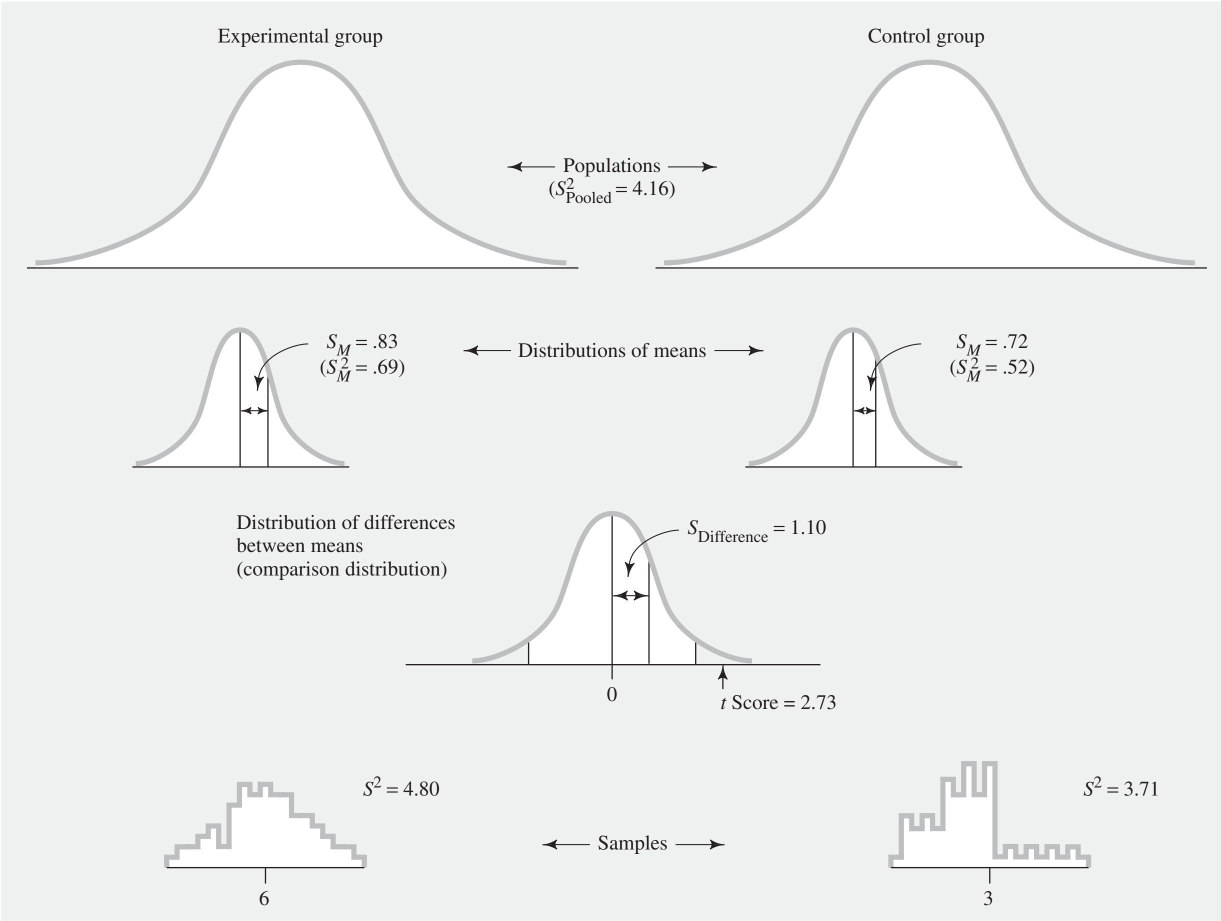


Figure 3 The distributions involved in the job skills example of a *t* test for independent means.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** The cutoff you need is for a two-tailed test (because the research hypothesis is nondirectional) at the usual .05 level, with 12 degrees of freedom. Looking this up on the *t* table, the cutoff *t* scores are 2.179 and −2.179.

❹ **Determine your sample's score on the comparison distribution.** The *t* score is the difference between the two sample means divided by the standard deviation of the distribution of differences between means. This comes out to a *t* of 2.73. (That is, $t = 3.00/1.10 = 2.73$.)

❺ **Decide whether to reject the null hypothesis.** The *t* score of 2.73 is more extreme than the cutoff *t* score of 2.179. Thus, the researchers can reject the null hypothesis. The research hypothesis is supported: The new special job skills program is more effective than the ordinary job skills program.

How are you doing?

1. List the ways in which hypothesis testing for a *t* test for independent means is different from a *t* test for dependent means in terms of (a) Step ❷, (b) Step ❸, and (c) Step ❹.
2. Using the .05 significance level, two-tailed, figure a *t* test for independent means for an experiment in which scores in an experimental condition are predicted to be lower than scores in a control condition. For the experimental condition, with $M = 8$, $S^2 = 12$. 26 participants, $M = 5$, $S^2 = 10$; for the control condition, with 36 participants, (a) Use the five steps of hypothesis testing. (b) Sketch the distributions involved.

Answers

1. (a) The comparison distribution for a *t* test for independent means is a distribution of differences between means.
   (b) The degrees of freedom for a *t* test for independent means is the sum of the degrees of freedom for the two samples.
   (c) The *t* score for a *t* test for independent means is based on differences between means (divided by the standard deviation of the distribution of differences between means).
2. (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 given the experimental procedure.
   Population 2: People given the control procedure.

   The research hypothesis is that the mean of Population 1 is less than the mean of Population 2. The null hypothesis is that the mean of Population 1 is not less than the mean of Population 2.

   ❷ **Determine the characteristics of the comparison distribution.**
   a. Its mean will be 0.
   b. Figure its standard deviation.
   Ⓐ **Figure the estimated population variances based on each sample.** $S_1^2 = 10$ and $S_2^2 = 12$.
   Ⓑ **Figure the pooled estimate of the population variance.** $S^2_{\text{Pooled}} = (25/60)(10) + (35/60)(12) = 4.17 + 7.00 = 11.17$.
   Ⓒ **Figure the variance of each distribution of means.** $S^2_{M_1} = 11.17/26 = .43$ and $S^2_{M_2} = 11.17/36 = .31$.
   Ⓓ **Figure the variance of the distribution of differences between means.** $S^2_{\text{Difference}} = .43 + .31 = .74$.
   Ⓔ **Figure the standard deviation of the distribution of differences between means.** $S_{\text{Difference}} = \sqrt{.74} = .86$.
   c. It is a *t* distribution with $df_{\text{Total}} = 60$.

   ❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** The *t* cutoff for the .05 level, one-tailed, $df = 60$ is -1.671. (The cutoff is a negative *t* score, because the research hypothesis is that the mean of Population 1 will be *lower* than the mean of Population 2.)

   ❹ **Determine your sample's score on the comparison distribution.** $t = (M_1 - M_2)/S_{\text{Difference}} = (5 - 8)/.86 = -3.49$.

   ❺ **Decide whether to reject the null hypothesis.** The *t* of -3.49 is more extreme than the cutoff *t* of -1.671. Therefore, reject the null hypothesis.

   b. The distributions involved are shown in Figure 4.

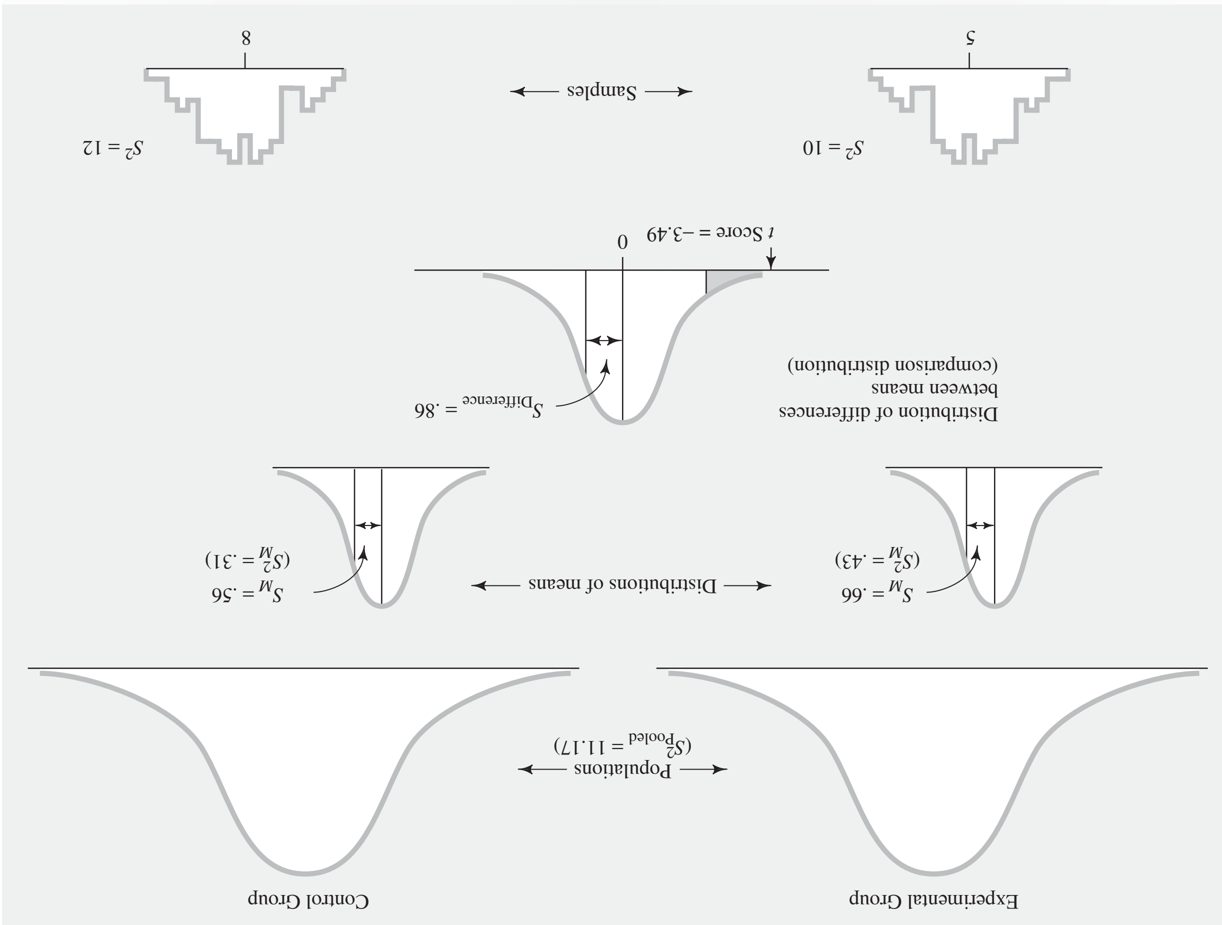


Figure 4 The distributions for a *t* test for independent means for the answer to the "How are you doing?" question 2.

Assumptions of the *t* Test for Independent Means

The first assumption for a *t* test for independent means is the same as that for any *t* test: Each of the population distributions is assumed to follow a normal curve. In practice, this is only a problem if you have reason to think that the two populations are dramatically skewed distributions, and in opposite directions. The *t* test holds up well even when the shape of the population distributions is fairly far from normal.

In a *t* test for independent means, you also have to be able to assume that the two populations have the same variance. (You take advantage of this assumption when you average the estimates from each of the two samples.) Once again, however, it turns out that in practice the *t* test gives pretty accurate results even when there are fairly large differences in the population variances, particularly when there are equal or near equal numbers of scores in the two samples.

However, the *t* test can give quite misleading results if (a) the scores in the samples suggest that the populations are very far from normal (highly skewed), (b) the variances are very different, or (c) there are both problems. In these situations, there are alternatives to the ordinary *t* test procedure.

Many computer programs for figuring the *t* test for independent means actually provide two sets of results. One set of results figures the *t* test assuming the population variances are equal. This method is the standard one, the one you have learned in this chapter. A second set of results uses a special alternative procedure that takes into account that the population variances may be unequal. (But it still assumes that the populations follow a normal curve.) An example of these two sets of results is shown in the "Using SPSS" section at the end of this chapter (see Figure 8). However, in most situations we can assume that the population variances are equal. Thus, researchers usually use the standard method. Using the special alternative procedure has the advantage that you don't have to worry about whether you met the equal population variance assumption. But it has the disadvantage that if you have met that assumption, with this special method you have less statistical power. That is, when you do meet the assumption, you would be less likely to get a significant result if you used the special method.

Effect Size and Power for the *t* Test for Independent Means

Effect Size

The effect size for the *t* test for independent means is the difference between the population means divided by the standard deviation of the population of individuals. When you have results of a completed study, you estimate the effect size as the difference between the sample means divided by the pooled estimate of the population standard deviation. (The pooled estimate of the population standard deviation is the square root of the pooled estimate of the population variance.) Stated as a formula,

The estimated effect size for a *t* test for independent means is the difference between the sample means divided by the pooled estimate of the population's standard deviation.

$$\text{Estimated Effect Size} = \frac{M_1 - M_2}{S_{\text{Pooled}}} \tag{9}$$

Cohen's (1988) conventions for the *t* test for independent means are the same as in all the situations we have considered so far: .20 for a small effect size, .50 for a medium effect size, and .80 for a large effect size.

Consider our example of the effectiveness of a special job skills program. The mean for the sample of individuals who participated in the special job skills program was 6.00. The mean for the sample of individuals who participated in the ordinary job skills program was 3.00. We figured the pooled estimate of the population variance to be 4.16; the pooled estimate of the population's standard deviation is thus 2.04. The difference in means of 3.00, divided by 2.04, gives an effect size of 1.47. This is a very large effect size. In terms of the formula,

$$\text{Estimated Effect Size} = \frac{M_1 - M_2}{S_{\text{Pooled}}} = \frac{6.00 - 3.00}{2.04} = 1.47$$

Power

Power for a *t* test for independent means can be determined using a power table, a power software package, or an Internet power calculator. The power table shown in Table 4 gives the approximate power for the .05 significance level for small, medium, and large effect sizes, and one-tailed or two-tailed tests.[2]

[2]Cohen (1988, pp. 28–39) provides more detailed tables in terms of number of participants, levels of effect size, and significance levels. Note 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 4 Approximate Power for Studies Using the *t* Test for Independent Means Testing Hypotheses at the .05 Significance Level

Number of Participants in Each Group	Effect Size		
	Small (.20)	Medium (.50)	Large (.80)
One-tailed test			
10	.11	.29	.53
20	.15	.46	.80
30	.19	.61	.92
40	.22	.72	.97
50	.26	.80	.99
100	.41	.97	*
Two-tailed test			
10	.07	.18	.39
20	.09	.33	.69
30	.12	.47	.86
40	.14	.60	.94
50	.17	.70	.98
100	.29	.94	*

*Nearly 1.

For example, suppose you have read a study using a *t* test for independent means that had a nonsignificant result using the .05 significance level, two-tailed. There were 40 participants in each group. Should you conclude that there is in fact no difference at all in the populations? This conclusion seems quite unjustified. Table 4 shows a power of only .14 for a small effect size. This suggests that, if such a small effect really exists in the populations, this study would probably not come out significant. Still, we can also conclude that if there is a true difference in the populations, it is probably not large. Table 4 shows a power of .94 for a large effect size. This suggests that if a large effect exists, it almost surely would have produced a significant result.

Power When Sample Sizes Are Not Equal

For a study with any given total number of participants, power is greatest when the participants are divided into two equal groups. Recall the example from the start of this chapter where the 42 participants were divided into 11 in the experimental group and 31 in the control group. This study has much less power than it would have if the researchers had been able to divide their 42 participants into 21 in each group.

There is a practical problem in figuring power from tables when sample sizes are not equal. (Power software packages and Internet power calculators require you to specify the sample sizes, which are then taken into account when they figure power.) Like most power tables, Table 4 assumes equal numbers in each of the two groups. What do you do when your two samples have different numbers of people in them? It turns out that in terms of power, the **harmonic mean** of the numbers of participants in two unequal sample sizes gives the equivalent sample size for what you would have with two equal samples. The harmonic mean sample size is given by this formula:

$$\text{Harmonic Mean} = \frac{(2)(N_1)(N_2)}{N_1 + N_2} \qquad (10)$$

harmonic mean Special average influenced more by smaller numbers; in a *t* test for independent means when the number of scores in the two groups differ, the harmonic mean is used as the equivalent of each group's sample size when determining power.

The harmonic mean is two times the first sample size times the second sample size, all divided by the sum of the two sample sizes.

BOX 1 Two Women Make a Point about Gender and Statistics

One of the most useful advanced statistics books written so far is *Using Multivariate Statistics* by Barbara Tabachnick and Linda Fidell (2007), two experimental psychologists who worked at California State University at Northridge. These two met at a faculty luncheon soon after Tabachnick was hired. Fidell recalls:

> I had this enormous data set to analyze, and out came lots of pretty numbers in nice neat little columns, but I was not sure what all of it meant, or even whether my data had violated any critical assumptions. That was in 1975. I had been trained at the University of Michigan; I knew statistics up through the analysis of variance. But none of us were taught the multivariate analysis of variance at that time. Then along came these statistical packages to do it. But how to comprehend them?

Both Fidell and Tabachnick had gone out and learned on their own, taking the necessary courses, reading, asking others who knew the programs better, trying out what would happen if they did this with the data, what would happen if they did that. Now the two women asked each other, why must this be so hard? Were others reinventing this same wheel at the very same time? They decided to put their wheel into a book.

"And so began [many] years of conflict-free collaboration," reports Fidell. (That is something to compare to the feuds recounted in other boxes in this book.) The authors had no trouble finding a publisher, and the book, now in its fifth edition (Tabachnick & Fidell, 2007), has sold "nicely." In Fidell's opinion, statistics is a field in which women seem particularly to excel and feel comfortable.

Source: Personal interview with Linda Fidell.

In our example with 11 in one group and 31 in the other, the harmonic mean is 16.24:

$$\text{Harmonic Mean} = \frac{(2)(N_1)(N_2)}{N_1 + N_2} = \frac{(2)(11)(31)}{11 + 31} = \frac{682}{42} = 16.24$$

Thus, even though you have a total of 42 participants, the study has the power of a study with equal sample sizes of only about 16 in each group. (This means that a study with a total of 32 participants divided equally would have had about the same power.)

Planning Sample Size

Table 5 gives the approximate number of participants needed for 80% power for estimated small, medium, and large effect sizes, using one-tailed and two-tailed tests, all using the .05 significance level.[3] Suppose you plan a study in which you expect a medium effect size and will use the .05 significance level, one-tailed. Based on Table 5, you need 50 people in each group (100 total)

Table 5 Approximate Number of Participants Needed in Each Group (Assuming Equal Sample Sizes) for 80% Power for the *t* Test for Independent Means, Testing Hypotheses at the .05 Significance Level

	Effect Size		
	Small (.20)	**Medium (.50)**	**Large (.80)**
One-tailed	310	50	20
Two-tailed	393	64	26

[3]Cohen (1988, pp. 54–55) provides fuller tables, indicating needed numbers of participants for levels of power other than 80%; for effect sizes other than .20, .50, and .80; and for other significance levels.

to have 80% power. However, if you did a study using the same significance level but expected a large effect size, you would need only 20 people in each group (40 total).

How are you doing?

1. List two assumptions for the t test for independent means. For each, give the situations in which violations of these assumptions would be a serious problem.
2. Why do you need to assume the populations have the same variance?
3. What is the predicted effect size for a study in which the sample drawn from Population 1 has a mean of 17, Population 2's sample mean is 25, and the pooled estimate of the population standard deviation is 20?
4. What is the power of a study using a t test for independent means, with a two-tailed test at the .05 significance level, in which the researchers predict a large effect size and there are 20 participants in each group?
5. What is the approximate power of a study using a t test for independent means, with a two-tailed test at the .05 significance level, in which the researchers predict a large effect size, and there are 6 participants in one group and 34 participants in the other group?
6. How many participants do you need in each group for 80% power in a planned study in which you predict a small effect size and will be using a t test for independent means, one-tailed, at the .05 significance level?

Answers

1. One assumption is that the two populations are normally distributed; this is mainly a problem if you have reason to think the two populations are strongly skewed in opposite directions. A second assumption is that the two populations have the same variance; this is mainly a problem if you believe the two distributions have quite different variances *and* the sample sizes are different.
2. You need to assume the populations have the same variance because you make a pooled estimate of the population variance. The relatively straightforward pooling procedure we use would not make sense if the estimates from the two samples were for populations with different variances.
3. Predicted effect size $= (17 - 25)/20 = -8/20 = -.40$.
4. The power is .69.
5. Harmonic mean $= (2)(6)(34)/(6 + 34) = 408/40 = 10.2$. Power for a study like this with 10 in each group $= .39$.
6. 310 participants.

Review and Comparison of the Three Kinds of *t* Tests

You should now be familiar with three kinds of t tests: the t test for a single sample, the t test for dependent means, and the t test for independent means. Table 6 provides a review and comparison of these three kinds of t tests. As you can see in Table 6, the population variance is not known for each test, and the shape of the comparison distribution for each test is a t distribution. The single sample t test is used for hypothesis testing when you are comparing the mean of a single sample to a known population mean. However, in most research in the behavioral and social sciences, you do not know the population mean. With an unknown population mean, the t test for dependent means

Table 6 Review of the Three Kinds of *t* Tests

	Type of *t* Test		
Feature of the *t* Tests	**Single Sample**	**Dependent Means**	**Independent Means**
Population variance is known	No	No	No
Population mean is known	Yes	No	No
Number of scores for each participant	1	2	1
t Test carried out on difference scores	No	Yes	No
Shape of comparison distribution	*t* distribution	*t* distribution	*t* distribution
Formula for degrees of freedom	$df = N - 1$	$df = N - 1$	$df_{Total} = df_1 + df_2$ $(df_1 = N_1 - 1;$ $df_2 = N_2 - 1)$
Formula for *t*	$t = (M - \text{Population } M)/S_M$	$t = (M - \text{Population } M)/S_M$	$t = (M_1 - M_2)/S_{Difference}$

TIP FOR SUCCESS
We recommend that you spend some time carefully going through Table 6. Test your understanding of the three kinds of *t* tests by covering up portions of the table and trying to recall the hidden information.

is the appropriate *t* test when each participant has two scores (such as a before score and an after score) and you want to see if, on average, there is a difference between the participants' pairs of scores. The *t* test for independent means is used for hypothesis testing when you are comparing the mean of scores from one group of individuals (such as an experimental group) with the mean of scores from a different group of individuals (such as a control group).

The *t* Test for Independent Means in Research Articles

A *t* test for independent means is usually described in research articles by giving the means (and sometimes the standard deviations) of the two samples, plus the usual way of reporting any kind of *t* test—for example, $t(18) = 4.72$, $p < .01$. (Recall that the number in parentheses is the total degrees of freedom.) The result of the study of the health effects of expressive writing might be written up as follows: "The mean level of self-reported health in the expressive writing group was 79.00 $(SD = 9.72)$, and the mean for the control writing group was 68.00 $(SD = 10.55)$, $t(18) = 2.42$, $p < .05$." (In most cases, articles do not say whether a test is two-tailed; they usually only mention tails when it is a one-tailed test.)

Here is another example. Muise, Christofides, and Desmarais (2009) conducted a study of use of the social networking Web site Facebook among 308 university students. The students completed survey items about their use of Facebook, including a 27-item measure of Facebook jealousy (with scores ranging from a low of 1 to a high of 7) created by the researchers (e.g., "How likely are you to monitor your partner's activities on Facebook?"). Here is an excerpt from the results section of the study: "Participants in the current sample reported spending an average of 38.93 minutes on Facebook each day $(SD = 32.13)$ and had between 25 and 1,000 Facebook friends $(M = 296.19, SD = 173.04)$. . . . Women, $M = 40.57, SD = 26.76$, in our sample spent significantly more time on Facebook than men, $M = 29.83, SD = 23.73$; $t(305) = -3.32$, $p < 0.01$, and women, $M = 3.29, SD = 1.24$, score significantly higher on Facebook

Table 7 Mean and Standard Deviation of Scores for Women and Men on Measures of Machismo, Attitudes toward Women, and Adoption Beliefs

	Women (n = 64)	Men (n = 88)	t	p
Machismo	1.17 ± .15	1.32 ± .20	4.77	$<.001$
AWSA	3.26 ± .31	2.98 ± .35	5.00	$<.001$
Adoption	3.10 ± .39	2.85 ± .41	3.07	$<.01$

Source: Gibbons, J. L., Wilson, S. L., & Rufener, C. A. (2006). Gender attitudes mediate gender differences in attitudes towards adoption in Guatemala. *Sex Roles, 54,* 139–145.

jealousy than men, $M = 2.81$, $SD = 1.09$; $t(305) = -3.32$, $p < 0.01$" (p. 442). How do these results compare to the experiences that you and your friends have using Facebook?

Table 7 is an example in which the results of several t tests are given in a table. This table is taken from a study conducted by Gibbons, Wilson, and Rufener (2006). In that study, 152 college students in Guatemala were surveyed on their beliefs about machismo (a strong sense of masculinity), their attitudes toward women, and their beliefs about adoption. As shown in Table 7, the researchers used three t tests for independent means to examine whether female and male students differed on these beliefs and attitudes. The scales were scored so that higher scores were for more positive attitudes about machismo, more egalitarian (equal) gender beliefs (which were measured using the Attitudes Towards Women Scale for Adolescents, abbreviated as AWSA in Table 7), and more favorable beliefs about adoption. The first line of the table shows that men (with a mean score of 1.32) had more positive attitudes about machismo than women (mean score of 1.17). The t score for this comparison was 4.77, and it was statistically significant at $p < .001$. The results in Table 7 also show that women had more positive attitudes toward women than men did and that women had more favorable beliefs regarding adoption than men. (The number after each ± sign is the standard deviation for that particular group.)

Learning Aids

1. A t test for independent means is used for hypothesis testing with scores from two entirely separate groups of people. The comparison distribution for a t test for independent means is a distribution of differences between means of samples. This distribution can be thought of as being built up in two steps: Each population of individuals produces a distribution of means, and then a new distribution is created of differences between pairs of means selected from these two distributions of means.
2. The distribution of differences between means has a mean of 0 and is a t distribution with the total of the degrees of freedom from the two samples. Its standard deviation is figured in several steps:
   Ⓐ Figure the estimated population variances based on each sample.
   Ⓑ Figure the pooled estimate of the population variance.

Ⓒ Figure the variance of each distribution of means.
Ⓓ Figure the variance of the distribution of differences between means.
Ⓔ Figure the standard deviation of the distribution of differences between means.

3. The assumptions of the t test for independent means are that the two populations are normally distributed and have the same variance. However, the t test gives fairly accurate results when the true situation is moderately different from the assumptions.
4. Estimated effect size for a t test for independent means is the difference between the samples' means divided by the pooled estimate of the population standard deviation. Power is greatest when the sample sizes of the two groups are equal. When they are not equal, you use the harmonic mean of the two sample sizes when looking up power on a table. Power for a t test for independent means can be determined using a table (see Table 4), a power software package, or an Internet power calculator.
5. t tests for independent means are usually reported in research articles with the means of the two groups plus the degrees of freedom, t score, and significance level. Results may also be reported in a table where each significant difference is shown by asterisks.

t test for independent means
distribution of differences between means
pooled estimate of the population variance (S^2_{Pooled})
weighted average
variance of the distribution of differences between means ($S^2_{\text{Difference}}$)
standard deviation of the distribution of differences between means ($S_{\text{Difference}}$)
harmonic mean

Figuring the Standard Deviation of the Distribution of Differences between Means

Figure $S_{\text{Difference}}$ for the following study: $N_1 = 40$, $S_1^2 = 15$; $N_2 = 60$; $S_2^2 = 12$.

Answer

Ⓐ **Figure the estimated population variances based on each sample:** $S_1^2 = 15$; $S_2^2 = 12$.

Ⓑ **Figure the pooled estimate of the population variance:** $df_1 = N_1 - 1 = 40 - 1 = 39$; $df_2 = N_2 - 1 = 60 - 1 = 59$; $df_{\text{Total}} = df_1 + df_2 = 39 + 59 = 98$.

$$S^2_{\text{Pooled}} = \frac{df_1}{df_{\text{Total}}}(S_1^2) + \frac{df_2}{df_{\text{Total}}}(S_2^2) = (39/98)(15) + (59/98)(12)$$
$$= 5.97 + 7.22 = 13.19$$

Ⓒ **Figure the variance of each distribution of means:** $S^2_{M_1} = S^2_{\text{Pooled}}/N_1 = 13.19/40 = .33$; $S^2_{M_2} = S^2_{\text{Pooled}}/N_2 = 13.19/60 = .22$

⓫ **Figure the variance of the distribution of differences between means:** $S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2} = .33 + .22 = .55$

⓬ **Figure the standard deviation of the distribution of differences between means:** $S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}} = \sqrt{.55} = .74$

Hypothesis Testing Using the *t* Test for Independent Means

A researcher randomly assigns five individuals to receive a new experimental procedure and five to a control condition. At the end of the study, all 10 are measured. Scores for those in the experimental group were 7, 6, 9, 7, and 6. Scores for those in the control group were 5, 2, 4, 3, and 6. Carry out a *t* test for independent means using the .05 level of significance, two-tailed. Use the steps of hypothesis testing and sketch the distributions involved.

Answer

The figuring is shown in Table 8; the distributions are shown in Figure 5. 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: People like those who receive the experimental procedure.
Population 2: People like those who receive the control procedure.

Table 8 Figuring for Example Worked-Out Problem for Hypothesis Testing Using the *t* Test Independent Means

Experimental Group			Control Group		
Score	Deviation From Mean	Squared Deviation From Mean	Score	Deviation From Mean	Squared Deviation From Mean
7	0	0	5	1	1
6	−1	1	2	−2	4
9	2	4	4	0	0
7	0	0	3	−1	1
6	−1	1	6	2	4
Σ: 35	0	6	20	0	10

$M_1 = 7$; $S^2_1 = 6/4 = 1.50$; $M_2 = 4$; $S^2_2 = 10/4 = 2.50$

$N_1 = 5$; $df_1 = N_1 - 1 = 4$; $N_2 = 5$; $df_2 = N_2 - 1 = 4$

$df_{\text{Total}} = df_1 + df_2 = 4 + 4 = 8$

$S^2_{\text{Pooled}} = \frac{df_1}{df_{\text{Total}}}(S^2_1) + \frac{df_2}{df_{\text{Total}}}(S^2_2) = \frac{4}{8}(1.50) + \frac{4}{8}(2.50) = .75 + 1.25 = 2.00$

$S^2_{M_1} = S^2_{\text{Pooled}}/N_1 = 2.00/5 = .40$

$S^2_{M_2} = S^2_{\text{pooled}}/N_2 = 2.00/5 = .40$

$S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2} = .40 + .40 = .80$

$S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}} = \sqrt{.80} = .89$

Needed t with $df = 8$, 5% level, two-tailed $= \pm 2.306$

$t = (M_1 - M_2)/S_{\text{Difference}} = (7 - 4)/.89 = 3/.89 = 3.37$

Decision: Reject the null hypothesis; the research hypothesis is supported.

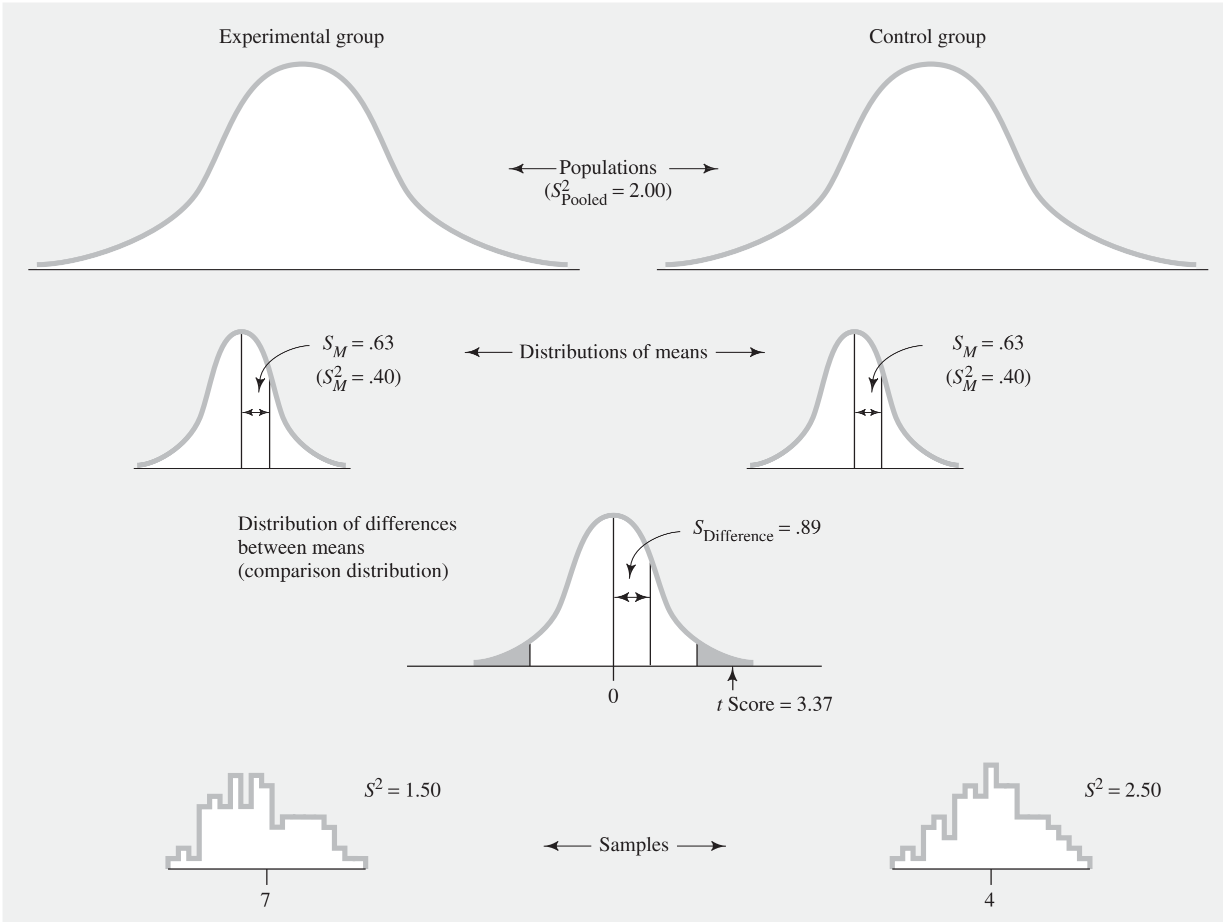


Figure 5 The distributions in the Example Worked-Out Problem for hypothesis testing using the *t* test for independent means.

The research hypothesis is that the means of the two populations are different. The null hypothesis is that the means of the two populations are the same.

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

a. The distribution of differences between means has a mean of 0.

b. Figure its standard deviation.

Ⓐ **Figure the estimated population variances based on each sample:** $S^2_1 = 1.50$; $S^2_2 = 2.50$.

Ⓑ **Figure the pooled estimate of the population variance:** $S^2_{\text{Pooled}} = 2.00$.

Ⓒ **Figure the variance of each distribution of means:** $S^2_{M_1} = .40$; $S^2_{M_2} = .40$.

Ⓓ **Figure the variance of the distribution of differences between means:** $S^2_{\text{Difference}} = .80$.

Ⓔ **Figure the standard deviation of the distribution of differences between means:** $S_{\text{Difference}} = .89$.

c. The shape of the comparison distribution is a *t* distribution with $df_{\text{Total}} = 8$.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** With significance level, two-tailed test, the cutoffs are 2.306 and −2.306.

❹ **Determine the sample's score on the comparison distribution.** $t = (7 - 4)/.89 = 3.37$.

❺ **Decide whether to reject the null hypothesis.** The t of 3.37 is more extreme than the cutoffs of $\pm$ 2.306. Thus, you can reject the null hypothesis. The research hypothesis is supported.

Finding Power When Sample Sizes Are Unequal

A planned study with a predicted small effect size has 22 in one group and 51 in the other. What is the approximate power for a one-tailed test at the .05 significance level?

Answer

$$\text{Harmonic mean} = \frac{(2)(N_1)(N_2)}{N_1 + N_2} = \frac{(2)(22)(51)}{22 + 51} = \frac{2244}{73} = 30.7.$$

From Table 4, for a one-tailed test with 30 participants in each group, power for a small effect size is .19.

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

1. Describe the core logic of hypothesis testing in this situation. Be sure to mention that the t test for independent means is used for hypothesis testing when you have scores from two entirely separate groups of people. Be sure to explain the meaning of the research hypothesis and the null hypothesis in this situation.
2. Explain the entire complex logic of the comparison distribution that is used with a t test for independent means (the distribution of differences between means). Be sure to explain why you use 0 as its mean. (This and point 3 will be the longest part of your essay.)
3. Outline the logic of estimating the population variance and the variance of the two distributions of means. Describe how to figure the standard deviation of the distribution of differences between means.
4. Explain why the shape of the comparison distribution that is used with a t test for independent means is a t distribution (as opposed to the normal curve).
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. For each of the following studies, say whether you would use a t test for dependent means or a t test for independent means.
   (a) An education researcher randomly assigns a group of 47 fourth-grade students to receive a new after-school math program and 48 other fourth-grade students to receive the standard after-school math program, and then measures how well they all do on a math test.
   (b) A marketing researcher measures 100 physicians' reports of the number of their patients asking them about a particular drug during the month before and the month after a major advertising campaign for that drug.
   (c) A researcher tests reaction time of each member of a group of 14 individuals twice, once while in a very hot room and once while in a normal-temperature room.
2. Figure $S_{\text{Difference}}$ for each of the following studies:

Study	N_1	S_1^2	N_2	S_2^2
(a)	20	1	20	2
(b)	20	1	40	2
(c)	40	1	20	2
(d)	40	1	40	2
(e)	40	1	40	4

3. For each of the following experiments, decide whether the difference between conditions is statistically significant at the .05 level (two-tailed).

	Experimental Group			Control Group		
Study	N	M	S^2	N	M	S^2
(a)	30	12.0	2.4	30	11.1	2.8
(b)	20	12.0	2.4	40	11.1	2.8
(c)	30	12.0	2.2	30	11.1	3.0

4. A communication researcher randomly assigned 82 volunteers to one of two experimental groups. Sixty-one were instructed to get their news for a month only from television, and 21 were instructed to get their news for a month only from the Internet. (Why the researcher didn't assign equal numbers to the two conditions is a mystery!) After the month was up, all participants were tested on their knowledge of several political issues. The researcher did not have a prediction as to which news source would make people more knowledgeable. That is, the researcher simply predicted that there is some kind of difference. These were the results of the study: TV group: $M = 24$, $S^2 = 4$; Internet group: $M = 26$, $S^2 = 6$. Using the .01 level, what should the researcher conclude? (a) Use the steps of hypothesis testing; (b) sketch the distributions involved; and (c) explain your answers to someone who is familiar with the t test for a single sample, but not with the t test for independent means.
5. A teacher was interested in whether using a student's own name in a story affected children's attention span while reading. Six children were randomly assigned to read a story under ordinary conditions (using names like Dick and Jane). Five other children read versions of the same story, but with each child's own name

substituted for one of the children in the story. The teacher kept a careful measure of how long it took each child to read the story. The results are in the following table. Using the .05 level, does including the child's name make any difference? (a) Use the steps of hypothesis testing; (b) sketch the distributions involved; and (c) explain your answers to someone who has never had a course in statistics.

Ordinary Story		Own-Name Story	
Student	**Reading Time**	**Student**	**Reading Time**
A	2	G	4
B	5	H	8
C	7	I	10
D	9	J	9
E	6	K	8
F	7		

6. A developmental researcher compares 4-year-olds and 8-year-olds on their ability to understand the analogies used in stories. The scores for the five 4-year-olds tested were 7, 6, 2, 3, and 8. The scores for the three 8-year-olds tested were 9, 2, and 5. Using the .05 level, do older children do better? (a) Use the steps of hypothesis testing; (b) sketch the distributions involved; and (c) explain your answers to someone who understands the t test for a single sample, but does not know anything about the t test for independent means.
7. Figure the estimated effect size for problems (a) 4, (b) 5, and (c) 6. (d) Explain what you have done in part (a) to someone who understands the t test for independent means but knows nothing about effect size.
8. Figure the approximate power of each of the following planned studies, all using a t test for independent means at the .05 significance level, one-tailed, with a predicted small effect size:

Study	N_1	N_2
(a)	3	57
(b)	10	50
(c)	20	40
(d)	30	30

9. What are the approximate numbers of participants needed for each of the following planned studies to have 80% power, assuming equal numbers in the two groups and all using the .05 significance level? (Be sure to give the total number of participants needed, not just the number needed for each group.)

	Expected Means		Expected	
Study	M_1	M_2	S_{Pooled}	**Tails**
(a)	107.0	149.0	84.0	1
(b)	22.5	16.2	31.5	2
(c)	14.0	12.0	2.5	1
(d)	480.0	520.0	50.0	2

10. Van Aken and Asendorpf (1997) studied 139 German 12-year-olds. All of the children completed a general self-worth questionnaire and were interviewed about the supportiveness they experienced from their mothers, fathers, and classmates. The researchers then compared the self-worth of those with high and low levels of support of each type. The researchers reported that "lower general self-worth was found for children with a low-supportive mother ($t(137) = 4.52$, $p < .001$, d [effect size] $= 0.78$) and with a low-supportive father ($t(137) = 4.03$, $p < .001$, d [effect size] $= 0.69$).... A lower general self-worth was also found for children with only low-supportive classmates ($t(137) = 2.04$, $p < .05$, d [effect size] $= 0.35$)." ("d" in the above is a symbol for effect size.) (a) Explain what these results mean to a person who has never had a course in statistics. (b) Include a discussion of effect size and power. (When figuring power, you can assume that the two groups in each comparison had about equal sample sizes.)
11. Gallagher-Thompson, Dal Canto, Jacob, and Thompson (2001) compared 27 wives who were caring for their husbands who had Alzheimer's disease to 27 wives in which neither partner had Alzheimer's disease. The two groups of wives were otherwise similar in terms of age, number of years married, and socioeconomic status. Table 9 (reproduced from their Table 1) shows some of their results. Focusing on the Geriatric Depression Scale (the first row of the table) and the Mutuality Scale for Shared Values (the last row in the table), explain these results to a person who knows about the t test for a single sample, but is unfamiliar with the t test for independent means.

Table 9 Comparison of Caregiving and Noncaregiving Wives on Select Psychosocial Variables

	Caregiving Wives ($n = 27$)			Noncaregiving Wives ($n = 27$)				
	M	*SD*	Range	*M*	*SD*	Range	*t*	*p*
Geriatric Depression Scale[a]	9.42	6.59	1–25	2.37	2.54	0–8	5.14	.0001
Perceived Stress Scale[b]	22.29	8.34	6–36	15.33	6.36	7–30	3.44	.001
Hope questionnaire[c]								
Agency	11.88	1.63	9–16	13.23	1.39	10–16	3.20	.002
Resilience	11.89	0.91	10–14	13.08	1.60	10–16	3.31	.002
Total	23.77	2.03	21–29	26.31	2.56	22–31	3.97	.0001
Mutuality Scale[d]								
Closeness	3.51	.81	.33–4	3.70	.41	2.67–4	−1.02	.315
Reciprocity	2.25	1.19	.17–4	3.25	.55	1.67–4	−3.68	.001
Shared pleasures	2.65	1.00	0–4	3.52	.61	1.75–4	−3.66	.001
Shared values	3.15	.89	0–4	3.46	.45	2.4–4	−1.51	.138

Note: For all measures, higher scores indicate more of the construct being measured.
[a]Maximum score is 30.
[b]Maximum score is 56.
[c]Four questions in each subscale, with a maximum total score of 32.
[d]Maximum mean for each subscale is 4.
Source: Gallagher-Thompson, D., Dal Canto, P. G., Jacob, T., & Thompson, L. W. (2001). A comparison of marital interaction patterns between couples in which the husband does or does not have Alzheimer's disease. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 56,* S140–S150. Copyright © 2001 by the Gerontological Society of America. Reprinted by permission of the publishers.

Set II

12. Make up two examples of studies (not in the book or from your lectures) that would be tested with a t test for independent means.
13. For each of the following studies, say whether you would use a t test for dependent means or a t test for independent means.
    (a) A researcher measures the heights of 40 university students who are the first-born in their families and compares the 15 who come from large families to the 25 who come from smaller families.
    (b) A researcher tests performance on a math skills test of each of 250 individuals before and after they complete a one-day seminar on managing test anxiety.
    (c) A researcher compares the resting heart rate of 15 individuals who have been taking a particular drug to the resting heart rate of 48 other individuals who have not been taking this drug.
14. Figure $S_{\text{Difference}}$ for each of the following studies:

Study	N_1	S_1^2	N_2	S_2^2
(a)	30	5	20	4
(b)	30	5	30	4
(c)	30	5	50	4
(d)	20	5	30	4
(e)	30	5	20	2

15. For each of the following experiments, decide whether the difference between conditions is statistically significant at the .05 level (two-tailed).

	Experimental Group			Control Group		
Study	N	M	S^2	N	M	S^2
(a)	10	604	60	10	607	50
(b)	40	604	60	40	607	50
(c)	10	604	20	40	607	16

16. A researcher theorized that people can hear better when they have just eaten a large meal. Six individuals were randomly assigned to eat either a large meal or a small meal. After eating the meal, their hearing was tested. The hearing ability scores (high numbers indicate greater ability) are given in the following table. Using the .05 level, do the results support the researcher's theory? (a) Use the steps of hypothesis testing, (b) sketch the distributions involved, and (c) explain your answers to someone who has never had a course in statistics.

Big Meal Group		Small Meal Group	
Subject	Hearing	Subject	Hearing
A	22	D	19
B	25	E	23
C	25	F	21

17. Twenty students randomly assigned to an experimental group receive an instructional program; 30 in a control group do not. After 6 months, both groups are tested on their knowledge. The experimental group has a mean of 38 on the test (with an estimated population standard deviation of 3); the control group has a mean of 35 (with an estimated population standard deviation of 5). Using the .05 level, what should the experimenter conclude? (a) Use the steps of hypothesis testing; (b) sketch the distributions involved; and (c) explain your answer to someone who is familiar with the *t* test for a single sample but not with the *t* test for independent means.
18. A study of the effects of color on easing anxiety compared anxiety test scores of participants who completed the test printed on either soft yellow paper or on harsh green paper. The scores for five participants who completed the test printed on the yellow paper were 17, 19, 28, 21, and 18. The scores for four participants who completed the test on the green paper were 20, 26, 17, and 24. Using the .05 level, one-tailed (predicting lower anxiety scores for the yellow paper), what should the researcher conclude? (a) Use the steps of hypothesis testing; (b) sketch the distributions involved; and (c) explain your answers to someone who is familiar with the *t* test for a single sample but not with the *t* test for independent means.
19. Figure the estimated effect size for problems (a) 16, (b) 17, and (c) 18. (d) Explain your answer to part (a) to a person who understands the *t* test for independent means but is unfamiliar with effect size.
20. What is the approximate power of each of the following planned studies, all using a *t* test for independent means at the .05 significance level, two-tailed, with a predicted medium effect size?

Study	N_1	N_2
(a)	90	10
(b)	50	50
(c)	6	34
(d)	20	20

21. What are the approximate numbers of participants needed for each of the following planned studies to have 80% power, assuming equal numbers in the two groups and all using the .05 significance level? (Be sure to give the total number of participants needed, not just the number needed for each group.)

Study	Expected Means M_1	Expected Means M_2	Expected S_{Pooled}	Tails
(a)	10	15	25	1
(b)	10	30	25	1
(c)	10	30	40	1
(d)	10	15	25	2

22. Escudero, Rogers, and Gutierrez (1997) videotaped 30 couples discussing a marital problem in their laboratory. The videotapes were later systematically rated for various aspects of the couple's communication, such as domineeringness and the positive or negative quality of affect (emotion) expressed between them. A major interest of their study was to compare couples who were having relationship problems with those who were not. The 18 couples in the group having problems were recruited from those who had gone to a marital clinic for help; they were called the Clinic group. The 12 couples in the group not having problems were recruited through advertisements and were called the Nonclinic group. (The two groups in fact had dramatically different scores on a standard test of marital satisfaction.) Table 10 presents some of their results. (You can ignore the arrows and plus and minus signs, which have to do with how they rated the interactions. Also, ignore the note at the bottom about "arcsine transformation.") (a) Focusing on Domineeringness and Submissiveness, explain these results to a person who has never had a course in statistics. (b) Include a discussion of effect size and power. (When figuring power, you can assume that the two groups in each comparison had about equal sample sizes.)
23. Jackson, Ervin, Gardner, and Schmitt (2001) gave a questionnaire about Internet usage to university students. Table 11 shows their results comparing men and women. (a) Select one significant and one nonsignificant result and explain these two results to a person who understands the *t* test for a single sample but does not know anything about the *t* test for independent means. (b) Include a discussion of effect size and power (note that the sample sizes for the male and female groups are shown in the table footnote).

Table 10 Base-Rate Differences between Clinic and Nonclinic Couples on Relational Control and Nonverbal Affect Codes Expressed in Proportions (*SD*s in Parentheses)

	Couple Status		Between-Group Differences
	Clinic Mean	Nonclinic Mean	*t*
Domineeringness (↑)	.452 (107)	.307 (.152)	3.06*
Levelingness (→)	.305 (.061)	.438 (.065)	5.77**
Submissiveness (↓)	.183 (.097)	.226 (.111)	1.12
Double-codes	.050 (.028)	.024 (.017)	2.92*
Positive affect (+)	.127 (.090)	.280 (.173)	3.22*
Negative affect (−)	.509 (.192)	.127 (.133)	5.38**
Neutral affect (0)	.344 (.110)	.582 (.089)	6.44**
Double-codes (+/−)	.019 (.028)	.008 (.017)	2.96*

Note: Proportions of each control and affect code were converted using arcsine transformation for use in between-group comparisons. $^{*}p < .01$, $^{**}p < .001$ (d.f. = 28).
Source: Escudero, V., Rogers, L. E., & Gutierrez, E. (1997). Patterns of relational control and nonverbal affect in clinic and nonclinic couples. *Journal of Social and Personal Relationships, 14,* 5–29.

Table 11 Gender Differences in Internet Use and Potential Mediators

	Males[a]	Females[b]	*t*-value	*df*	*p*-value
E-mail use	4.16 (0.66)	4.30 (0.57)	2.81	626	.005
Web use	3.57 (0.67)	3.30 (0.67)	−4.84	627	.000
Overall Internet use	3.86 (0.58)	3.80 (0.53)	−1.44	627	.130
Computer anxiety	1.67 (0.56)	1.80 (0.57)	4.03	612	.000
Computer self-efficacy	3.89 (0.52)	3.71 (0.62)	−3.49	608	.001
Loneliness	2.06 (0.64)	1.96 (0.64)	−1.88	607	.061
Depression	1.22 (0.32)	1.28 (0.34)	2.36	609	.019
E-mail privacy	4.04 (0.78)	4.10 (0.69)	−0.97	609	.516
E-mail trust	3.50 (0.77)	3.46 (0.75)	−0.65	610	.516
Web privacy	4.06 (0.74)	4.09 (0.71)	0.62	623	.534
Web trust	3.14 (0.73)	3.12 (0.73)	−0.28	624	.780
Web search success	4.05 (0.85)	4.13 (0.81)	1.12	568	.262
Importance of computer skills	2.54 (1.03)	2.31 (0.90)	−2.57	477	.011
Computers cause health problems	2.67 (1.00)	3.00 (1.08)	3.36	476	.001
Gender stereotypes about computer skills	3.45 (1.15)	4.33 (0.96)	−8.95	476	.000
Racial/ethnic stereotypes about computer skills	3.63 (1.17)	3.99 (1.07)	3.40	477	.001
Computers are taking over	3.08 (1.19)	2.87 (1.08)	−1.89	476	.059

Note: For the attitude items, 1 = strongly agree, 2 = strongly disagree. For gender, 1 = male, 2 = female. Numbers in parentheses are standard deviations.

[a] $n = 227$.

[b] $n = 403$.

Source: Jackson, L. A., Ervin, K. S., Gardner, P. D., & Schmitt, N. (2001). Gender and the Internet: Women communicating and men searching. *Sex Roles, 44,* 363–379.

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 Independent Means

It is easier to learn these steps using actual numbers, so we will use the expressive writing example from earlier in the chapter. The scores for that example are shown in Table 1.

❶ Enter the scores into SPSS. SPSS assumes that all scores in a row are from the same person. In this example, each person is in only one of the two groups (*either* the expressive writing group *or* the control writing group). Thus, in order to tell SPSS which person is in each group, you should enter the numbers as shown in Figure 6. In the first column (labeled "group"), we used the number "1" to indicate that a person is in the expressive writing group and the number "2" to indicate that a person is in the control writing group. Each person's score on the health measure is listed in the second column (labeled "health").

❷ ✐ *Analyze.*

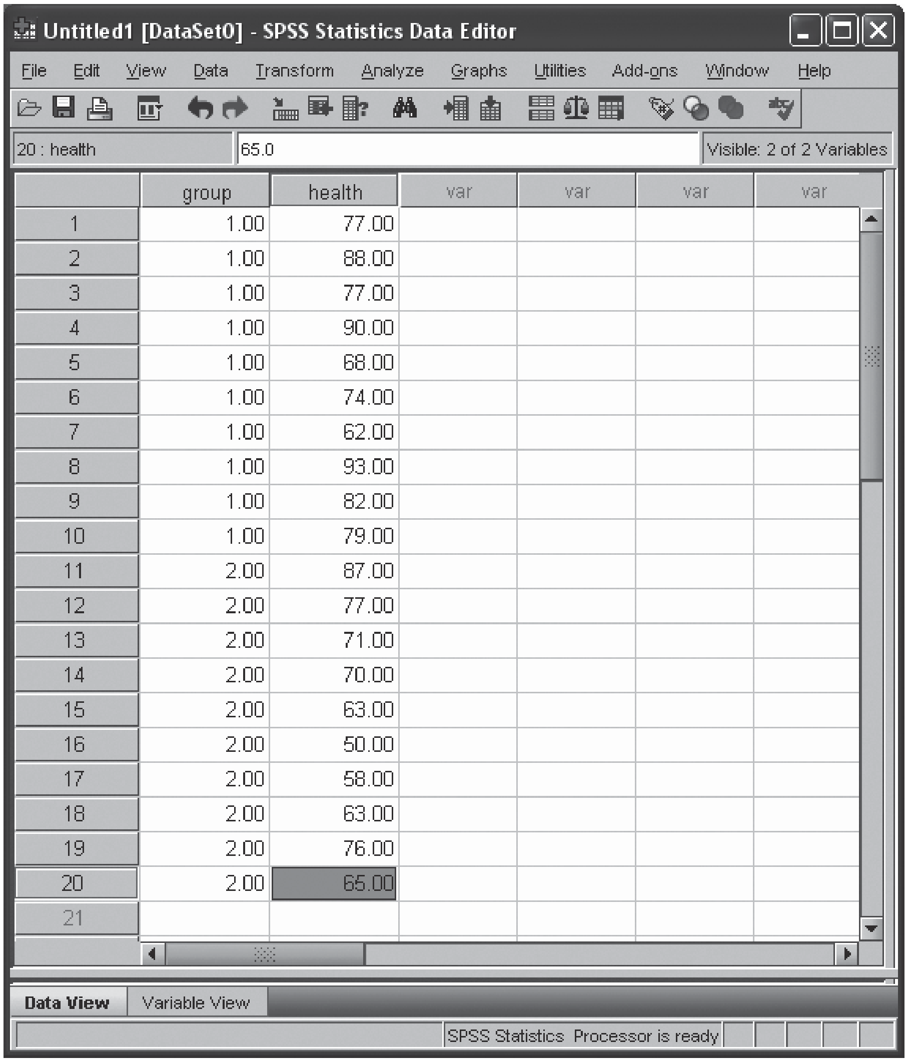


Figure 6 SPSS data editor window for the expressive writing example (in which 20 students were randomly assigned to be in an expressive writing group or a control writing group).

3 *Compare means.*
4 *Independent-Samples T Test* (this is the name SPSS uses for a *t* test for independent means).
5 on the variable called "health" and then the arrow next to the box labeled "Test Variable(s)." This tells SPSS that the *t* test should be carried out on the scores for the "health" variable.
6 the variable called "group" and then the arrow next to the box labeled "Grouping Variable." This tells SPSS that the variable called "group" shows which person is in which group. *Define Groups.* You now tell SPSS the values you used to label each group. Put "1" in the Group 1 box and put "2" in the Group 2 box. Your screen should now look like Figure 7. *Continue.*
7 *OK.* Your SPSS output window should look like Figure 8.

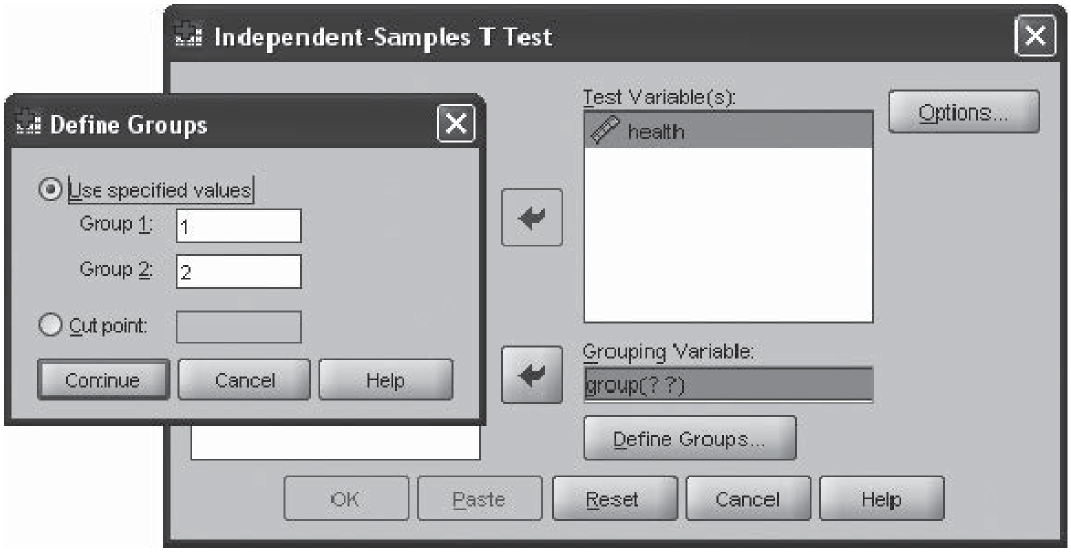
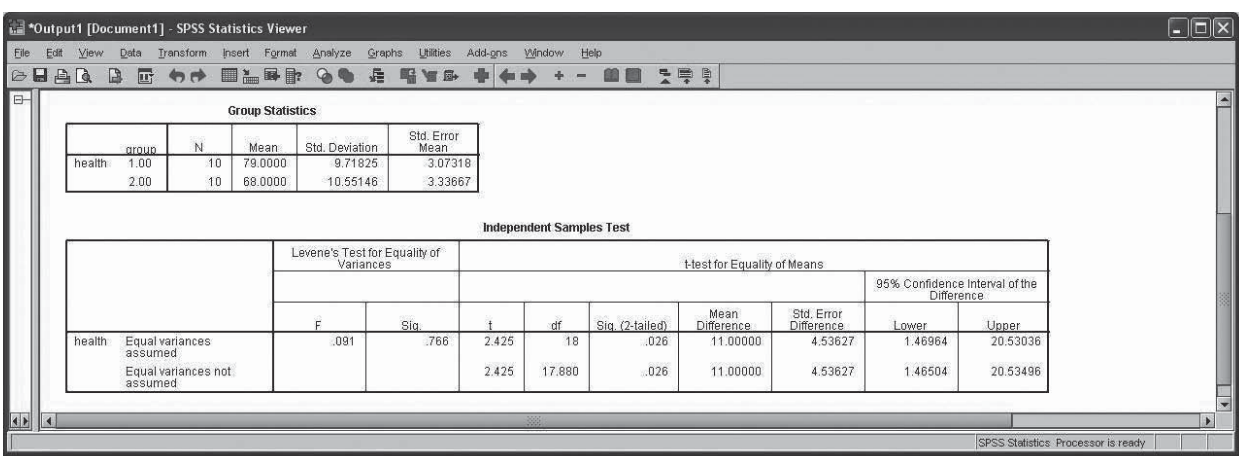


Figure 7 SPSS independent means *t* test window for the expressive writing example.

The first table in the SPSS output provides information about the two variables. The first column gives the values of the grouping variable (1 and 2, which indicate the expressive writing group and the control writing group, respectively). The second, third, and fourth columns give, respectively, the number of individuals (N), mean (M), and estimated population standard deviation (S) for each group. The fifth column, labeled "Std. error mean," is the standard deviation of the distribution of means, S_M, for each group. Note that these values for the standard error of the mean are based on each population variance estimate and not on the pooled estimate, so they are not quite the same for each group as the square root of each S_M^2 figured in the text. (See Table 1 for the figuring for this example.)


*Output1 [Document1] - SPSS Statistics Viewer

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
health	1.00	10	79.0000	9.71825	3.07318
	2.00	10	68.0000	10.55146	3.33667

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
health	Equal variances assumed	.091	.766	2.425	18	.026	11.00000	4.53627	1.46964	20.53036
	Equal variances not assumed			2.425	17.880	.026	11.00000	4.53627	1.46504	20.53496

SPSS Statistics Processor is ready

Figure 8 SPSS output window for a *t* test for independent means for the expressive writing example.

The second table in the SPSS output shows the actual results of the t test for independent means. Before the t test results, SPSS shows the results of "Levene's Test for Equality of Variances," which is a test of whether the *variances* of the two populations are the same. This test is important mainly as a check on whether you have met the assumption of equal population variances. (Recall the earlier section in this chapter, "Assumptions of the t Test for Independent Means.") If Levene's test is significant (that is, if the value in the "Sig." column is less than .05), this assumption of equal population variances is brought into question. However, in this example, the result is clearly not significant (.766 is well above .05), so we have no reason to doubt the assumption of equal population variances. Thus, we can feel more confident that whatever conclusion we draw from the t test will be accurate.

The t test results begin with the column labeled "t." Note that there are two rows of t test results. The first row (a t of 2.425, *df* of 18, etc.), labeled "Equal variances assumed" (on the left-hand side of the table), shows the t test results assuming the population variances are equal. The second row (a t of 2.425, *df* of 17.880, etc.), labeled "Equal variances not assumed," shows the t test results if we do not assume that the population variances are equal. In the present example (as in most real-life cases) Levene's test was not significant, so we use the t test results assuming equal population variances. Compare the values for "t" (the sample's t score), "df" (degrees of freedom), and "Std. Error Difference" (the standard deviation of the distribution of differences between means, $S_{\text{Difference}}$) from Figure 8 with their respective values we figured by hand in Table 1. The column labeled "Sig. (2-tailed)" shows the exact significance level of the sample's t score. The significance level of .026 is less than our .05 cutoff for this example, which means that you can reject the null hypothesis and the research hypothesis is supported. (You can ignore the final two columns of the table, listed under the heading "95% Confidence Interval of the Difference." These columns refer to the raw scores corresponding to the t scores at the bottom 2.5% and the top 2.5% of the t distribution.) Note that SPSS does not know if you are doing a one-tailed or a two-tailed test. So it always gives results for a two-tailed test. If you are doing a one-tailed test, the true significance level is exactly half of what is given by SPSS. (For example, if the significance level given were .16, this would be equivalent to .08 for a one-tailed test.)

1. (a) Independent; (b) dependent; (c) dependent.
2. (a) $S^2_{\text{Pooled}} = (df_1/df_{\text{Total}})(S^2_1) + (df_2/df_{\text{Total}})(S^2_2)$
   $= (19/38)(1) + (19/38)(2) = 1.5;$
   $S^2_{M_1} = S^2_{\text{Pooled}}/N_1 = 1.5/20 = .075$; $S^2_{M_2} = .075$;
   $S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2} = .075 + .075 = .15$;
   $S_{\text{Difference}} = .39$.
   (b) .35; (c) .32; (d) .27; (e) .35.
3. (a) t needed $(df = 58, p < .05, \text{two-tailed}) = -2.004$, 2.004;
   $S^2_{\text{Pooled}} = (df_1/df_{\text{Total}})(S^2_1) + (df_2/df_{\text{Total}})(S^2_2)$
   $= (29/58)(2.4) + (29/58)(2.8) = 2.6;$
   $S^2_{M_1} = S^2_{\text{Pooled}}/N_1 = 2.6/30 = .087$;
   $S^2_{M_2} = .087$;
   $S^2_{\text{Difference}} = S^2_{M_2} + S^2_{M_2} = .087 + .087 = .174$.
   $S_{\text{Difference}} = .417$;
   $t = (M_1 - M_2)/S_{\text{Difference}} = (12 - 11.1)/.417 = 2.16$.
   Conclusion: Reject the null hypothesis. The difference is significant. (b) $S^2_{\text{Pooled}} = 2.67$; $S_{\text{Difference}} = .45$; $t = 2.00$; do not reject the null hypothesis. (c) $S^2_{\text{Pooled}} = 2.6$; $S_{\text{Difference}} = .416$; $t = 2.16$; reject the null hypothesis.
4. (a)
   ❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are two populations of interest:

   Population 1: People who get their news from TV.
   Population 2: People who get their news from the Internet.

The research hypothesis is that the two populations have different means. The null hypothesis is that the two populations have the same mean.

❷ **Determine the characteristics of the comparison distribution.** $S^2_{\text{Pooled}} = (60/80)(4) + (20/80)(6) = 4.5$; comparison distribution (distribution of differences between means): $M = 0$; $S_{\text{Difference}} = .54$; shape $= t(df = 80)$.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** t needed $(df = 80, p < .01$, two-tailed$) = -2.639, 2.639$.

❹ **Determine your sample's score on the comparison distribution.** $t = (24 - 26)/.54 = -3.70$.

❺ **Decide whether to reject the null hypothesis.** -3.70 is more extreme than -2.639; reject the null hypothesis; the prediction that there will be some difference is supported by the experiment.

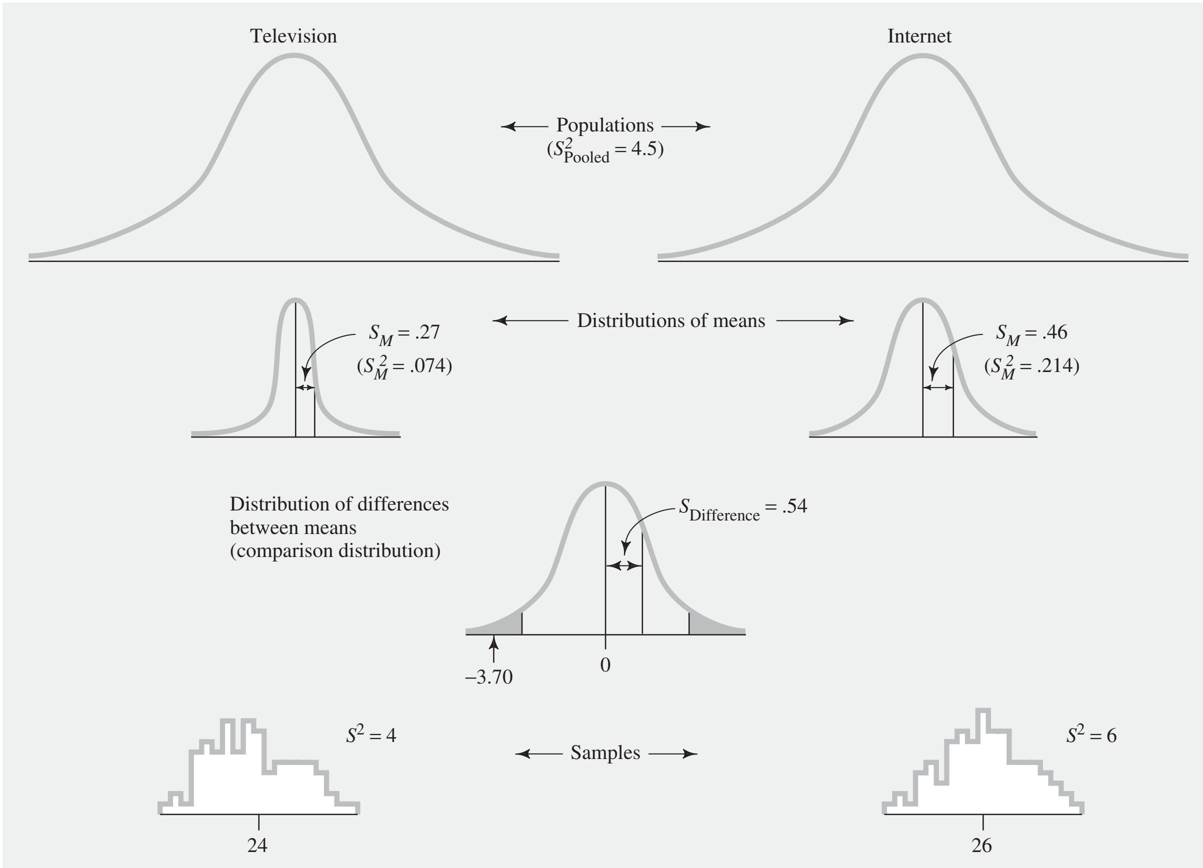
(b) See figure on top of next page.

(c) In this situation, I am testing whether the two samples come from identical populations. I have two estimates of those identical populations, one from each sample. Thus, to get the most accurate overall estimate of the population variance, I can average the two estimates of the population variance. In order to give more weight to the estimate based on the larger degrees of freedom, I figure a weighted average, multiplying each estimate by its proportion of the total degrees of freedom, then adding up the results. This pooled estimate of the population variance comes out to 4.5.

I was interested not in individual scores but in the difference between the mean of a group of 61 and the mean of another group of 21. Thus, I needed to figure out what would be the characteristics of a distribution of differences between means, specifically a distribution of differences between means of groups of 61 and 21 that are randomly taken from the two identical populations whose variance I just estimated. This required two steps. First, I figured the characteristics of the distribution of means in the usual way for the population associated with each sample, but using my pooled estimate for the variance of each population of individuals. This came out to .074 for the TV group and .214 for the Internet group. The second step is directly about the distribution of differences between means. It is like a distribution you would get if you took a mean from the distribution of means for the TV group and took one from the distribution of means for the Internet group and figured the difference between these two means. After doing this many times, the distribution of these differences between two means would make up a new distribution, called a distribution of differences between means. This distribution is the comparison distribution in a t test for independent means. We are assuming (if Internet vs. TV made no difference) that the two original populations have the same means. Thus, on average, the difference between a mean taken from the TV group and a mean taken from the Internet group should come out to 0. (This is because sometimes one will be bigger and sometimes the other, but in the long run these random fluctuations should balance out.) The variance of this distribution of differences between means is affected by the variation in both distributions of means. In fact, it is just the sum of the two. Thus, its variance is .074 plus .214, or .288. Its square root, the standard deviation of the distribution of differences between means, is .54.

Because this whole process is based on estimated variances, the distribution of means is a t distribution. Specifically, it will be a t distribution with degrees of freedom equal to the total number of degrees of freedom that went into the two estimates. In this example, $df = 80$. Looking this up on the t table for .01 two-tailed gives a cutoff needed of ± 2.639. The t for my sample is the particular difference between the means of my two groups, divided by the standard deviation of the distribution of differences between means. This comes out to $(24 - 26)/.54 = -3.70$. This is more extreme than the cutoff, so I can reject the null hypothesis.

5. (a), (b), and (c). Hypothesis-testing steps, sketch, and explanation similar to 4 above (except for c you also need to include basic material).
   t needed $= -2.262, 2.262$; $S^2_{\text{Pooled}} = (5/9)(5.60) + (4/9)(5.20) = 5.42$; $S^2_{\text{Difference}} = .90 + 1.08 = 1.98$; $S_{\text{Difference}} = 1.41$; $t = (6 - 7.8)/1.41 = -1.28$. Do not reject the null hypothesis; the experiment is inconclusive as to whether including the child's name makes a difference.
6. (a), (b), and (c). Hypothesis-testing steps, sketch, and explanation similar to 4 above. t needed $= 1.943$; $S^2_{\text{Pooled}} = (4/6)(6.7) + (2/6)(12.33) = 8.58$; $S^2_{\text{Difference}} = 2.86 + 1.72 = 4.58$; $S_{\text{Difference}} = 2.14$; $t = (5.33 - 5.2)/2.14 = .06$. Do not reject the null hypothesis; the experiment is inconclusive as to whether older children do better.
7. (a) Estimated effect size $= (M_1 - M_2)/S_{\text{Pooled}} = (24 - 26)/\sqrt{4.5} = -.94$; (b) $-.77$; (c) .04. (d) Effect size represents the degree to which two populations do not overlap. The less that two populations overlap, the larger the effect size. In the behavioral and social sciences, we often want to know not just if a result is significant, but how big the effect is. Effect size gives you a measure of how big the effect is. Effect size for the t test for independent means is the difference between the population means, divided by the standard deviation of the population of individuals. However, you do not know the population means. Thus, you estimate them using the sample means. You also do not know the standard deviation of the population of individuals. Thus, you have to estimate it too. You do this by using the pooled estimate of the population standard deviation. So, in part (a), the effect size was the difference between the sample means (24 minus 26, which gives -2), divided by the pooled estimate of the population standard deviation (which was the square root of 4.5, or 2.12). This gave an estimated effect size of $-.94$, which is a large effect size according to Cohen's effect size conventions for the t test for independent means.
8. (a) Harmonic mean $= 2(N_1)(N_2)/(N_1 + N_2) = (2)(3)(57)/(3 + 57) = 5.7$; from Table 4 approximate power $= .11$; (b) harmonic mean $= 16.7$, power $= .15$; (c) harmonic mean $= 26.7$, power $= .19$; (d) power $= .19$.
9. (a) Estimated effect size $= (M_1 - M_2)/S_{\text{Pooled}} = (107 - 149)/84 = -50$, medium effect size; needed N from Table 5 $= 50$ per group, 100 total. (b) Estimated effect size $= .20$; needed $N = 393$ per group, 786 total. (c) Estimated effect size $= .80$; needed $N = 20$ per group, 40 total. (d) Estimated effect size $= -.80$; needed $N = 26$ per group, 52 total.



10. (Along with the following, include a full explanation of all terms and concepts as in 4c .) (a) and (b) This study shows that using a conventional .05 significance level, German children who receive low levels of support—whether from their mother, father, or classmates—showed lower levels of self-worth. Furthermore, the effect sizes were fairly large (.78 and .69) with regard to support from mother or father; however, the effect size was only small to moderate (.35) with regard to support from classmates. This would seem to imply that support from parents is more important than support from classmates in terms of a child's feeling of self-worth. The power of the study for a large effect size is .98. (This assumes there were about equal numbers of children in the high and low support groups, that the test is two-tailed, and uses the figure for 50 in each group.) The power for a medium effect size is .70. Because we already know that the results are significant and we know the effect sizes, the power calculations are not very important.
11. Similar to 4c above, but focusing on the results of this study.

***t* test for independent 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)** Its mean will be 0. **(b)** Figure its standard deviation:

Ⓐ For each population, $S^2 = [\Sigma(X - M)^2]/(N - 1)$

Ⓑ $S^2_{Pooled} = (df_1/df_{Total})(S^2_1) + (df_2/df_{Total})(S^2_2)$; $df_1 = N_1 - 1$

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

Ⓒ $S^2_{M_1} = S^2_{\text{Pooled}}/N_1$; $S^2_{M_2} = S^2_{\text{Pooled}}/N_2$

Ⓓ $S^2_{\text{Difference}} = S^2_{M_1} + S^2_{M_2}$

Ⓔ $S_{\text{Difference}} = \sqrt{S^2_{\text{Difference}}}$

(c) t distribution, degrees of freedom $= df_{\text{Total}}$.

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.** $t = (M_1 - M_2)/S_{\text{Difference}}$

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