9

Beyond the Null Hypothesis

Our discussion of the null hypothesis began two chapters ago with the notion that the null hypothesis is a statement of equality (no difference) or chance. We traveled the road of hypothesis-testing logic, dealing with calculated test statistics, levels of significance, critical values, regions of rejection, and Type I and II errors. As I mentioned before, though, there is still more logic to consider, so that's where we turn now.

In doing so, however, we'll abandon the format we've relied on in the last two chapters. Instead of busying ourselves with calculations, computations, and such, we'll enter a more conceptual world. First, we'll examine the role of alternative or research hypotheses in the field of statistical analysis. That, in turn, will lead us to a discussion of the difference between one-tailed and two-tailed test situations. Finally, we'll deal with the matters of power and effect.

© CourseSmart

Before We Begin

This chapter rests on the assumption that you're familiar with the process of hypothesis testing; so much so that you have an almost intuitive understanding of the process—stating the null, selecting a level of significance, identifying the critical value, calculating the test statistic, and making a decision. There's also an assumption made that you fully understand the concept of making a Type I error.

With that as a background, we now go a bit further. For example, we'll round out your understanding of null hypotheses by giving you a look at alternative hypotheses. Also, you'll be given a chance to further your understanding of Type I errors through the discussion of Type II errors. And so it goes with this chapter. The material is highly conceptual in nature, and it deals with some of finer points of interpretation of findings.

Research or Alternative Hypotheses

To understand the concept of a research or alternative hypothesis, think about our earlier example involving exposure to a film on the dangers of recreational drug use. In that example, we tested the null hypothesis that drug awareness test scores would not change following exposure to a film on the dangers of recreational drug use. Even though we tested the null hypothesis (the expectation of no difference), it's hard to imagine that we wouldn't have expected some change or difference in the scores after exposure to the film. After all, that's what often drives our research in the first place—the expectation that we'll find a difference.

If we shift our thinking away from the null hypothesis and toward an expected difference of some sort, we can consider several alternatives. As it turns out, those alternatives are referred to as **alternative hypotheses** or **research hypotheses.** Recalling our earlier example, think about the various alternative or research hypotheses that we might have advanced:

H_1: Drug awareness test scores after exposure to the film will be higher than prior to exposure to the film.
(In other words, we expect the scores on the drug awareness test to *increase* following exposure to the film.)

H_2: Drug awareness test scores after exposure to the film will be lower than prior to exposure to the film.
(In other words, we expect the scores on the drug awareness test to *decrease* following exposure to the film.)

H_3: Drug awareness test scores before and after exposure to the film will be different or will change in some way.
(In other words, we expect the scores to *change*, but we don't know if the scores will increase or decrease.)

© CourseSmart

Each of these statements (H_1, H_2, and H_3) represents an alternative or research hypothesis. Each stands in opposition to the null hypothesis (H_0). Each statement asserts something other than the null.

LEARNING CHECK

Question: What is an alternative or research hypothesis?
Answer: An alternative or research hypothesis is a hypothesis that stands in opposition to the null hypothesis.

Let's start with a close look at H_1 and H_2. Those are **directional hypotheses,** in the sense that they specify the nature or direction of the change or difference that we expect. H_1 is a statement that we expect the test scores to increase; H_2 is a statement that we expect the scores to decrease. H_3, on the other hand, is a **non-directional hypothesis.** It doesn't specify the direction of the expected difference; H_3 merely states that we expect to find a difference. It's still an alternative or research hypothesis, in the sense that it stands in opposition to the null, but the direction of the expected difference isn't specified.

LEARNING CHECK

Question: What is a directional hypothesis, and what is a non-directional hypothesis?
Answer: A directional hypothesis specifies the nature or direction of a hypothesized difference. It asserts that there will be a difference or a change in a particular direction (increase or decrease). A non-directional hypothesis does not specify the nature or direction of an expected difference. It simply asserts that a difference will be present.

© CourseSmart

Up to this point we've been considering drug awareness scores, but the notion of a research or alternative hypothesis applies in a host of research situations. In one research situation, for example, we might expect alcohol consumption to be higher in one population than another. In another research situation, we might expect one group to be more productive than another. Still another research effort might find us back in a classic before/after research design, once again expecting scores to change in a particular direction. Regardless of the specifics of the research situation, it's common to approach the task with some set of expectations in place—some expectation other than the null. And that brings us back to the notion of an alternative or research hypothesis.

© CourseSmart

The research or alternative hypothesis we settle on is often a function of past research results or a particular theoretical perspective. Maybe previous research or a body of theory suggests that scores will increase or that one population will score higher than another. Maybe previous research or theoretical grounding suggests just the opposite. In some cases, previous research or theoretical statements may conflict, leaving us to expect a difference of some sort, but without any notion as to the direction of that difference.

When researchers have good reason to expect a change or difference in a particular direction, they are likely to opt for a directional alternative or research hypothesis. When there are conflicting results from previous research or contradictory suggestions from a body of theory, though, a non-directional alternative or research hypothesis is typically employed. Yes, it's still the null hypothesis that's tested, but there's also a research or alternative hypothesis in play, whether it's stated or not. In reality, it's often the research or alternative hypothesis that's driving the research in the first place. With that as a background, we can now explore the link between the alternative or research hypothesis and the topic of one-tailed and two-tailed tests.

© CourseSmart

One-Tailed and Two-Tailed Test Scenarios

In Chapters 7 and 8, we considered a variety of hypothesis-testing situations, but there was something that I neglected to tell you at the time. We approached those hypotheses as though they were *two-tailed test situations;* I just didn't bother to tell you that's what we were doing. I deliberately delayed any discussion of that for two reasons.

First, I firmly believe there's a limit to how much material anybody can absorb at once. Chapters 7 and 8 covered a lot of conceptual material, along with a hefty amount of calculations and computations. In short, you deserved a break. Second, I think it's wise to develop a solid grounding in the logic of hypothesis testing before dealing with the difference between a one-tailed and a two-tailed test. Assuming you now have a solid grasp of concepts such as the null hypothesis, calculated test statistics, and regions of rejection, we are ready to move ahead.

To illustrate the difference between the two approaches, let's take a simple example. Let's say we have a questionnaire that measures levels of religious participation (scores can range from 0 to 100). Assume we've collected scores from a sample of urban residents and a sample of rural residents and have obtained the results shown in Table 9-1.

As shown in Table 9-1, we have information from two samples (urban and rural residents), and we have a mean for each group (66.45 for the urban residents and 79.27 for the rural residents). Let's also assume that we've already calculated an estimate of the standard error (4.42). This test situation is appropriate for a difference of means test for independent samples—one in which we'd eventually calculate a *t* value for the difference of means. You'll recall that part of the *t* test procedure requires that we find the difference between the two

Table 9-1 Religious Participation Scores for Urban and Rural Residents

Urban Residents	Rural Residents
59	83
77	93
74	91
69	79
53	77
68	54
70	65
71	92
72	68
56	88
62	82

$\overline{X}$ for Urban Residents = 66.45

$\overline{X}$ for Rural Residents = 79.27

Estimate of the standard error of the difference of means = 4.42

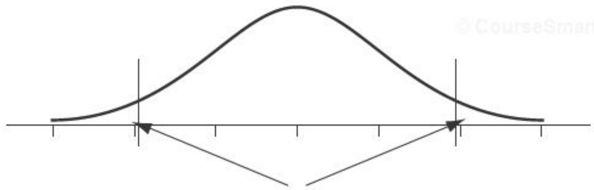
means, which we'll eventually divide by the estimate of the standard error of the difference of means to obtain our calculated test statistic. In reality, it isn't important how we go about finding the difference. We can subtract the mean of the rural residents from the mean of the urban residents; or we can approach the problem in the opposite fashion, subtracting the mean of the urban residents from the mean of the rural residents.

In this example, the mean of the rural residents is higher than the mean of the urban residents. In this case, if we subtract the mean of the urban residents from the mean of the rural residents, we'll get a difference with a positive sign (+). If we subtract the mean of the rural residents from the mean of the urban residents, we'll get a difference with a negative sign (–). As I mentioned before, it isn't important how we go about the subtraction process, just as long as we mentally keep track of which mean was subtracted from the other and how that relates to the sign of the difference. Assuming we can do that, let's look at some different possibilities.

Testing a Non-directional Research Hypothesis

Let's start with the assumption that we really don't know what to expect. Maybe some previous research suggests rural residents would have higher participation levels, but other research suggests that urban residents would have higher participation levels. In a case like that, it would make sense to approach the problem with a non-directional alternative or research hypothesis in mind.

In terms of the actual hypothesis test, our alternative or research hypothesis is a statement that we expect to find an extreme difference—a difference that's located somewhere in the outer regions of the distribution of possible differences. Since our alternative or research hypothesis is non-directional, however, it's actually a statement that we're open to that extreme difference being found at either tail of the distribution (see Figure 9-1). It can be an extreme difference that has a positive sign (+), or it can be an extreme difference that has a negative sign (–). We're open to a difference indicating that



In a two-tailed test situation, we're looking for a test statistic that falls in either of the critical regions—the positive region (+) at the upper end of the distribution or the negative region (–) at the lower end of the distribution.

Figure 9-1 Critical Regions in a Two-Tailed Test Situation

rural residents have higher scores, but we're also open to a difference indicating that urban residents have higher scores. When we're working with a non-directional research hypothesis—when we're open to finding a significant difference at either end of the distribution (in the region of rejection of either tail of the distribution)—we're in a **two-tailed test scenario.**

> **LEARNING CHECK**
>
> *Question:* What is a two-tailed test scenario, and when is it appropriate?
>
> *Answer:* A two-tailed test scenario is a research situation in which the researcher is looking for an extreme difference that could be located at either end of the distribution.
> A two-tailed test is appropriate when the alternative or research hypothesis is non-directional.

To develop a better understanding of all of this, let's take the problem all the way from a statement of the null and a non-directional research hypothesis, along with the level of significance, through the calculation of the test statistic and interpretation of results. Consider the following information.

Null hypothesis: There is no difference between the means of rural and urban residents; the means are equal.

Research hypothesis: There is a difference between the means of rural and urban residents.

Rural residents: mean ($\overline{X}_{\text{Rural}}$) = 79.27 Sample size ($n_{\text{Rural}}$) = 11
Urban residents: mean ($\overline{X}_{\text{Urban}}$) = 66.45 Sample size ($n_{\text{Urban}}$) = 11
Difference between the means (79.27 – 66.45) = 12.82

Estimate of the standard error of the difference of means = 4.42
Level of significance (alpha or α) = .05

Appropriate test is the t test for difference of means.
Degrees of freedom = 20
Critical value = 2.086

© CourseSmart

Note: The critical value was obtained from Appendix B: Family of t Distributions (Two-Tailed Test). Since we don't care about the direction of any difference (we don't care if it is positive or negative), we're in a two-tailed test situation.

Following the procedures outlined in Chapter 8 (for the difference of means tests with independent samples), we calculate the test statistic (t):

$$t = \frac{\overline{X}_{Rural} - \overline{X}_{Urban}}{s_{\overline{X}_{Rural} - \overline{X}_{Urban}}}$$

$$t = \frac{12.82}{4.42}$$

$$t = 2.90$$

Given the calculated test statistic of 2.90 and the critical value (from Appendix B) of 2.086, we conclude that we should reject the null hypothesis. In rejecting the idea of no difference, we find support for our research or alternative hypothesis—that there is a difference. By now, the basics of this process should be quite familiar to you: State the null, calculate the test statistic, check the critical value, and make a decision.

What's new at this point, though, is the role of a two-tailed test in the process. We found support for the idea that there's a difference by relying on a two-tailed test scenario. We set up the research in a way that allowed us to find a significant difference in either tail of the distribution. We were looking for an extreme t value—one that was so extreme that it would fall somewhere on the most extreme 5% of the distribution. Because we were operating on the basis of a two-tailed test, the 5% was equally divided between the upper and lower tails of the distribution (2.5% in the upper tail and 2.5% in the lower tail). And that, in short, is the essence of a two-tailed test scenario. It's a situation in which an extreme score can be located at either end of the distribution (see Figure 9-2).

© CourseSmart

Testing a Directional Research Hypothesis

Now let's consider a different situation—one involving a different alternative or research hypothesis. In this instance, let's assume that we expect to discover that rural residents have *higher* religious participation scores than urban residents. Since we're now hypothesizing (in the form of the research or alternative hypothesis) that the rural residents will have higher religious participation scores than the urban residents, we are specifying the direction

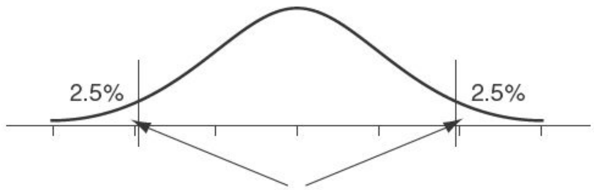


Figure 9-2 Allocation of the Extreme 5% of the Distribution in a Two-Tailed Test Scenario

of the expected difference. Therefore, we're relying on a *directional* research or alternative hypothesis.

As it turns out, the selection of a directional alternative or research hypothesis (as opposed to a non-directional one), results in a few changes in how we approach the test. These changes can be summarized as follows:

- The null hypothesis changes slightly. Instead of the null being a statement that we expect the two means to be equal, the null is now a statement that either the two means are equal or the mean score of the rural residents is *lower* than the mean of the urban residents. Remember, our alternative hypothesis in this case is that we expect the mean score for rural residents to be *higher* than the mean score for urban residents. Therefore, the null hypothesis (if it truly stands in opposition to the research or alternative hypothesis) is that we expect the mean score of the rural residents to be equal to or lower than the mean score of the urban residents.
- We're no longer looking for an extreme difference at either end of the distribution. Instead, we're looking for a *t* value at only one end of the curve—a *t* value that falls in only *one tail* of the distribution. Remember: We're asserting that we expect the mean score of the rural residents to be *higher* than the mean score of the urban residents. Assuming that we calculate the difference by subtracting the mean of the urban residents from the mean of the rural residents, we'll be looking for a positive difference (a difference that carries a + sign).
- Since we're only willing to accept a difference in a certain direction, we're in what is referred to as a **one-tailed test scenario.** To find our critical value, therefore, we'll use Appendix C: Family of *t* Distributions (One-Tailed Test).
- The critical value for a one-tailed test (found in Appendix C) will be different from the critical value for a two-tailed test (Appendix B).

To carry out our test, we follow the same formula as before. As a reminder, here are the same steps, repeated again:

© CourseSmart

$$t = \frac{\overline{X}_{Rural} - \overline{X}_{Urban}}{s_{\overline{X}_{Rural} - \overline{X}_{Urban}}}$$

$$t = \frac{12.82}{4.42}$$

$$t = 2.90$$

Note that our calculated t test statistic carries a positive sign because we're subtracting the mean of the urban residents from the mean of the rural residents. If we'd subtracted the mean of the rural residents from the mean of the urban residents, however, we'd have a negative t test statistic.

As noted previously, we're no longer looking for a significant t value at either extreme of the distribution of all possible t values. Instead, we're expecting it to be found in only one extreme region—in the upper end of the distribution, the area related to the positive t values. Remember: We're looking for a t value that falls in only one of the extreme regions; that's why we say we're in a one-tailed test situation.

Working at the .05 level of significance, with 20 degrees of freedom, the table for a one-tailed test (Appendix C) shows a critical value of 1.725. In this instance, the extreme 5% of the area under the curve is not divided between the two tails of the distribution. Rather, the extreme 5% is to be found *either* at the lower end of the distribution *or* at the upper end of the distribution. In other words, the extreme area in a one-tailed scenario is found on only one side of the distribution (see Figure 9-3).

Since our calculated t value is 2.90 and our critical value is 1.725, it appears that we're in good shape; we're on our way to rejecting the null hypothesis. At this point, though, we want to carefully consider the nature or direction of the difference that we found between the two means. In this case, the difference is consistent with our alternative or research hypothesis; the mean score for the rural residents was higher than the mean score for the urban residents. We're in a position to reject the null.

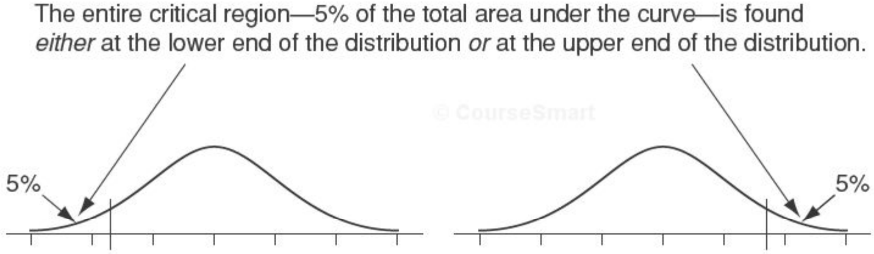


Figure 9-3 Allocation of the Extreme 5% of the Distribution in a One-Tailed Test Scenario

© CourseSmart

LEARNING CHECK

Question: What is a one-tailed test scenario, and when is it appropriate?

Answer: A one-tailed test scenario is a research situation in which the researcher is looking for an extreme difference that is located on only one side of the distribution. A one-tailed test is appropriate when the alternative or research hypothesis is directional.

In summary, we used a one-tailed test scenario, we found significant results, and the results were in the hypothesized direction (according to the alternative or research hypothesis we had advanced). If, however, we'd been working with a different alternative or research hypothesis—that urban residents would have higher participation scores than rural residents—we would have found ourselves facing a researcher's worst nightmare. We would have found a difference between the mean score of the urban and rural residents, to be sure, but the difference would have been opposite to the direction we hypothesized.

The logic involved in a one-tailed test can clearly get a little tricky, particularly when you get to the matter of whether or not the direction of the difference is consistent with the direction specified by the alternative or research hypothesis. Still, that's no reason to let it throw you. Just accept the fact that the logic is a little tricky—the sort of thing that eventually takes hold with repeated application.

Beyond that, you might also take some consolation in the fact that many researchers steadfastly avoid the use of one-tailed tests, and they do so for what they consider to be very good reason. To understand why some researchers, as a matter of course, just avoid using a one-tailed test, think back to the different critical values we encountered. When working in the two-tailed test scenario, our critical value was 2.086. In the one-tailed test situation, though, the critical value was 1.725. And that difference—the difference between the two critical values—brings us to the heart of the matter, at least in terms of how some researchers see the issue.

When you compare those values, one thing should become clear. Assuming any observed difference is consistent with the direction suggested by the alternative or research hypothesis, it's easier to achieve significant results in a one-tailed than in a two-tailed test scenario. As Gravetter and Wallnau (2002) note:

> A one-tailed test allows you to reject the null hypothesis when the difference . . . is relatively small, provided the difference is in the specified direction. A two-tailed test, on the other hand, requires a relatively large difference independent of direction. (p. 189)

© CourseSmart

In simple terms, some researchers prefer to work in a situation in which the bar of proof, so to speak, is as high as reasonably possible. They shy away from the easier situations (the ones made easier by the one-tailed test scenario) and opt instead for more demanding ones—the two-tailed test scenarios. Indeed, many researchers will resort to a one-tailed test situation only under very unusual circumstances, if at all. The conclusion to all this? It's simple. Don't be terribly surprised if you never encounter a one-tailed test in the remainder of your statistical journey. It's one of the finer points of hypothesis-testing logic, but one that you should know about, nonetheless. If you do encounter a one-tailed test along the way, at least you'll know what it's all about.

That said, we can leave the present discussion. Before we close out the chapter, though, we'll consider two final concepts: power and effect. Though not routinely considered by all statisticians, these concepts are particularly relevant to the field of experimental psychology, as well as other disciplines that frequently rely on experimental research designs.

Power and Effect

We'll begin with a brief review of Type I and Type II errors. First, take a moment to think about the difference between Type I and Type II errors. Here they are, spelled out for you again:

Type I Error: Rejecting the null hypothesis when it is true.
Type II Error: Failing to reject a null hypothesis when it is false.

Now, instead of thinking about the errors in terms of their definitions, think about them in terms of the possibilities we encounter whenever we approach a null hypothesis. First and foremost, always remember that there are two possibilities with respect to the null hypothesis: A null hypothesis is either true, or it is false. In other words, it doesn't make any difference what our research results eventually lead us to conclude. The fact remains that the null hypothesis, in reality, is either true or false.

Given that, our test of a null hypothesis can lead to four possibilities. We can either reject or fail to reject a true null hypothesis. By the same token, we can either reject or fail to reject a false null hypothesis. Table 9-2 illustrates these four possibilities.

Since the logic involved in all of this can get a little confusing at first, let's take apart Table 9-2, cell by cell. At first you may find the explanation I'm about to give you to be a little silly, but trust me on this: A little bit of silliness at this point can serve you well. Let's start with the information on the left-hand side of the table.

© CourseSmart

Table 9-2 Logical Possibilities in the Test of a Null Hypothesis

		The Null Hypothesis Is . . .	
		True	False
Based on the Test, We Either . . .	Fail to Reject the Null	***RESEARCH OBJECTIVE***	***TYPE II ERROR***
	Reject the Null	***TYPE I ERROR***	***RESEARCH OBJECTIVE***

The left-hand side of the table represents a situation in which the null hypothesis is actually true. To fully grasp this point, pretend for a moment that we're all-knowing, and we're watching someone else undertake a research problem. Because we're all-knowing, we're in a position to know that the null hypothesis is true. Maybe the null in this case is a statement that there's no difference between the means of two populations. The point is, we know that the null is in fact true. The team of researchers who're about to test the null, however, don't know what we know. All they know is that the null says there's no difference.

Now let's say the researchers test the null hypothesis, and the test results cause them to fail to reject the null. Remember: We know that the null is true. In a case like that—we know the null is true, and the researchers have failed to reject it—we might be inclined to pat the researchers on the back: Good job! The null was true, and you failed to reject it! You reached your objective! When a null hypothesis is true, researchers want to be in a position to fail to reject it.

Now, though, let's consider a different outcome. Let's say it's still the case that the null is true, but let's say the test results led the researchers to reject the null. In other words, the researchers found an extreme difference and rejected the null. Oops! The researchers just committed a Type I error! The researchers wouldn't know it, but we would. We know that the null is true, and the researchers just rejected the null.

In summary, there are two possible outcomes when a null hypothesis is true. One outcome is desirable; the other one isn't. If researchers fail to reject a true null hypothesis, the researchers are on solid ground. If, on the other hand, the researchers reject a true null hypothesis, they have committed a Type I error. So much for null hypotheses that are true. Now let's turn to the case of a false null hypothesis.

This time, focus on the right-hand side of the table. In this case, we know (because we're all-knowing) that the null hypothesis is false. In other

© CourseSmart

words, the null hypothesis says that there's no difference between two populations, but we know that there is a difference. As before, there are two possible outcomes.

Let's say a group of researchers test the null hypothesis and find a significant difference. As a result, they reject the null hypothesis. As before, we'd be inclined to pat the researchers on the back: Job well done! The null was false and you rejected it! That was a job well done in the sense that it would be consistent with another research objective. Just as researchers want to fail to reject true null hypotheses, they also want to reject false null hypotheses.

Now let's consider the other possibility. Let's say that our researchers failed to find a significant difference. We (because we're all-knowing) know there's a significant difference between the two populations, but the researchers failed to detect that difference. As a result, they failed to reject the null when, in fact, it's false. In other words, the researchers committed a Type II error.

Now all of that amounts to quite a bit of logic to digest, but digest it you must. Let me suggest that you take some time for a dark room moment to contemplate what we just covered. Imagine a situation in which you're hovering above a team of researchers and watching them work.

First, imagine a situation in which you know that the null hypothesis is true. Imagine that you know that there's no difference between two population means. Visualize the researchers failing to reject the null. Then imagine the researchers rejecting the null. Think about your reactions to what they have done.

Now imagine a situation in which you know that the null hypothesis is false. Imagine that you know that there is a difference between two population means. Visualize the researchers rejecting the null. Then imagine the researchers failing to reject the null. Think about your reactions to what they have done.

Let me suggest that you go through this visualization exercise over and over—thinking through all the possibilities time and time again, to the point that you're totally comfortable with them. Draw your own diagrams, or think up your own examples. If necessary, repeat the process to the point that you have a near-intuitive understanding of the logic. Assuming you feel comfortable with the full logic of Type I and Type II errors, we can move forward.

Remember what one of the major objectives is in the research process. If the null is false, a researcher will want to reject it. And that's where the concepts of *power* and *effect* come into play. First, we'll consider the matter of power; then we'll turn to the concept of effect.

The **power** of a statistical test is the ability of the test to reject a false null hypothesis. It's represented by the cell in the lower right-hand corner of Table 9-2. Remember: There are two possible outcomes to a statistical test when the null hypothesis is false. If we fail to reject a false null, we commit a Type II or beta (β) error. If we reject a false null hypothesis, however, we've reached our objective. Since β represents the probability of committing a Type II error, we can define the power of a test as $1 - \beta$ (or 1 minus the probability of a Type II error). To better understand all of this, consider the following example.

> **LEARNING CHECK**
>
> *Question:* What is the power of a test?
> *Answer:* Power is the ability of a test to reject a false null hypothesis.

© CourseSmart

Let's say that we want to know whether or not sleep deprivation has an effect on the amount of time required to complete a task. We could approach the problem as follows:

- Assemble two groups of research participants—one sleep deprived and one not.
- Ask members of each group to complete a task.
- Record the amount of time each participant spent completing the task.
- Perform the necessary calculations—a difference of means test—and arrive at a conclusion.
- Either reject the null or fail to reject the null.

As before, this process should be very familiar to you by now. What may not be apparent to you, however, is how the concepts of power and effect are involved. In truth, the goal of the research outlined above would be to detect the effect of sleep deprivation on task completion time. In the context of research, **effect** is the change in a measurement that is attributable to a treatment condition or stimulus of some sort.

To demonstrate that sleep deprivation has an effect, we'd have to find a significant difference between the two sets of recorded times and reject the null. It stands to reason that the larger the effect, the greater the likelihood that we'd do just that. A slight difference between the means would probably result in our failing to reject the null. A large difference, though, would probably result in our rejecting the null. All factors being equal, the larger the difference between the scores, the greater the likelihood that the null will be rejected. When the null is rejected, there's support for the notion that sleep deprivation has an effect on task completion time.

Of course there's always a possibility that sleep deprivation had an effect on task completion time, but we failed to pick up on that. In other words, there's always some probability of a Type II error—the null was false, but we failed to reject it. That possibility brings us to the heart of the issue.

> **LEARNING CHECK**
>
> *Question:* What is the definition of effect?
> *Answer:* Effect is the change in measurement that is attributable to a treatment condition or stimulus of some sort.

© CourseSmart

If a treatment condition of some sort has an effect on an outcome (that is, the null hypothesis is false), we want to be in a position to detect it. We want to detect the effect and subsequently reject the null. What we don't want is a situation in which an effect is present, but we've failed to detect it.

So how might we guard against such a situation? There are certain steps that we, as researchers, can take to increase the likelihood that we'll detect or pick up on the effect of a treatment condition. There are certain things we can do at the outset of a research problem to increase the *power* of the test—the likelihood that we will reject the null hypothesis when it's false.

First, we can increase our sample size. Assuming, for example, that the design of our research has us looking for a difference of some sort between two groups of research participants, an increase in the size of the samples would increase the likelihood that we'd detect any difference between the groups that actually exists.

Second, we can opt for a one-tailed test scenario. In this case, the entire region of rejection is found at one end of the distribution, and the necessary value to reject the null hypothesis (the critical value) is reduced slightly. In turn, there is an increase in the likelihood that our results will be significant (assuming they are in the hypothesized direction).

Similarly, we can increase our level of significance (regardless of whether we were working with a one-tailed or two-tailed test scenario), and thereby increase the likelihood of rejecting the null hypothesis. Unfortunately, this option has an associated cost—namely, an increase in the probability of a Type I error.

Finally, we can, to the extent possible, strive for highly controlled research situations—for example, situations in which participants in two groups are matched on relevant variables. By matching participants on a host of variables, we reduce the possibility that any difference might be *masked* by the influence of extraneous or outside variables. In short, highly controlled research designs increase the possibility that a treatment effect will, in fact, shine through.

Chapter Summary

We've explored quite a bit of conceptual material in this chapter, so it's probably appropriate to undertake a quick review of what we've just covered and a quick check of where you should be in your statistical education. For example, by now you should be totally comfortable with the role of the null hypothesis in scientific research and how it stacks up against an alternative or research hypothesis. Similarly, you should now be comfortable with the concept of an alternative or research hypothesis and how it is incorporated into your research efforts.

With your knowledge of alternative or research hypotheses, the notions of one-tailed and two-tailed test scenarios should now make sense. Similarly, the mystery as to why there are two tables for the distribution of t (Appendix B and Appendix C) should now be solved. Even if you never opt to use a one-tailed test, at least you'll be familiar with the logic that's involved in the application if you see or hear reference to it.

© CourseSmart

Finally, the more in-depth exploration into the logic of Type I and Type II errors should have given you a better understanding of research objectives in the larger sense—particularly the objective of rejecting false null hypotheses. With that understanding as a base, the concepts of power and effect should have taken on some meaning.

As we close this chapter and prepare for the next, we leave the more conceptual world and return to the world of calculations, computations, critical values, and such. At the same time, though, the logical underpinnings of hypothesis testing, including much of the material we just covered, should remain part of your thinking.

Some Other Things You Should Know

Some of the material we covered in this chapter will have more relevance to some readers than to others. For example, the material on power and effect has particular relevance for those in the field of experimental psychology. These issues are typically of minimal consequence in fields such as sociology or political science, which rely on large-scale surveys (and, consequently, large sample sizes). When power and effect are of consequence, however, additional resources should be consulted. For example, excellent discussions are to be found in Dunn (2001), Hurlburt (1998), Pagano (2001), and Howell (1995).

As to one- versus two-tailed tests, you should note that there are many situations in which the choice isn't even available. For example, the ANOVA procedure we will cover in the next chapter involves a comparison of three or more means. In a case like that, only one alternative or research hypothesis is appropriate—namely, a non-directional research hypothesis stating that the means vary across the different groups. The alternative or research hypothesis is not a statement that one mean will be higher or lower than another.

Finally, you should be aware of how the logic of one-tailed and two-tailed tests is dealt with when working with Z and the Table of Areas Under the Normal Curve. Should you find yourself in that situation, you can approach the problem with the assurance that the same logic applies. For example, the table presented in this text (Appendix A) really contains one-tailed values, because it only deals with one half of the normal curve. Note that a Z value of 1.96 actually has an associated proportion of .4750 (a percentage of 47.50%). We interpret a Z value of 1.96 as encompassing 95% of the area under the normal curve, but that's because we're mentally taking into account the area between Z values of -1.96 and $+1.96$.

If we were calculating Z and working in a one-tailed test scenario at the .05 level of significance, we would want to find the Z value (the critical value) that corresponded to *either* the upper *or* lower 5% of the area. The Z value associated with the upper or lower area (but not both) is the Z value that corresponds to .4500, or 45%. That Z value turns out to be approximately 1.64.

Key Terms

alternative hypothesis
directional hypothesis
effect
non-directional hypothesis
one-tailed test scenario
power
research hypothesis
two-tailed test scenario

Chapter Problems

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

General Thought Questions

1. A(*n*) ______________ hypothesis is a hypothesis that stands in opposition to the null hypothesis.
2. An alternative or research hypothesis that specifies the nature or direction of a hypothesized difference is considered a _________________.
3. A _________ tailed test scenario is appropriate when the alternative or research hypothesis is non-directional in nature.
4. A _________ tailed test scenario is appropriate when the alternative or research hypothesis is directional in nature.
5. A Type I error involves ____________ a null hypothesis when it is ______________.
6. A Type II error involves ____________ a null hypothesis when it is __________.

Alternative or Research Hypotheses Application Questions/Problems

1. A researcher is examining the possibility of a difference between the grade point averages of on-campus students and commuter students.
   a. What would be an appropriate null hypothesis?
   b.–d. What alternative or research hypotheses could be advanced?
2. A criminologist is examining the possibility of a difference between the length of sentences handed out to white and non-white defendants in first-offense drug trafficking cases.
   a. What would be an appropriate null hypothesis?
   b.–d. What alternative or research hypothesis could be advanced?
3. A political scientist is examining the possibility of a difference in the levels of voter participation in rural and urban areas.
   a. What would be an appropriate null hypothesis?
   b.–d. What alternative or research hypotheses could be advanced?

4. A team of environmental geographers believe that there's no significant difference between levels of water pollution in creeks in the northern and southern parts of the state, but they still want to conduct a test to verify this belief.
   - **a.** What would be an appropriate null hypothesis?
   - **b.–d.** What alternative or research hypotheses could be advanced?

© CourseSmart

One-Tailed and Two-Tailed Application Questions/Problems

1. Given the information in Appendix A, identify the critical values for Z in the following situations:
   - **a.** .05 level of significance; two-tailed test situation
   - **b.** .05 level of significance; one-tailed test situation
   - **c.** .01 level of significance; two-tailed test situation
   - **d.** .01 level of significance; one-tailed test situation
2. Given the information in Appendix B and C identify the critical values for t in the following situations:
   - **a.** 15 degrees of freedom and the .05 level of significance in a two-tailed test
   - **b.** 21 degrees of freedom at the .05 level of significance in a one-tailed test
   - **c.** 18 degrees of freedom at the .10 level of significance in a two-tailed test
   - **d.** 18 degrees of freedom at the .05 level of significance in a one-tailed test

© CourseSmart
© CourseSmart
© CourseSmart
© CourseSmart