8

Hypothesis Testing With Two Samples (Mean Difference and Difference of Means)

In this chapter, we expand on the logic we covered in the last chapter. Instead of working with single sample research questions, we'll now take up the two sample case. The applications we're going to cover are typically referred to as difference of means tests. As before, we'll be looking at information collected from samples, but our real interest will extend to the realm of two populations. You already have the background to deal with difference of means tests on a conceptual level. From the formulation of a null hypothesis, to the identification of a critical value, through the conclusion and interpretation of results, you've traveled the road of hypothesis-testing basics. What's more, you'll likely discover that the research questions involving difference of means tests are very straightforward.

The research situations that call for a difference of means test typically fall into two categories: situations involving matched or related samples, and situations involving independent samples. That's the order we'll follow in this chapter.

Before We Begin

I'm going to make a suggestion that you may or may not appreciate. I'm going to ask you to go back to the previous chapter as soon as you finish this chapter. That's right; first complete this chapter and then take another look at the previous chapter. I know that may sound strange, but there's a reason why I suggest it. I believe that many students of statistics have an easier time understanding hypothesis testing involving two samples than hypothesis testing involving one sample. As to why that might be the case, maybe it's because the idea of comparing two groups (or more correctly, two samples) is such a commonplace activity, whether it is warranted or not. At any rate, my experience tells me that it is with the two sample applications that students typically start to get comfortable with hypothesis testing. Therefore, if you struggled with the last chapter or you still don't feel on solid ground with it, let me urge you to move ahead with this chapter. Once you're through, go back to the previous chapter.

Related Samples

To say that two samples are **matched or related** in some way is to say that the cases included in the samples were not selected independently of one another. Let me give you a few examples.

Let's say we give a group of students a drug awareness test, measuring their awareness of the dangers of recreational drug use. Then we show the students a film about the dangers of recreational drug use, and we administer the test again. In this case, we will be measuring the same people twice—once before exposure to the film and again after exposure to the film. This is an example of a classic before/after test situation—a situation in which the participants are matched against themselves, so to speak. Each person is tested twice, and the focus is on any change in the test scores that occurs between the first and second tests. The two samples are obviously related to each other; they involve the same people, measured twice.

Another example might involve measuring attitudes of couples who are engaged to one another. Let's say that we're collecting information about their expectations concerning marriage, placing the males in one sample and the females in another. In this situation, the cases or people in one sample are clearly linked to the cases in the other. If the research question looked at differences between men and women in general with respect to expectations about marriage, the problem could be approached without a matched or related sample design. When the focus is on couples, however, and how the male and female members of couples differ, the question calls for a matched or related sample approach.

For a final example, consider a research situation that involves comparing the productivity of two groups of workers in different environments (such as a 5-day, 8-hour work schedule versus a 4-day, 10-hour work schedule). If research participants are selected in such a way that they are matched, for example, on the basis of age, sex, pay grade, and length of time on the job, it would be appropriate to treat the research situation as a matched sample design.

☑ LEARNING CHECK

Question: What are matched or related samples?

Answer: Matched or related samples are samples involving cases or subjects that share certain characteristics in common.

In all of these examples, the two samples were somehow linked on a case-by-case basis. In the first instance, a person's score on the second drug awareness test was linked back to his/her score on the first test. The focus was on any difference between the score on the first test and the score on the second. In the marital expectation scenario, a person's responses concerning expectations about marriage were linked to those of another person to whom he/she was socially tied. The focus was on any difference between the scores of the two individuals who made up a couple. In the final example, each worker in one sample was related (by matching on relevant criteria) to another worker in the other sample. To more fully grasp the structure of such research situations, consider Table 8-1.

With the basic structure of the research situation in mind, we can now turn to the logic of the matched sample test application.

The Logic of the Test

The logic in this application begins with the notion of what constitutes a difference. Let's look at the drug awareness scenario again as a starting point. Consider the set of scores presented in Table 8-2, reflecting performance on

Table 8-1 The Focus of Interest With the *t* Test for Related or Matched Samples

These differences are found by subtracting one score from the other score (score in Sample 1 from score in Sample 2, or vice versa).

Sample 1		Sample 2	Differences
Score/Value	- - - ->	Score/Value	d
Score/Value	- - - ->	Score/Value	d
Score/Value	- - - ->	Score/Value	d
Score/Value	- - - ->	Score/Value	d
Score/Value	- - - ->	Score/Value	d
Score/Value	- - - ->	Score/Value	d
Score/Value	- - - ->	Score/Value	d
Score/Value	- - - ->	Score/Value	d
Score/Value	- - - ->	Score/Value	d
Score/Value	- - - ->	Score/Value	d
			$\overline{D}$

This is the mean of the differences ($\overline{D}$).

Table 8-2 Results From Drug Awareness Test

Subject	Test 1 T_1	Test 2 T_2	Difference d
1	50	55	5
2	77	79	2
3	67	82	15
4	94	90	–4
5	64	64	0
6	77	83	6
7	85	80	–5
8	52	55	3
9	81	79	–2
10	91	91	0
11	52	61	9
12	61	77	16
13	83	83	0
14	66	70	4
15	71	75	4

$\Sigma d = 53$

$\overline{D} = 53/15 = 3.53$

the test by a sample of 15 participants in a before/after test situation. First, each of 15 participants was administered a drug awareness test, and their scores were recorded. The participants were then shown a film concerning the dangers of recreational drug use. Following exposure to the film, the 15 participants were given the drug awareness test again, and these scores were recorded, as shown in Table 8-2.

The object of our interest is the differences listed in the right-hand column of the table (column labeled Difference or *d*). This column includes positive differences, negative differences, and zero differences. To find the mean of those differences, we simply add up all the differences and divide by 15: 53/15 = 3.53. If we use the symbol $\overline{D}$ to indicate the mean of the differences (*d*s), then

$$\overline{D} = \frac{\sum d}{n}$$

$$\overline{D} = \frac{53}{15}$$

$$\overline{D} = 3.53$$

Thus, we're in a position to say that the mean of our sample differences is 3.53.

Just as we can calculate a mean of the differences, we can also calculate a standard deviation of the differences as follows:

$$s_d = \sqrt{\frac{\sum (d - \overline{D})^2}{n - 1}}$$

$$s_d = \sqrt{\frac{525.72}{15 - 1}}$$

$$s_d = \sqrt{37.55}$$

$$s_d = 6.13$$

Don't let those calculations throw you. The goal is to calculate a standard deviation of the differences (the *d*s). Just think back to the formula that you used before. Recall that the formula involved an $\overline{X}$ and individual X values. This application is the same, except it involves a $\overline{D}$ and individual d values.

Just in case you're still feeling a little lost on this matter of how to calculate the standard deviation of the differences, let me urge you to do two things. First, review the material on the standard deviation that you encountered in Chapter 2, paying attention to each element in the process and how the $n - 1$ correction factor is used. Secondly, take a look at the steps outlined in the summary, below. The summary doesn't give you every calculation along the way, but there should be enough information to clear up any confusion.

1. Find the mean of the differences (the mean of the *d*s) and designate it as $\overline{D}$.
2. Find the deviation of each d from $\overline{D}(d - \overline{D})$. For example:

   $5 - 3.53 = 1.47$

   $2 - 3.53 = -1.53$

$15 - 3.53 = 11.47$

Continue through all of the *d*s.

3. Square all of the deviations. For example:

   $1.47 \times 1.47 = 2.16$

   $-1.53 \times -1.53 = 2.34$

   $11.47 \times 11.47 = 131.56$

   Continue through all of the squared $\overline{D} - d$ values (all the squared deviations).
4. Find the sum of the squared deviations (it will equal 525.72).
5. Divide the sum of the squared deviation by $n - 1$ (i.e., divide 525.72 by 14) and you have 37.55.
6. Find the square root of 37.55.

The square root of 37.55 = 6.13, so that's the answer. The standard deviation of the differences (the standard deviation of the *d*s) is symbolized by s_d. Therefore, $s_d = 6.13$.

Having determined that the standard deviation (s_d) of the differences (the *d* values) is 6.13, we now have quite a bit of information. We know, for example, that we have 15 observed differences and that the mean of the differences is equal to 3.53 ($\overline{D} = 3.53$). We also know that the standard deviation of the distribution of differences is equal to 6.13. As you might have guessed, we're right back where we've been before. It's probably a good idea to take another dark room moment right now to think about the situation I'm about to describe.

Imagine that you've taken a sample of 15 people, noted their scores at Time 1 and Time 2, and calculated the differences (*d* values) between the individual scores ($T_1 - T_2$ or $T_2 - T_1$; it doesn't matter which approach you take). Now imagine that you've calculated the mean of those differences ($\overline{D}$) and you've recorded that mean. Now imagine that you repeat that process a thousand times over, each time calculating a mean and recording it.

☑ LEARNING CHECK

Question: In this application, how do you calculate the *d* values?

Answer: Find the difference between the two scores or values associated with a given research subject or case (e.g., difference between score at Time 1 and Time 2).

By now you should know what the next step is in your mental picture—you're plotting all of those means to create the sampling distribution of mean differences. The question, of course, boils down to the same question you've asked yourself before: Where does my observed mean difference fall along a sampling distribution of all possible mean differences? With that thought in mind, we can now turn to the null hypothesis for this application.

The Null Hypothesis

The null hypothesis in this application is simply a statement that we expect a mean difference of 0. You'll recall from the previous chapter that the null is frequently a statement of no difference, and that notion applies squarely in the present instance. Regardless of what may be operating in the back of our minds—even though we may, in truth, expect a change in scores between the first and the second test—we will test the null. And the null is a statement that we expect the mean difference to be 0. In other words, the null is a statement that we expect there to be no change. Symbolically, we can state the null hypothesis as follows:

$$H_0\colon \mu_{\overline{D}} = 0$$

Mean of differences obtained from subtracting score at one time (T_1) from score at another time (T_2)

LEARNING CHECK

Question: What does the statistical symbol H_0 represent?
Answer: H_0 is the symbol that represents the null hypothesis.

Combining the Logic and the Null

It's a good idea at this point to think about the big picture, in the sense that we're not really interested in what's going on with the samples (the sample at T_1 and the sample at T_2). Rather, we're interested in what's really going on with the populations (the population at T_1 and the population at T_2). We're wondering what we'd discover if we tested the entire population of students at T_1 and retested that entire population at T_2. With the null, we're making the statement that we expect the mean of any differences in the populations to be 0.

Our goal, then, is to compare our observed mean difference to an assumed mean difference of 0. Since we're assuming that the mean difference is 0, we can assume that the mean of the sampling distribution of mean differences is also equal to 0. It's the Central Limit Theorem that allows us to make that assumption. As before, we'll eventually compare our observed mean difference (3.53) to the assumed mean difference of 0, and express the magnitude of any difference between the two values in standard error units. Remember: The estimated standard error in this case will be the estimated standard deviation of a sampling distribution of mean differences (the standard deviation of the distribution you mentally constructed if you took the time for the last dark room moment I suggested).

☑ LEARNING CHECK

Question:	What is the definition of the estimate of the standard error in the case of a t test for related sample means?
Answer:	It is an estimate of the standard deviation of the sampling distribution of mean differences.

The Estimate of the Standard Error of the Mean Difference

The estimate of the **standard error of the mean difference** is symbolized by $s_{\overline{D}}$ and is calculated as follows:

$$s_{\overline{D}} = \frac{s_d}{\sqrt{n}}$$

Where s_d represents the standard deviation of the distribution of differences

You should take note of how similar the formula is to the formula for the estimate of the standard error that you encountered in Chapter 7. Assuming you know of the standard deviation of the distribution of the *d* values (it was provided to you or you calculated it), all you have to do is divide that value by the square root of the sample size. The result is the estimate of the standard error of the mean difference. In the example we're considering here, it would be as follows:

$$s_{\overline{D}} = \frac{6.13}{\sqrt{15}}$$

$$s_{\overline{D}} = \frac{6.13}{3.87}$$

$$s_{\overline{D}} = 1.58$$

☑ LEARNING CHECK

Question:	How is the estimate of the standard error calculated in the case of a t test for related sample means?
Answer:	The standard deviation of the distribution of differences (in other words, the distribution of the *d* values) is divided by the square root of *n*.

Applying the Test

The application is now fairly straightforward. All we have to do is compare our observed mean difference to the assumed mean difference of 0, and divide the result by our estimate of the standard error. This procedure produces a t ratio, similar to those you've encountered before. A complete formula for the calculation of the t ratio would be as follows:

$$t = \frac{\overline{D} - \mu_{\overline{D}}}{s_{\overline{D}}}$$

Since the second part of the formula ($\mu_{\overline{D}}$) is assumed to be equal to 0, we can drop that element from the formula, leaving ourselves with only the following to consider:

$$t = \frac{\overline{D}}{s_{\overline{D}}}$$

$$t = \frac{3.53}{1.58}$$

$$t = 2.23$$

Working through the calculations, we arrive at a calculated test statistic (t ratio) of 2.23. Now all that remains is to evaluate the calculated test statistic in light of the critical value—the value that our calculated test statistic must meet or exceed if we are to reject the null hypothesis.

☑ LEARNING CHECK

Question: What does the value of t represent in the t test for related sample means?

Answer: The t value is the calculated test statistic. It is a ratio that expresses how far the observed mean difference departs from the assumed mean difference of 0 in standard error units.

Interpreting the Results

Assuming we had set our level of significance at .05 in advance of our test, our next step is to identify the critical value and move toward a conclusion and interpretation. In a t test for related samples, the degrees of freedom will be our sample size (n) minus 1—in this case, 15 – 1, or 14. Working with 14 degrees of freedom, at the .05 level of significance, we find the critical value of t in Appendix B to be 2.15. Recall that our calculated test statistic (the t ratio) is 2.23.

Since the calculated t value exceeds the critical value, we're in a position to reject the null hypothesis. In other words, we reject the notion that there is no difference between the two populations. We reject the idea that the two populations do not differ from one another (T_1 and T_2). That, of course, is another way of saying that it appears that the two populations do differ. Exposure to the drug awareness film appears to have some effect on test scores.

In rejecting the null hypothesis at the .05 level of significance, we're acknowledging that there's always a chance that we've made a mistake. Yes, we've rejected the null—the expectation that there is no difference. It's possible, though, that there really is no difference and our observed difference was, just by chance, a very extreme case along the sampling distribution of all possible mean differences. If that's what was going on, we've made a Type I error—we've rejected the null when, in fact, it's true. As before, however, we never know whether or not we've committed a Type I error. All we know is the probability of having committed a Type I error, and that probability is our level of significance. Once again, that's why I think it's advisable at this point in your statistical education to offer an interpretation along the following lines: We reject the null with the knowledge that there's a 5% chance that we've committed a Type I error.

Later on, when you're totally comfortable with the logic of hypothesis testing, you can feel free to drop the reference to Type I errors when announcing your conclusion. Until you're comfortable, though, let me suggest you stick with a format that makes reference to the probability of making a Type I error in your conclusion. It's a good way to continually hammer home the logic of hypothesis testing.

Some Additional Examples

Just to make certain that you're on the right track with the t test for related or matched samples, give some time and thought to each of the following questions, paying particular attention to each element that goes into the calculation of the final test statistic (the t ratio).

1. Assume two matched samples, each involving 10 cases ($n = 10$) and the following information:

   Mean of the differences ($\overline{D}$) = 11.65

   Estimated standard error of the mean difference ($s_{\overline{D}}$) = 3.39

   Calculate t. Assuming a .05 level of significance ($\alpha = .05$), do you reject or fail to reject the null hypothesis?

2. Assume two related samples, each involving 10 cases ($n = 10$) and the following information:

   Mean of the differences ($\overline{D}$) = 2.70

   Estimated standard error of the mean difference ($s_{\overline{D}}$) = 0.97

   Calculate t. Assuming a .01 level of significance ($\alpha = .01$), do you reject or fail to reject the null hypothesis?

3. Assume two matched samples, each involving 14 subjects ($n = 14$) and the following information:

   Mean of the differences ($\overline{D}$) = 3.42

   Standard deviation of the differences (s_d) = 3.46

Recall the formula for determining the estimated standard error of the mean using n and s_d. Calculate t. Assuming a .05 level of significance ($\alpha = .05$), do you reject or fail to reject the null hypothesis?

4. Assume two matched samples each involving 19 subjects ($n = 19$) and the following information:

   Mean of the differences ($\overline{D}$) = 7.11

   Standard deviation of the differences (s_d) = 13.84

   Calculate t. Assuming a .05 level of significance ($\alpha = .05$), do you reject or fail to reject the null hypothesis?

5. Assume two related samples involving the following information:

   $n = 22$

   $\overline{D} = 3.52$

   $s_d = 5.26$

   $\alpha = .05$

   Discuss your results in terms of the null hypothesis.

Answers:

1. $t = 3.44$
   Reject the null hypothesis at the .05 level of significance.
2. $t = 2.78$
   Fail to reject the null hypothesis at the .01 level of significance.
3. $t = 3.68$
   Reject the null hypothesis at the .05 level of significance.
4. $t = 2.24$
   Reject the null hypothesis at the .05 level of significance.
5. $t = 3.14$
   Reject the null hypothesis at the .05 level of significance.

So much for the t test for related samples. Now we turn our attention to another application involving the t ratio with a focus on differences. In the next application, however, we encounter a very different meaning of *difference*.

Independent Samples

You've just covered the idea of sample differences (and, therefore, population differences) in one context. Now you'll encounter a different context—one with a new notion as to what constitutes a difference. While the last application is still fresh in your mind, let's turn to the next application.

Suppose that I asked you how you'd go about determining whether or not there's a difference between the drinking habits of fraternity members and nonfraternity members. My guess is that you'd probably give me a fairly good explanation. For example, you'd very likely tell me that you'd start by getting a random sample of fraternity members and a random sample of nonfraternity members. Then you might tell me that you'd get information on their drinking habits (maybe by asking them how many drinks they typically consume in a week). Maybe you'd go so far as to tell me you'd calculate the mean ($\overline{X}$) number of drinks for each sample, just so you'd have a starting point to compare the two groups. Assuming you outlined the problem that way, I'd say "Congratulations! You're on the right track." You've just outlined the basic elements in an independent sample research design.

The key notion in all of that—**independent samples**—is the idea that the sample of fraternity members is selected independently of the sample of nonmembers. In other words, the selection of cases for one sample in no way affects the selection of cases in the other sample. Even if we selected a single sample of students at random and then divided that sample into two groups (fraternity members and nonmembers), we would be dealing with the same principle of independence.

☑ LEARNING CHECK

Question: What does the phrase *independent samples* mean?

Answer: Independent samples are samples selected in such a way that the selection of cases or subjects included in one sample has no connection to or influence on the selection of cases or subjects in the other sample.

With all of that in mind, let's assume that we undertake the research you have outlined, and let's say we obtain the data regarding number of drinks per week found in Table 8-3.

The Logic of the Test

The present application allows us to compare the means of two samples with an eye toward whether or not any difference is a reflection of a true difference between the populations. So much for the central goal of the test. Now let's turn to the elements that make up the structure of the test. As a prelude, take a close look at Table 8-4.

Table 8-3 Data Regarding Drinks per Week

Fraternity	Nonfraternity
6	0
3	5
2	3
4	4
5	3
6	6
7	3
5	6
4	5
5	4
4	4
8	2
6	$n = 12$ $\overline{X} = 3.75$ $s = 1.71$
7	
$n = 14$ $\overline{X} = 5.14$ $s = 1.66$	

Study based on a sample of 14 fraternity members and a sample of 12 nonmembers

Mean of nonmembers = 3.75.

Mean of fraternity members = 5.14.

Standard deviation of nonmembers = 1.71.

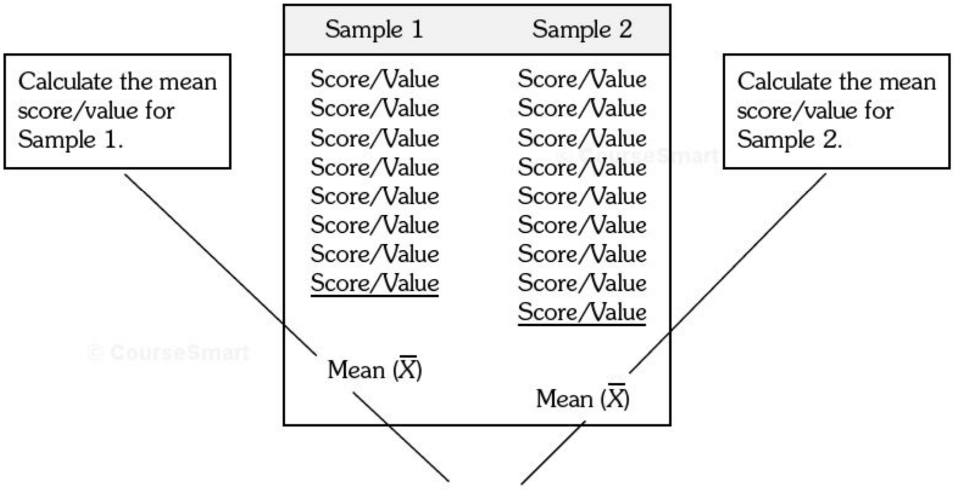
Standard deviation of fraternity members = 1.66.

LEARNING CHECK

Question:	In the *t* test for difference of means for independent samples, which difference is the object of interest?
Answer:	The test focuses on the difference between the mean of one sample and the mean of another sample, with an eye toward the extent to which any observed difference represents a true difference between population means.

You're already familiar with the notion of a sampling distribution of sample means when dealing with a single sample. You know, for example, that you could take sample after sample after sample, and each sample is apt to yield a slightly different sample mean. As it turns out, the same logic applies in a situation involving the difference between two samples. Returning to our example

Table 8-4 The Focus of Interest With the *t* Test for Independent Samples



Calculate the mean score/value for Sample 1.

Sample 1	Sample 2
Score/Value	Score/Value
Score/Value	Score/Value
Score/Value	Score/Value
Score/Value	Score/Value
Score/Value	Score/Value
Score/Value	Score/Value
Score/Value	Score/Value
Score/Value	Score/Value
	Score/Value
Mean ($\overline{X}$)	Mean ($\overline{X}$)

Calculate the mean score/value for Sample 2.

Find the difference between the two means.

The null is a statement that we expect the difference to be 0.

involving the drinking habits of fraternity members and nonmembers, consider the following:

- You could take a sample of fraternity members and a sample of nonmembers. You could calculate the mean of each sample, calculate the difference between the two means, and record that difference.
- You could then take another two samples, calculate the mean of each sample, and record the difference between those means.
- You could repeat the process over and over again—each time calculating the mean of each sample, noting any difference between the two means, and recording that difference.
- You could eventually produce a plot or distribution of all of those differences. That plot or distribution would be known as a *sampling distribution of the differences of means.*

If necessary, run that material through your mind again as a dark room moment. Take the time to mentally focus on the difference between the means that you get each time you select two samples. Focus on the fact that you're interested in the difference between the means of pairs of samples.

As you might expect, the question will ultimately come down to a matter of how extreme our observed difference between two means is in comparison to all possible differences that we might observe. It's similar to the question you've faced before—whether or not an observation is an extreme value along a sampling distribution (in this case, of the difference of means).

The Null Hypothesis

Assuming you took a careful look at Table 8-4, you noticed a central element in the logical underpinnings of the *t* test for independent samples—namely, a reference to the null hypothesis that there is no difference between the means. More specifically, the null hypothesis is a statement that there is no difference between the means of the two populations. As before, we may be looking at samples, but our interest is in what's going on with the populations from which the samples were taken. In symbolic terms, we can state the null hypothesis in two different ways:

$$H_0\colon \underset{\substack{\swarrow\!\!\nearrow \\ \text{Mean of fraternity} \\ \text{members}}}{\mu_1} = \underset{\substack{\nwarrow \\ \text{Mean of} \\ \text{nonmembers}}}{\mu_2} \qquad \text{or} \qquad H_0\colon \underset{\substack{\nearrow \\ \text{Mean of fraternity} \\ \text{members}}}{\mu_1} - \underset{\substack{\nwarrow \\ \text{Mean of} \\ \text{nonmembers}}}{\mu_2} = 0$$

Combining the Logic and the Null

Now we're back in familiar territory. If the null hypothesis is true, we'd expect the mean of the sampling distribution of the difference between means to equal 0. Moreover, we'd expect any observed difference between two sample means to be fairly close to 0. It all boils down to whether or not our observed difference is extreme.

As before, we'll assume we're working at the .05 level of significance. The question now turns on what constitutes an extreme difference and the magnitude of any observed difference, expressed in standard error units. Remember: It doesn't matter whether we're considering differences in salaries (expressed in dollars), weights (expressed in pounds or ounces), test scores (from 0 to 100), drinks per week, or anything else. It's always a question of how far our observed difference departs from the assumed difference of 0 in standard error units.

The Estimate of the Standard Error of the Difference of Means

In the previous application (the one involving related samples and the drug awareness test), the calculation of the estimate of the standard error was straightforward. We used the standard deviation of the differences as the basis for our estimate of the standard error of the mean difference. We had a distribution of differences (obtained by calculating the difference in test scores for each person), and we calculated the standard deviation of that distribution of differences. It was then a short step to divide the standard deviation by the square root of our sample size (the number of pairs of scores). The result was the estimate of the standard error of the mean difference.

The calculation of the standard error in this application (independent samples) isn't quite so straightforward. In the present case, we're no longer dealing with just one distribution (as we did in the case of the lone distribution of differences). Instead, we're dealing with two distributions. We have a distribution of scores for fraternity members and a distribution of scores for nonmembers. Thus, the estimate of the standard error becomes a bit more complex, and at times, confusing. Here's why.

Complexity and Some Possible Confusion. First, we have to consider that we're dealing with two samples. Besides that, we have to consider that we're often dealing with different sample sizes. The present situation, for example, involves 14 fraternity members but only 12 nonmembers. In other research situations, you might find yourself working with 25 cases in one sample and 45 in another.

It shouldn't take you long to realize that any measure of the variation that's present in a sample is, in part, a function of sample size. For example, you're not apt to pick up much of the variation that exists in a population if you're working with a sample of only 2 cases. A sample of 100 cases, though, will probably provide a fairly good representation of the variation. Given that, it stands to reason that our formula for the estimate of the standard error of the difference of means will be sensitive to the number of cases in each sample.

Another source of complexity (and perhaps confusion) stems from the fact that the formula begins with a consideration of the *variance* for each sample. In the previous application (involving related samples), as well as various applications in the last chapter, we used the *standard deviation* to develop our estimate of the standard error. (We simply took the standard deviation and divided it by the square root of the sample size.) In the present application, though, we'll first look at sample variances as we set out to calculate the estimate of the standard error.

The reason why we'll start with the variances traces back to the previous point—the need to take into account different sample sizes. As you probably recall, the standard deviation is the square root of the variance. Conversely, the variance is merely the standard deviation squared (s^2). Given that relationship, you might be inclined to approach the problem of different sample sizes by weighting the standard deviations of the samples—that is, multiplying the standard deviations by different values to reflect the different sample sizes.

The truth of the matter, though, is that the variance is always calculated first. We first calculate a variance, and then we take the square root of it to obtain the standard deviation. It's true that you often see statistical problems or data summaries presented with only the standard deviation given (and not the variance), but you can rest assured of one thing: The variance was calculated first.

Besides that, you should give some thought to the effect that weighting has on a variance, as opposed to a standard deviation. Let's say, for example, that we want to weight some values by a factor of 10. A *variance* of 9, for example, is equal to a *standard deviation* of 3 (the standard deviation being the square root of the variance). The variance times 10 equals 90, and the square root of 90 is 9.487. But 9.487 is hardly the same value as the standard deviation (3) times 10.

With those considerations in mind, you're in a better position to understand why the formula for the estimate of the standard error of the difference of means begins with weighted sample variances. I can assure you that the formula eventually readjusts, so to speak, into an expression of standard deviations, but it begins with a consideration of weighted variances.

All of those issues aside, the real point is what the estimate of the standard error of the difference of means allows us to do. As the standard error has done before in the other applications we've tackled, it allows us to eventually express the magnitude of the observed difference of means in a standardized way. Now let's see how it is calculated.

The Formula. The formula we'll use to estimate the **standard error of the difference of means** is as follows:

$$s_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \bullet \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

Note: As noted previously, the formula presented for the estimate of the standard error is based on the assumption that the variance and/or standard deviation of each sample was originally calculated by using $n - 1$ in the denominator.

There's no reason to let this formula overwhelm you. Remember: A fair amount of the complexity traces back to the fact that we're initially having to deal with variances (the s^2 values), and we have two samples under consideration. Just to recap the information we have so far concerning the sample results, consider the following summary (see Table 8-3):

Fraternity members	$\overline{X}_1 = 5.14$	$n_1 = 14$	$s_1 = 1.66$
Nonmembers	$\overline{X}_2 = 3.75$	$n_2 = 12$	$s_2 = 1.71$

You'll note that the value of the standard deviation (s) is given. Recalling what we recently covered, you know that all you have to do to obtain the variance is to square the standard deviation. Thus, we determine the variance (s^2) for each of the two samples as follows:

Fraternity members	$s_1^2 = 1.66^2 = 2.76$
Nonmembers	$s_2^2 = 1.71^2 = 2.92$

Armed with the value of the variance and the number of cases for each sample, we're now in a position to develop an estimate of the standard error of the difference of means. Recalling the formula previously presented, we can calculate the estimate of the standard error as follows:

$$s_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \bullet \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

$$s_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{(14 - 1)2.76 + (12 - 1)2.29}{14 + 12 - 2} \bullet \left[\frac{1}{14} + \frac{1}{12}\right]}$$

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{(13)2.76 + (11)2.92}{24} \bullet [0.07 + 0.08]}$$

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{35.88 + 32.12}{24} \bullet [0.15]}$$

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{68}{24} \bullet [0.15]}$$

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{2.83[0.15]}$$

$$s_{\overline{x}_1-\overline{x}_2} = \sqrt{0.42}$$

$$s_{\overline{x}_1-\overline{x}_2} = 0.65.$$

Applying the Test

Now we're in a position to put everything together. To do that, we'll return to the information we now have:

Fraternity members	$\overline{X}_1 = 5.14$	$n_1 = 14$	$s_1^2 = 2.76$
Nonmembers	$\overline{X}_2 = 3.75$	$n_2 = 12$	$s_2^2 = 2.92$

The estimate of the standard error of the difference of means $(s_{\overline{x}_1-\overline{x}_2}) = .65$.

For the sake of this example, we'll assume we're working at the .05 level of significance.

Our goal is to compare any observed difference between the two means (fraternity members and nonmembers) to an assumed difference of 0, and then to convert the magnitude of any observed difference into standard error units. Accordingly, here's the way the formula looks. As before, we're converting the comparison to a *t* ratio—hence, the symbol *t* at the outset of the formula.

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{s_{\overline{x}_1-\overline{x}_2}}$$

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{s_{\overline{x}_1-\overline{x}_2}}$$

Working our way through the formula, we obtain the following test statistic (our calculated value of *t*):

$$t = \frac{(5.14 - 3.75)}{0.65}$$

$$t = \frac{1.39}{0.65}$$

$$t = 2.14$$

Interpreting the Results

Now we have a calculated test statistic, t, with a value of 2.14. The question, of course, is whether it represents a significant result. In other words, does the calculated test statistic equal or exceed the critical value? To determine the answer to that question, we return to Appendix B. As before, we focus on the column for the .05 level of significance.

In the case of the t test for the difference of means, the number of degrees of freedom has to reflect the number of cases in each sample. Up to this point we've calculated the number of degrees of freedom as $n - 1$, but now we're dealing with two independent samples. Now we calculate the number of degrees of freedom for each sample, and add the two together. In other words, the number of degrees of freedom in the difference of means test is as follows:

$$(n_1 - 1) + (n_2 - 1) \qquad \text{or} \qquad (n_1 + n_2) - 2$$

In the present example, we're working with one sample of 14 cases and another sample of 12 cases. Therefore, we're working with (14 – 1) + (12 – 1), or 24 degrees of freedom.

Armed with the knowledge that we're working with 24 degrees of freedom at the .05 level of significance, all that remains is to identify the critical value—found by locating the intersection of the appropriate row (degrees of freedom) and column (.05 level of significance) in Appendix B. We note that the critical value is 2.06.

We find that our calculated test statistic ($t = 2.14$) exceeds the critical value of 2.06. Therefore, we can reject the null hypothesis, with the knowledge that there is a 5% chance that we've made a Type I error. By rejecting the null hypothesis, we're rejecting the notion that the two populations (the population of fraternity members and nonmembers) are equal in terms of levels of drinking. In other words, we've found support for the assertion that the means of the two populations are, in fact, different.

Some Additional Examples

Just to increase your familiarity with the process involved in the hypothesis test for the independent samples difference of means test, consider the following problems. Additional problems are included at the end of the chapter.

1. Assume two independent samples and the following information:

   Mean of Group 1 ($\overline{X}_1$) = 53.92　　Mean of Group 2 ($\overline{X}_2$) = 50.0
   Sample size of Group 1 (n_1) = 11　　Sample size of Group 2 (n_2) = 17

   Estimated standard error of the difference between means ($s_{\bar{x}_1 - \bar{x}_2}$) = 2.80

   Calculate t. Assuming a .05 level of significance ($\alpha = .05$), do you reject or fail to reject the null hypothesis?

2. Assume two unrelated samples and the following information:

   Mean of Group 1 ($\overline{X}_1$) = 7.39 Mean of Group 2 ($\overline{X}_2$) = 4.50

   Sample size of Group 1 (n_1) = 18 Sample size of Group 2 (n_2) = 8

   Estimated standard error of the difference between means ($s_{\overline{X}_1-\overline{X}_2}$) = 0.90

   Calculate t. Assuming a .01 level of significance ($\alpha = .01$), do you reject or fail to reject the null hypothesis?

3. Assume two independent samples and the following information:

   $\overline{X}_1 = 24.53$ $n_1 = 24$

   $\overline{X}_2 = 26.28$ $n_2 = 18$

   $s_{\overline{X}_1-\overline{X}_2} = 0.67$ $\alpha = .01$

   Calculate t. Do you reject or fail to reject the null hypothesis?

4. Assume two unrelated samples and the following information:

   Mean of Group 1 ($\overline{X}_1$) = 6.93 Mean of Group 2 ($\overline{X}_2$) = 4.38

   Variance of Group 1 (s_1^2) = 2.86 Variance of Group 2 (s_2^2) = 5.06

   Sample size of Group 1 (n_1) = 14 Sample size of Group 2 (n_2) = 16

   Recall the formula for calculating the estimated standard error of the difference between means using sample sizes and variances. Calculate t. Assuming a .05 level of significance ($\alpha = .05$), do you reject or fail to reject the null hypothesis?

5. Assume two independent samples and the following information:

   $\overline{X}_1 = 10.81$ $s_1^2 = 1.85$ $n_1 = 15$

   $\overline{X}_2 = 13.14$ $s_2^2 = 3.84$ $n_2 = 11$

   $\alpha = .05$

   Calculate t. Do you reject or fail to reject the null hypothesis?

Answers:

1. $t = 1.40$
   Fail to reject the null hypothesis at the .05 level of significance.

2. $t = 3.21$
   Reject the null hypothesis at the .01 level of significance.
3. $t = -2.61$
   Fail to reject the null hypothesis at the .01 level of significance.
4. $t = 3.49$
   Reject the null hypothesis at the .05 level of significance.
5. $t = -3.53$
   Reject the null hypothesis at the .05 level of significance.

Chapter Summary

This chapter introduced you to one of the most commonly encountered research situations—those based on two samples. As you worked your way through the material, you discovered several important ideas and considerations that are brought to bear in two sample research situations.

First, you explored the variations in research designs as you considered related sample designs and then independent sample designs. In the process, you learned that each looks at the concept of a difference in its own way. Accordingly, you learned that different research designs call for different approaches to the calculation of the *t* ratio.

Despite those fundamental differences, you should have been aware of the central theme that is present in both research situations—namely, the basic logic of hypothesis-testing situations. By now, the rather uniform approach should be solidified in your mind. For example, you probably sense by now that the applications always begin with the formulation of a null hypothesis and selection of a level of significance. From that point, you move to the calculation of a test statistic and comparison of that test statistic to a critical value. Based on your evaluation of the test statistic in light of the critical value, you arrive at a conclusion (you either reject or fail to reject the null hypothesis).

The importance of this hypothesis-testing process and logic can't be overstated. You'll encounter the same sort of logic in most statistical test situations. The research situations will vary. The particular test that is called for will vary. But the underlying logical structure remains similar across the applications.

Some Other Things You Should Know

The difference of means test is such a widely used statistical procedure that you deserve to know a few more things about it. Toward that end, we'll consider some of the assumptions that underlie application of the difference of means test. We'll also take an abbreviated look at how to approach the test when large samples are available and what to do to test hypotheses about differences between proportions.

The *t* tests both for independent and for related or matched samples rest on the assumption that the populations in question are normally distributed. This particular assumption is of more importance with small samples and can be relaxed somewhat in situations involving large samples. There's also an assumption that the populations involved in the independent samples application have equal variances. For an excellent discussion of both of these assumptions, as well as how the assumption of equality of variance can be tested, see Gravetter and Wallnau (2000).

For the independent samples application with large samples, the appropriate test statistic is Z. Here's what that means in practical terms:

- Instead of calculating t, you would calculate Z if the combined number of cases ($n_1 + n_2$) is greater than 100.
- The formula remains the same, except that Z replaces t at the beginning of the formula:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2)}{s_{\overline{X}_1 - \overline{X}_2}}$$

- If you are calculating Z, you use the Table of Areas Under the Normal Curve (the sampling distribution of Z) to obtain the appropriate critical value. Accordingly, degrees of freedom are of no consideration.

Finally, let me call your attention to yet another two sample test situation—the difference of proportions test. Situations calling for such a test would include questions about whether the proportion of people favoring a certain candidate has changed significantly over time or whether the proportion of students favoring evening courses is significantly different among students in state universities and private universities. This test shares the same logic as the difference of means tests you've just encountered, but the focus is on proportions as opposed to means. For a discussion of the difference of proportions test, see Healy (2002).

Key Terms

independent samples
matched or related samples
standard error of the difference of means
standard error of the mean difference

Chapter Problems

Fill in the blanks, calculate the requested values, or otherwise supply the correct answer.

General Thought Questions

1. In a matched samples design (the test involving the mean difference) the number of cases in each sample must be equal. True or false?
2. In the test for the difference of means, independent sample design, the number of cases in each sample must be equal. True or false?
3. In the matched samples design (the design based on the mean difference), the sampling distribution at issue is the sampling distribution of ____________________.

4. In the independent samples design (the design based on the difference of means), the sampling distribution at issue is the sampling distribution of ________________.

Application Questions/Problems: Matched/Related Samples Design

1. Assume you are working with the results of a research situation based on a matched sample design involving 15 participants (the same participants in a before/after test situation). The mean difference ($\overline{D}$) = 14.66, with an estimate of the standard error of the mean difference ($s_{\overline{D}}$) = 5.21. Assume that you're working at the .05 level of significance.
- **a.** Formulate an appropriate null hypothesis.
- **b.** Calculate t.
- **c.** Identify the critical value.
- **d.** State your conclusion.

2. Assume you are working with the results of a research situation based on a matched sample design involving 25 participants (the same participants in a before/after test situation). The mean difference ($\overline{D}$) = 9.72, with an estimate of the standard error of the mean difference ($s_{\overline{D}}$) = 6.33. Assume that you're working at the .05 level of significance.
- **a.** Formulate an appropriate null hypothesis.
- **b.** Calculate t.
- **c.** Identify the critical value.
- **d.** State your conclusion.

3. Thirty program participants have been given a test designed to measure their reading comprehension skill levels. Following a two-week course, designed to improve reading comprehension skills, the same participants are re-tested. The mean difference ($\overline{D}$) = 5.43, with an estimate of the standard error of the mean difference ($s_{\overline{D}}$) = 2.11. Assume that you're working at the .05 level of significance.
- **a.** Formulate an appropriate null hypothesis.
- **b.** Calculate t.
- **c.** Identify the critical value.
- **d.** State your conclusion.

4. Professor Johnson administers a 50 point test to the students in her class (n = 29) at the beginning of the semester (T_1) to measure their understanding of basic sociological concepts. She administers the same test to the students at the conclusion of the semester (T_2), and records each student's T_1 and T_2 scores. She obtains the following:

> The mean of the distribution of differences ($\overline{D}$) between the students' T_1 and T_2 scores is equal to 6.57, with an estimate of the standard error of the mean difference ($s_{\overline{D}}$) = 2.88.

Assume that you're working at the .05 level of significance.
- **a.** Formulate an appropriate null hypothesis.
- **b.** Calculate t.

c. Identify the critical value.
d. State your conclusion.

5. A geographer selects 30 counties from the north and 30 counties from the south, matching the two samples on the basis of urban/rural status (whether the county is inside or outside of a metropolitan area), dominant economic activity (manufacturing, retail, service, etc.), voter pattern in the last presidential election (whether the majority voted for the Republican or Democratic candidate). Focusing on the percentage of registered voters who actually voted in the last presidential election, the researcher obtains the following results:

> The mean of the distribution of differences ($\overline{D}$) between northern counties and southern counties for voter turnout is equal to 4.00%, with an estimate of the standard error of the mean difference ($s_{\overline{D}}$) = 2.35%.

Assume that you're working at the .05 level of significance.
a. Formulate an appropriate null hypothesis.
b. Calculate t.
c. Identify the critical value.
d. State your conclusion.

Application Questions/Problems: Independent Sample Design

1. Calculate t for the following research situation involving two independent samples, Sample A and Sample B. (Assume that you are subtracting the Mean of Sample B from the Mean of Sample A).

Mean of Sample A = 12.65;	$n = 15$
Mean of Sample B = 10.42;	$n = 18$

Estimate of the standard error of the difference of means ($s_{\bar{x}_1-\bar{x}_2}$) = .75. Assume that you're working at the .05 level of significance.
a. Formulate an appropriate null hypothesis.
b. Calculate t statistic.
c. Identify the critical value.
d. State your conclusion.

2. Consider a research situation involving two independent samples, Sample A and Sample B. (Assume that you are subtracting the Mean of Sample B from the Mean of Sample A).

Mean of Sample A = 30.45	$n = 25$
Mean of Sample B = 26.54	$n = 27$

Estimate of the standard error of the difference of means ($s_{\bar{x}_1-\bar{x}_2}$) = 2.15. Assume that you're working at the .05 level of significance.
a. Formulate an appropriate null hypothesis.
b. Calculate t statistic.
c. Identify the critical value.
d. State your conclusion.

3. Consider a research situation involving two independent samples, Sample A and Sample B. (Assume that you are subtracting the Mean of Sample B from the Mean of Sample A).

Mean of Sample A = 4.12 $n = 15$

Mean of Sample B = 6.23 $n = 13$

Estimate of the standard error of the difference of means ($s_{\bar{x}_1 - \bar{x}_2}$) = 1.44. Assume that you're working at the .05 level of significance.

a. Formulate an appropriate null hypothesis.
b. Calculate t statistic.
c. Identify the critical value.
d. State your conclusion.

4. A criminologist is interested in possible disparities between sentences given to males and females convicted in murder-for-hire cases. Selecting 14 cases involving men convicted of trying to solicit someone to kill their wives and 16 cases involving women convicted of trying to solicit someone to kill their husbands, the criminologist finds the following:

Mean length of sentence for males = 7.34 years with a standard deviation of 2.51 years

Mean length of sentence for females = 9.19 years with a standard deviation of 3.78 years

Assume that you're working at the .05 level of significance.

a. Formulate an appropriate null hypothesis.
b. Calculate t statistic.
c. Identify the critical value.
d. State your conclusion.

5. A political scientist is interested in the question of whether or not there is a difference between Republicans and Democrats when it comes to their involvement in voluntary associations. Using a 25 point scale to measure involvement in voluntary associations, and collecting information from a random sample of 22 Republicans and 17 Democrats, he/she discovers the following:

Republicans: Mean of 12.56 with a standard deviation of 3.77

Democrats: Mean of 16.43 with a standard deviation of 4.21

Assume that you're working at the .05 level of significance.

a. Formulate an appropriate null hypothesis.
b. Calculate t statistic.
c. Identify the critical value.
d. State your conclusion.