STATISTICS
IN EVERYDAY LIFE

Think of hypothesis testing as analogous to gambling. Imagine you have been invited to participate in a game of chance involving coin flips: heads you win; tails your opponent wins. You would agree to participate in such a game only if you could **assume** that the coin was honest and that the probability of heads and tails was 0.5. What if your opponent flips 10 tails in a row? Think of these flips as a **test** of your original assumption that the coin is honest. At some point, as the coin shows tail after tail, you must compare the outcomes with your original assumption of honesty and make a **decision**: Either the coin is weighted toward tails or you have just witnessed a very rare series of events. If you conclude that the game is rigged, you have rejected the null hypothesis that there is no difference in the probabilities of heads and tails. While no one could blame you for walking away, note that there is a slight chance that your decision was wrong: It is possible (although very unlikely) that the game is not rigged and the coin is not weighted toward tails. In hypothesis testing, we also **make assumptions** (steps 1 through 3), **test** these assumptions in step 4, and make a **decision** based on probabilities in step 5.

Choosing a One-Tailed or Two-Tailed Test

The five-step model for hypothesis testing is fairly rigid, and the researcher has little room for making choices. Nonetheless, the researcher must still deal with two choices. First, he or she must decide between a one-tailed and a two-tailed test. Second, an alpha level must be selected. In this section, we discuss the former decision; we discuss the latter in the next section.

Choosing a One- or Two-Tailed Test. The choice between a one- and two-tailed test is based on the researcher's expectations about the population from which the sample was selected. These expectations are reflected in the research hypothesis (H_1), which is contradictory to the null hypothesis and usually states what the researcher believes to be "the truth." In most situations, the researcher will wish to support the research hypothesis by rejecting the null hypothesis.

The format for the research hypothesis may take either of two forms depending on the relationship between what the null hypothesis states and what the researcher believes to be the truth. The null hypothesis states that the population has a specific characteristic. In the example that has served us throughout this chapter, the null hypothesis stated, in symbols, "All treated alcoholics have the *same* absentee rate (7.2 days) as the community." The researcher might believe that the population of treated alcoholics actually has *less* absenteeism (their population mean is *lower than* the value stated in the null hypothesis) or *more* absenteeism (their population mean is *greater than* the value stated in the null hypothesis) or he or she might be unsure about the direction of the difference.

If the researcher is unsure about the direction, the research hypothesis would state only that the population mean is "not equal" to the value stated in the null hypothesis. The research hypothesis stated in our example ($\mu \neq 7.2$) was in this format. This is called a **two-tailed test** of significance because it means that the researcher will be equally concerned with the possibility that the true population value is greater than *or* less than the value specified in the null hypothesis.

In other situations, the researcher might be concerned only with differences in a specific direction. If the direction of the difference can be predicted or if the researcher is concerned only with differences in one direction, a **one-tailed test** can be used. A one-tailed test may take one of two forms depending on the researcher's expectations about the direction of the difference. If the researcher believes that the true population value is greater than the value specified in the null hypothesis, the research hypothesis would use the ">" or "greater than" symbol. In our example, if we had predicted that treated alcoholics had *higher* absentee rates than the community (or averaged *more* days of absenteeism), our research hypothesis would have been:

$$(H_1\colon \mu > 7.2)$$

where > signifies "greater than"

If we predicted that treated alcoholics had lower absentee rates than the community (or averaged *fewer* days of absenteeism than 7.2), our research hypothesis would have been

$$(H_1\colon \mu < 7.2)$$

where < signifies "less than"

One-tailed tests are often appropriate when programs designed to solve a problem or improve a situation are being evaluated. For example, if the program for treating alcoholics made them *less* reliable workers, the program would be a failure—at least on that criterion. In this situation, the researcher may focus only on outcomes that would indicate that the program is a success (i.e., when treated alcoholics have lower rates) and conduct a one-tailed test with a research hypothesis in the form $H_1\!:\!\mu < 7.2$. Or consider the evaluation of a program designed to reduce unemployment. The evaluators would be concerned only with outcomes that show a decrease in the unemployment rate. If the rate shows no change or if unemployment increases, the program is a failure and both of these outcomes might be considered equally negative by the researchers. Thus, the researchers could legitimately use a one-tailed test that stated that unemployment rates for graduates of the program would be less than (<) rates in the community.

One- versus Two-Tailed Test. In terms of the five-step model, the choice of a one-tailed or two-tailed test determines what we do with the critical region in step 3. As you recall, in a two-tailed test, we split the critical region equally into the upper and lower tails of the sampling distribution. In a one-tailed test, we place the entire critical area in one tail of the sampling distribution. If we believe that the population characteristic is greater than the value stated in the null hypothesis (if the H_1 includes the > symbol), we place the entire critical region in the upper tail. If we believe that the characteristic is less than the value stated in the null hypothesis (if the H_1 includes the < symbol), the entire critical region goes in the lower tail.

For example, in a two-tailed test with alpha equal to 0.05, the critical region begins at $Z(\text{critical}) = \pm 1.96$. In a one-tailed test at the same alpha level, the $Z(\text{critical})$ is $+1.65$ if the upper tail is specified and -1.65 if the lower tail is specified. Table 7.2 summarizes the procedures to follow in terms of

TABLE 7.2 One- Vs. Two-Tailed Tests, $\alpha = 0.05$

If the Research Hypothesis Uses	The Test Is	And Concern Is With	Z(critical) =
$\neq$	Two-tailed	Both tails	± 1.96
$>$	One-tailed	Upper tail	$+1.65$
$<$	One-tailed	Lower tail	-1.65

the nature of the research hypothesis. The difference in placing the critical region is graphically summarized in Figure 7.5, and the critical Z scores for the most common alpha levels are given in Table 7.3 for both one- and two-tailed tests.

Note that the critical Z values for one-tailed tests at all values of alpha are closer to the mean of the sampling distribution. Thus, a one-tailed test is more likely to reject the H_0 without changing the alpha level (assuming that we have specified the correct tail). One-tailed tests are a way of statistically having and eating your cake and should be used whenever (1) the direction of the difference can be confidently predicted or (2) the researcher is concerned only with differences in one tail of the sampling distribution. An example should clarify these procedures.

Using a One-Tailed Test. A sociologist has noted that sociology majors seem more sophisticated, charming, and cosmopolitan than the rest of the student body. A "Sophistication Scale" test has been administered to the entire student body and to a random sample of 100 sociology majors, and these results have been obtained:

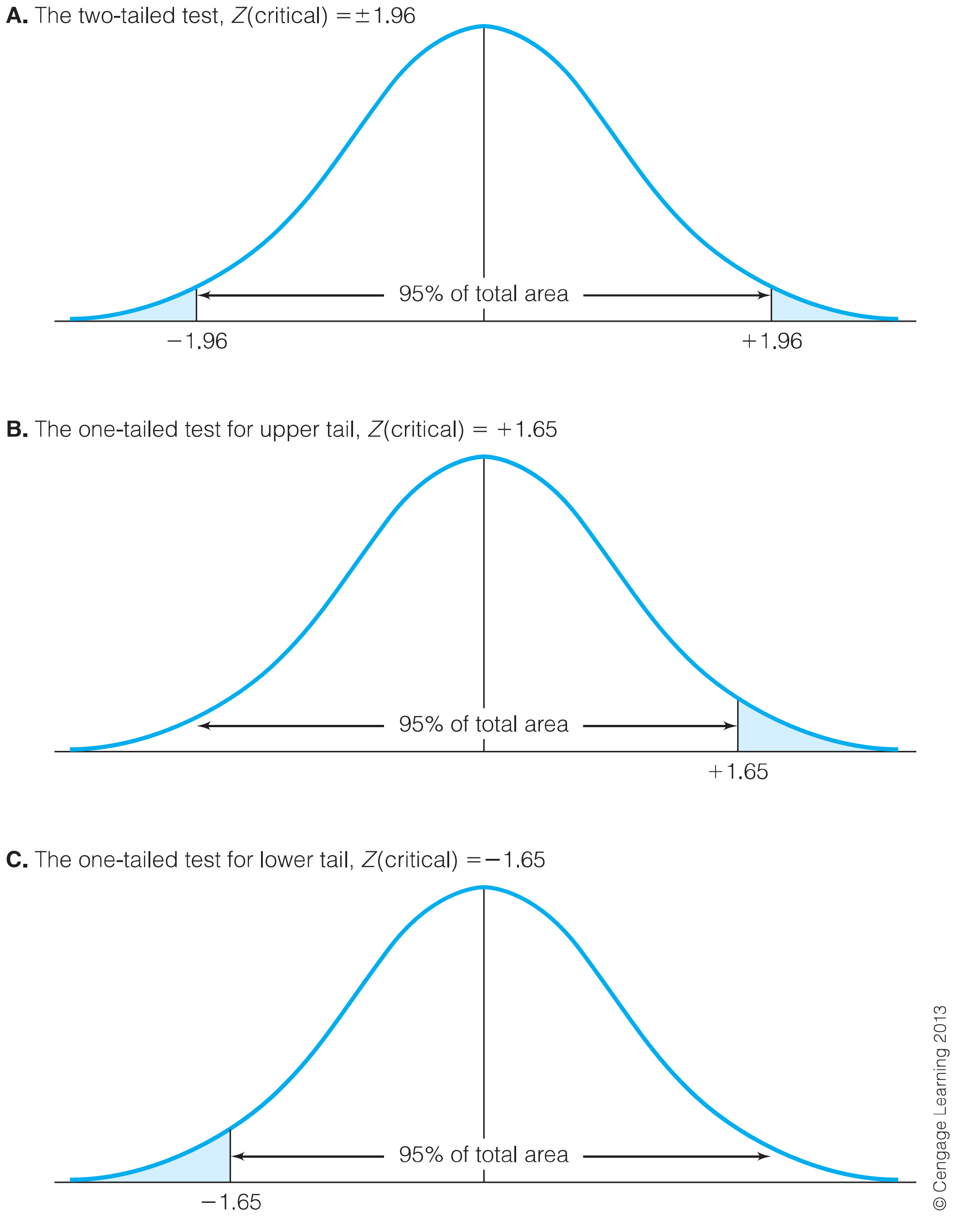
Student Body	Sociology Majors
$\mu = 17.3$	$\bar{X} = 19.2$
$\sigma = 7.4$	$N = 100$

We will use the five-step model to test the H_0 of no difference between sociology majors and the general student body.

TABLE 7.3 Finding Critical Z Scores for One-Tailed Tests

		One-Tailed Value	
Alpha	Two-Tailed Value	Upper Tail	Lower Tail
0.10	± 1.65	$+1.29$	-1.29
0.05	± 1.96	$+1.65$	-1.65
0.01	± 2.58	$+2.33$	-2.33
0.001	± 3.32	$+3.10$	-3.10
0.0001	± 3.90	± 3.70	-3.70

FIGURE 7.5 Establishing the Critical Region, One-Tailed Tests versus Two-Tailed Tests (alpha = 0.05)



Step 1. Making Assumptions and Meeting Test Requirements. Because we are using a mean to summarize the sample outcome, we must assume that the Sophistication Scale generates interval-ratio-level data. With a sample size of 100, the central limit theorem applies and we can assume that the sampling distribution is normal in shape.

Model: Random sampling
Level of measurement is interval-ratio
Sampling distribution is normal

Step 2. Stating the Null Hypothesis (H_0). The null hypothesis states that there is no difference between sociology majors and the general student body. The research hypothesis (H_1) will also be stated at this point. The researcher has predicted a direction for the difference ("Sociology majors are *more* sophisticated"), so a one-tailed test is justified. The one-tailed research hypothesis asserts that sociology majors have a higher (>) score on the Sophistication Scale. The two hypotheses may be stated as:

$$H_0: \mu = 17.3$$
$$(H_1: \mu > 17.3)$$

Step 3. Selecting the Sampling Distribution and Establishing the Critical Region. We will use Appendix A to find areas under the sampling distribution. If alpha is set at 0.05, the critical region will begin at the Z score $+1.65$. That is, the researcher has predicted that sociology majors are *more* sophisticated and that this sample comes from a population that has a mean *greater than* 17.3, so he or she will be concerned only with sample outcomes in the upper tail of the sampling distribution. If sociology majors are *the same as* other students in terms of sophistication (if the H_0 is true) or if they are *less* sophisticated (and come from a population with a mean less than 17.3), the theory is disproved. These decisions may be summarized as:

$$\text{Sampling distribution} = Z \text{ distribution}$$
$$\alpha = 0.05$$
$$Z(\text{critical}) = +1.65$$

Step 4. Computing the Test Statistic.

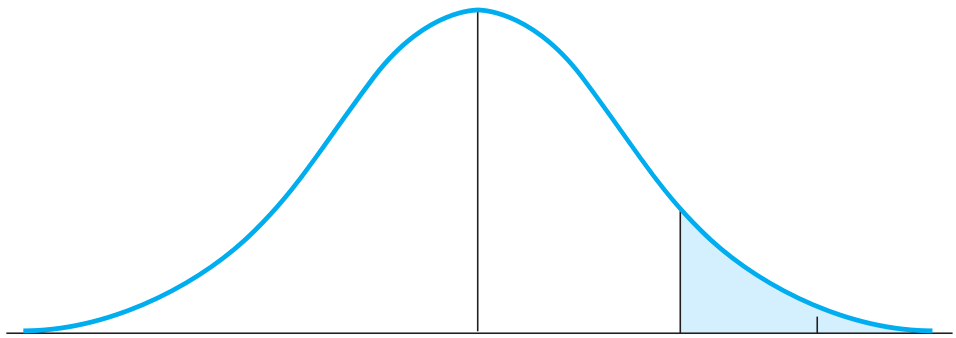
$$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$
$$Z(\text{obtained}) = \frac{19.2 - 17.3}{7.4/\sqrt{100}}$$
$$Z(\text{obtained}) = +2.57$$

Step 5 Making a Decision and Interpreting Test Results. In this step, we will compare the Z(obtained) with the Z(critical):

$$Z(\text{critical}) = +1.65$$
$$Z(\text{obtained}) = +2.57$$

The test statistic falls into the critical region, and this outcome is depicted in Figure 7.6. We will reject the null hypothesis because if H_0 were true, a difference of this size would be very unlikely. There is a significant difference between sociology majors and the general student body in terms of sophistication. Because the null hypothesis has been rejected, the research hypothesis (sociology majors are more sophisticated) is supported. *(For practice in dealing with tests of significance for means that may call for one-tailed tests, see problems 7.2 to 7.5, 7.8, 7.10, 7.13, 7.18, and 7.19).*

FIGURE 7.6 ***Z*(Obtained) Versus *Z*(Critical) (alpha = 0.05, one-tailed test)**



Selecting an Alpha Level

In addition to deciding between one-tailed and two-tailed tests, the researcher must select an alpha level. We have seen that the alpha level plays a crucial role in hypothesis testing. When we assign a value to alpha, we define what we mean by an "unlikely" sample outcome. If the probability of the observed sample outcome is lower than the alpha level (if the test statistic falls into the critical region), then we reject the null hypothesis.

How can we make reasonable decisions about the value of alpha? Recall that in addition to defining what will be meant by *unlikely,* the alpha level is the probability that the decision to reject the null hypothesis if the test statistic falls into the critical region will be incorrect. In hypothesis testing, the error of incorrectly rejecting the null hypothesis—or rejecting a null hypothesis that is actually true—is called **Type I error** (or **alpha error**). To minimize this type of error, use very small values for alpha.

To elaborate: When an alpha level is specified, the sampling distribution is divided into two sets of possible sample outcomes. The critical region includes all unlikely or rare sample outcomes. The remainder of the area consists of all sample outcomes that are not rare. The lower the level of alpha, the smaller the critical region and the greater the distance between the mean of the sampling distribution and the beginnings of the critical region. For the sake of illustration, compare the alpha levels and values for *Z*(critical) for two-tailed tests presented in Table 7.4. As you may recall, Table 6.2 also presented this information.

As alpha goes down, the critical region becomes smaller and moves farther away from the mean of the sampling distribution. The lower the alpha level,

TABLE 7.4 The Relationship Between Alpha and *Z*(critical) for a Two-Tailed Test

If Alpha Equals:	The Two-Tailed Critical Region Will Begin at *Z*(critical) Equal to
0.10	±1.65
0.05	±1.96
0.01	±2.58
0.001	±3.32

the harder it will be to reject the null hypothesis, and because a Type I error can occur only if our decision in step 5 is to reject the null hypothesis, the lower the probability of a Type I error. To minimize the probability of rejecting a null hypothesis that is in fact true, use very low alpha levels.

However, there is a complication. As the critical region decreases in size (as alpha levels decrease), the noncritical region—the area between the two *Z*(critical) scores in a two-tailed test—becomes larger. All other things being equal, the lower the alpha level, the less likely that the sample outcome will fall into the critical region. This raises the possibility of a second type of incorrect decision called **Type II error** (or **beta error**): failing to reject a null that is, in fact, false. The probability of a Type I error decreases as the alpha level decreases, but the probability of a Type II error increases. Thus, the two types of error are inversely related, and it is not possible to minimize both in the same test. As the probability of one type of error decreases, the probability of the other increases and vice versa.

It may be helpful to clarify in table format the relationships between decision making and errors. Table 7.5 lists the two decisions we can make in step 5 of the five-step model: We either reject or fail to reject the null hypothesis. Table 7.5 also lists the two possible conditions of the null hypothesis: It is either actually true or actually false. The table combines these possibilities into a total of four possible outcomes—two of which are desirable ("OK") and two of which indicate that an error has been made.

Let us consider the two desirable ("OK") outcomes first. We want to reject false null hypotheses and fail to reject true null hypotheses. The goal of any scientific investigation is to verify true statements and reject false statements.

The remaining two combinations are errors or situations we wish to avoid. If we reject a null hypothesis that is actually true, we are saying that a true statement is false. Likewise, if we fail to reject a null hypothesis that is actually false, we are saying that a false statement is true. Obviously, we would always prefer to wind up in one of the boxes labeled "OK"—and always reject false statements and accept the truth when we find it. However, remember that hypothesis testing always carries an element of risk and that it is not possible to minimize the chances of Type I and Type II errors simultaneously.

What all this means, then, is that you must think of selecting an alpha level as an attempt to balance the two types of error. Higher alpha levels will minimize the probability of a Type II error (saying that false statements are true), and lower alpha levels will minimize the probability of a Type I error (saying that true statements are false). Normally, in social science research, we will want to minimize a Type I error, and lower alpha levels (.05, .01, .001, or lower) will be used. The 0.05 level in particular has emerged as a generally recognized indicator

TABLE 7.5 Decision Making and the Five-Step Model

	Decision	
The H_0 Is Actually:	Reject	Fail to Reject
True	Type I, or α, error	OK
False	OK	Type II, or β, error