7

Hypothesis Testing With a Single Sample Mean

In the last chapter, you entered the world of inferential statistics when you learned how to make an inference about the population on the basis of what you knew about a sample. In this chapter, you'll find yourself using some of that same logic, but you'll go beyond a mere inference about a population value. In this chapter, you'll learn how statisticians formulate research questions, how they structure those questions, and how they put those questions to a test. In short, you'll learn about the world of *hypothesis testing*.

As we explore the world of hypothesis testing, we'll follow a path similar to the one we traveled in the last chapter. First we'll tackle hypothesis tests about a sample mean ($\overline{X}$) when we know the value of the standard deviation of the population (σ). Then we'll turn to tests about a sample mean ($\overline{X}$) when the population standard deviation (σ) is unknown. In the process, we'll make the same shift as we did before. First, we'll work with Z values and make a direct calculation of the standard error of the mean. In the second approach, we'll rely on t values and estimate the standard error of the mean.

In addition to learning about a particular statistical application, you'll learn about hypothesis testing in a general sense. In the process, you'll learn that the world of hypothesis testing has a language and a logical structure of its own. My guess is that you'll find that it's very different from anything you've ever experienced before. That's why it's a good idea to ease into the concepts gradually.

Before We Begin

To get right to the point, think about what you just covered. You dealt with confidence intervals. You dealt with concepts such as the mean, the standard deviation, and the standard error (calculated and estimated). You used those concepts when constructing confidence intervals. Now, though, we're getting ready to shift gears. Yes, we're going to rely on many of the same concepts, but our purpose will be very different. We're about to move into the world of hypothesis testing.

Before we start, let me emphasize three major points. First, hypothesis testing involves an approach to logic that may strike you as a little strange. I just ask you to remember that as you work your way through the chapter. Secondly, you need to have an objective, open mind if you really want to understand hypothesis testing. If you're inclined to hold opinions or make statements in the absence of facts, you might find the next chapter a bit bothersome. Finally, the material that you're about to encounter should probably be taken in bits and pieces. My advice is that you read about a concept or notion, think about that concept or notion, and then reread and rethink again. The concepts are important enough to warrant that sort of approach.

Setting the Stage

Researchers may want to compare a sample mean to a population mean for any number of reasons. Consider the following examples.

Let's say a researcher is about to analyze the results of a community survey, based on the responses of 50 registered voters. Assuming he/she has some knowledge about the entire population (for example, the mean age of all registered voters in the community), the researcher might start by comparing the mean age of the sample and the population, just to determine if the sample is reasonably representative of the population.

Maybe a criminologist is interested in the average sentence length handed out to first-time offenders in drug possession cases. A national study, now almost two years old, reports that the average sentence length is 23.4 months, but the criminologist wants to verify that the reported average still applies.

In yet another example, maybe a team of industrial psychologists is interested in the productivity of assembly line workers. Historical data, based on the performance of all workers over the past three years, indicate that workers will (on average) produce 193.80 units per day. The psychologists, however, believe that the level of productivity may be different for workers who've been given the option of a flextime schedule. Taking a sample of productivity records for those working on a flextime schedule, the psychologists can compare the sample mean with the historical population mean.

Those are just some of the situations appropriate for a hypothesis test involving a single sample mean. There are actually many different hypothesis-testing procedures—some involving a single sample mean, some based on two sample means, and still others that deal with three or more sample means. For the moment, though, we'll deal with the single sample situation. It's a fairly straightforward sort of application and well-suited as an introduction to the logic of hypothesis testing.

A Hypothesis as a Statement of Your Expectations: The Case of the Null Hypothesis

You've probably heard of or used the word **hypothesis** before, and you may have the notion that a hypothesis is a statement that you set out to prove. That understanding may work when it comes to writing a term paper or an essay, but it's far removed from the technical meaning of a hypothesis in a statistical sense. In truth, a statistician isn't interested so much in a hypothesis as in the **null hypothesis.**

Statisticians are forever attempting to put matters to a test, and they use a null hypothesis to set up the test. That's where we'll begin—with the notion of the null hypothesis. To be fair, though, you deserve an advance warning. You may think the logic behind the null hypothesis is totally backwards and, at times, convoluted. If that's the way it strikes you, rest assured your reaction isn't unusual. Indeed, my experience tells me that many students find the logic of hypothesis testing to be a little rough going at the outset. You may have to go over it again and again and again. What's more, you may have to take some time out for a few dark room moments along the way. Let me encourage you—do whatever you need to do. The logic of hypothesis testing is an essential element in the world of inferential statistics.

Assuming you're ready to move forward, let's take a closer look at the concept of a null hypothesis. As it turns out, the null hypothesis is a statement that can take many forms. In some cases, the null hypothesis is a statement of *no difference* or a statement of *equality*. In other cases, though, it's a statement of *no relationship*. How a null hypothesis is stated is a function of the specific research problem under consideration. In general, though, and to get on the road to understanding what the null is all about, it's probably best to begin by thinking of it as a *statement of chance*.

LEARNING CHECK

Question: What is a null hypothesis, and how might it be expressed?

Answer: A null hypothesis is the hypothesis that is tested. It can be a statement of no difference, a statement of chance, or a statement of no relationship.

Whether you realize it or not, you're already fairly familiar with the concept of chance or probability. For example, if I asked you to tell me the probability of pulling the ace of spades out of a deck of 52 cards, you'd tell me it's 1 out of 52 (since there is only one ace of spades in the deck). If I asked you to tell me the probability of having a head turn up on the flip of a coin, you'd likely tell me it's 50%—there's a 50/50 chance of it being a head. Of course, all of this assumes an honest deck of cards, or an honest coin.

In short, all of us occasionally operate on the basis of a system of probabilities—we know what to expect in the case of chance. In fact, that's frequently the only thing we know. For example, we don't have one set of probabilities for a slightly dishonest coin and another set of probabilities for an even more dishonest coin. All we have is a set or system of probabilities based on chance.

Now, to consider yet another example of a statement of chance, think about the normal curve. It is, after all, a *probabilistic distribution;* it gives you a statement of probabilities associated with various portions of the curve. For example, there's a 99% chance, or probability, that a score in a normal distribution will fall somewhere between 2.58 standard deviations above and below the mean. By the same token, there's only a 1% chance that a score would fall *beyond* ±2.58 standard deviations from the mean.

To convince yourself of this, think about what you already know about a Z score of, let's say, −2.01. You already know that it would be an extremely low Z score (and therefore has a low probability of occurring). You know that for the following reasons:

- The Z values of +1.96 and −1.96 enclose 95% of the area under the curve.
- Therefore, only 5% of the area under the curve falls outside those values.

- Only 5 times out of 100 would you expect to get a Z value of more than +1.96 or less than –1.96.
- The extreme 5% would actually be split between the two tails of the distribution—2.5% in one tail and 2.5% in the other tail.
- Since a Z value of –2.01 is beyond the value of –1.96, you know that the probability of such a Z score occurring is fairly rare—indeed, it would have a probability of occurring less than 2.5 times out of 100 (<2.5%).

Assuming all of that made sense, let me urge you to begin thinking of extreme scores or values as nothing more or less than a score or value that has a very low probability of occurring. When a statistician views a score or value as *extreme,* it means that it has a low probability of occurrence.

> **LEARNING CHECK**
>
> *Question:* What is an extreme score or value?
> *Answer:* A score or value that has a low probability of occurrence.

The material you just covered is the sort of thing you'll want to keep in the back of your mind. For the moment, though, it shouldn't concern you how it relates to where we're going. In fact, you might do well to recall what I mentioned earlier—namely, that statistical analysis is sometimes best learned when you don't know where the road is leading. That said, let's take the next step along the road—in fact, a giant step. Let's jump headlong into a statistical application.

Single Sample Test With σ Known

Let's begin with a closer look at an earlier example. Assume for the moment that we're part of a team of industrial psychologists interested in how the introduction of a flextime program may have affected the productivity of assembly line workers. Instead of a set shift for every worker (such as 9:00 to 5:00), the flextime program allows each worker to select a specific shift (for example, 7:00 to 3:00, 8:00 to 4:00, or 10:00 to 6:00). Our historical information on worker productivity over the past three years shows that workers, on average, assemble 193.80 units per day, with a standard deviation of 31.55 units. Since those values are the result of complete records over the past three years, we can treat them as population values: $\mu = 193.80$ and $\sigma = 31.55$.

Let's also say that we've selected a random sample of productivity reports on 50 workers who took advantage of the flextime option. Our interest is in whether or not there's a significant difference between the productivity of flextime workers and the historical level of productivity. Calculating the mean level of productivity for our sample ($\overline{X}$), we determine that it's 202.94. There's obviously a difference between the two means (μ and $\overline{X}$). After all, the mean of the population (the population of all workers over the past three years) is 193.80 units, and the mean of the sample is 202.94 units.

Population of All Workers (Past Three Years)	**Sample of 50 Flextime Workers**
Mean or μ = 193.80 σ = 31.55	Mean or $\overline{X}$ = 202.94

μ as opposed to $\overline{X}$

A non-statistician might think about the example we're considering and say, "OK, I get it. We've compared a sample mean to a population mean to see if there's a difference. There is a difference, so that's that." With our background in statistics, however, we know there's far more to the situation than meets the eye. Once again, we're back where we've been before. We have only one sample mean in front of us—one mean out of an infinite number of sample means that are possible. The real question is whether or not our sample mean is *significantly* different from the population mean. As you'll soon discover, a significant difference is one that's so great that it has a low probability of having occurred by chance (or sampling error).

To understand all of this, let's start with the notion that we have to exercise a bit of caution. If the sample mean ($\overline{X}$) is higher or lower than the population mean (μ), we can't just jump to the conclusion that the level of productivity for flextime workers is really higher or lower than the historical mean. After all, our sample mean is just one mean, and it's subject to sampling error. So how do we determine whether the difference between the sample and population mean is noteworthy? How do we determine whether the difference is significant or just a matter of sampling error? We'll get the answers to those questions by testing a hypothesis.

LEARNING CHECK

Question: What is a significant difference?
Answer: A significant difference is one that is so great that it has a low probability of having occurred by chance.

Refining the Null and Phrasing It the Right Way

As I mentioned before, statisticians test hypotheses by testing the *null hypothesis*. Since a null hypothesis is often a statement of equality (no difference), let's see how that plays out in the present example.

In the case we're considering here, the null hypothesis would be a statement that there's no difference between the mean of the population of flextime workers and the historical mean of the population of workers in general. At first glance, there appears to be a difference—after all, the population mean (μ) is 193.80 and the sample mean ($\overline{X}$) is 202.94—but the question really goes deeper. The real question is:

> **How likely is it that we would have obtained a sample mean ($\overline{X}$) of 202.94 from a population having a mean (μ) of 193.80 and a standard deviation (σ) of 31.55?**

In other words, if flextime workers are really part of the general population of workers (that is, they're not significantly different), then how likely is it that a sample of flextime workers would exhibit a sample mean of 202.94?

To answer this question, we'll eventually compare the sample mean with our expectation. If the sample mean is reasonably close to what we'd expect (based on chance or sampling error), we can attribute the difference to sampling error. If, on the other hand, the sample mean isn't reasonably close to what we might expect (based on chance or sampling error), we'll have reason to believe that the productivity of flextime workers is significantly different from the historical pattern. The question, therefore, is really whether or not a particular sample mean (whatever it might be) could reasonably come from a population with a mean (μ) of 193.80 and a standard deviation (σ) of 31.55.

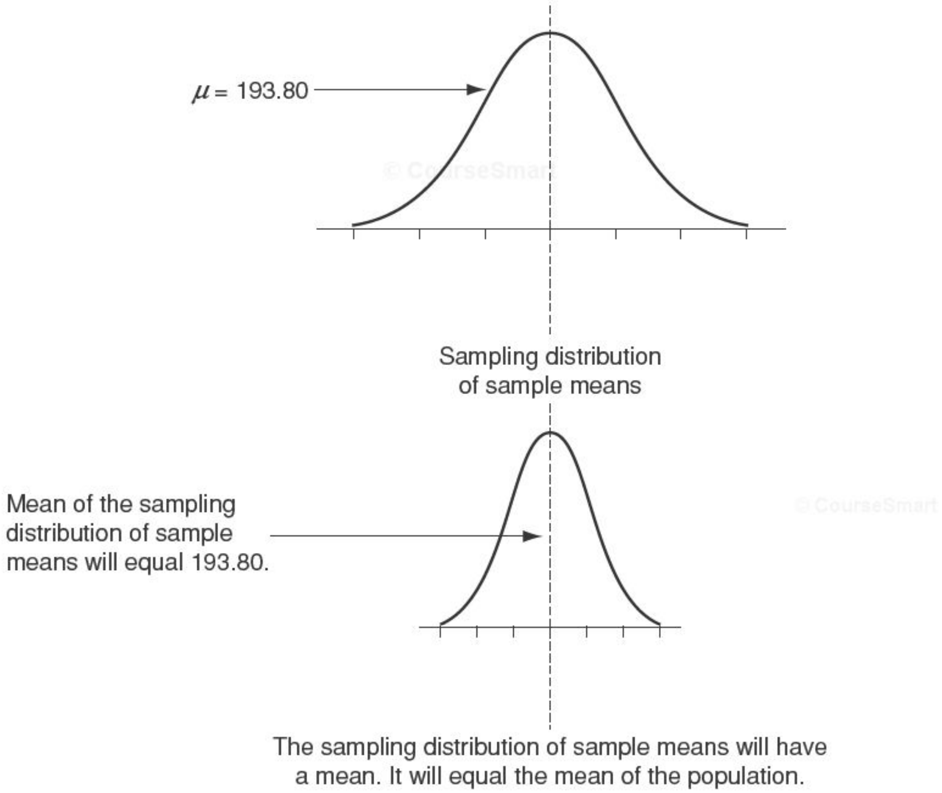
Right now we're focused on the null hypothesis, and the null hypothesis is a statement of our expectation. Therefore, our null hypothesis is a statement that the mean of the population (μ) is equal to 193.80, or

$$H_0: \mu = 193.80$$

In other words, we're advancing the null hypothesis that the mean of the population of flextime workers is equal to 193.80 (the same mean as the mean of the historical population of workers). In doing so, we're also advancing the notion that the mean of the sampling distribution of sample means is equal to 193.80. Note that the null hypothesis is designated as H_0.

The Logic of the Test

Assuming you've digested all of that, let's return to the fact that we have a sample mean ($\overline{X}$) of 202.94, and we're back in familiar territory. We have to determine where our sample mean is located in terms of a distribution of many different means. The Central Limit Theorem tells us that the mean of a sampling distribution of sample means will equal the mean of the population (μ). The null hypothesis states that the mean of the population (μ) is 193.80, so (recalling what the Central Limit Theorem tells us) we would expect the mean of the sampling distribution of sample means to equal 193.80. Take a close look at Figure 7-1 to better grasp the logic underlying our approach here.



The central question: Where does our sample mean of 202.94 fall in relation to all other sample means in a sampling distribution of sample means?

Figure 7-1 What The Central Limit Theorem Tells Us and The Central Question

If you're inclined to just move ahead, without taking a serious look at Figure 7-1, let me caution against that. The illustration is there to demonstrate a basic element in the underlying logic that's in play here. Here's the logic:

- There is a known historical mean level of productivity for all workers (193.80 units per day).
- That known historical mean is treated as the mean of the population.
- We want to know if there is a difference between the level of productivity of flextime workers and the historical mean level of productivity.
- We start with the assumption that the mean for the population of flextime workers would be the same mean as the historical mean for all workers (i.e., we assume that the mean for the population of flextime workers would be 193.80 units).
- Given that assumption (and based on the Central Limit Theorem), we assume that the mean of the sampling distribution of sample means (based

on repeated samples of flextime worker records) would equal the historical mean of the population (193.80).

- We'll compare our sample mean (202.94) to the assumed mean of the sampling distribution of sample means (193.80), and we'll calculate the difference.
- We'll evaluate the difference by expressing the difference in standard error units.
- To express the difference in standard error units, we'll simply divide the difference by the standard error of the mean.

Our task is to determine where our sample mean of 202.94 would fall along a sampling distribution of different means—the many different means that we might get if we were to construct a sampling distribution of sample means. More specifically, we're going to evaluate our sample mean on the basis of a sampling distribution of sample means that has an assumed mean of 193.80. Remember: Our null hypothesis is a statement that we expect the mean of the population of flextime workers to equal 193.80. The Central Limit Theorem tells us that the mean of the sampling distribution of sample means will equal the mean of the population, so we're really working with an assumption that the mean of the sampling distribution of sample means is equal to 193.80.

If we discover that our sample mean ($\overline{X}$) of 202.94 isn't that unusual (compared to all sample means that would be possible), we can attribute the observed difference to chance (or more correctly, sampling error), and conclude that the null hypothesis is true. If, on the other hand, it appears that our sample mean is fairly extreme (in comparison to an assumed mean of 193.80), then we'd be inclined to believe that flextime workers do exhibit a significantly different level of productivity. Accordingly, we'd be inclined to reject the null hypothesis.

Now all of that represents a central notion in the matter of hypothesis testing, so let me emphasize it again. If the observed difference is relatively small, we'd be inclined to believe that the null hypothesis is true. If, however, the difference is relatively large, we'd be in a position to reject the null hypothesis. In doing so, we'd be rejecting the idea that the mean level of productivity for the flextime workers is equal to the historical level of productivity. In other words, we'd actually be suggesting that the population of flextime workers is somehow different from the population of all workers.

If that's where we end up (i.e., we reject the null hypothesis), we can say that our results are statistically significant. In other words, we can say that we found a significant difference between the two values. We'll eventually get around to learning more about how statisticians use the term *significant*. For the moment, though, let's return to the problem at hand.

Applying the Test

The central question obviously turns on the difference between the two values ($\overline{X}$ and μ) and whether the difference is extreme. Would a difference of 5.89 units per day be enough to call it extreme? What about a difference of 14.10 units per day? What if the sample mean were 212.29 units per day?

What about a mean of 239.88? Would that be different enough from the population mean to call it extreme?

As it turns out, there's a single answer to all those questions. Whether or not a sample mean represents an extreme departure from the population mean is a relative matter. It's relative in terms of how far the sample mean departs from the population mean (which is also the mean of the sampling distribution of sample means), in terms of standard deviation units.

You've learned, for example, that ±1.96 standard deviations on a normal curve (Z values of ±1.96) will take you pretty far out in the distribution or along the baseline. As a matter of fact, ±1.96 standard deviation units from the mean will take you far enough along the baseline of the distribution to encompass 95% of the area or cases. To refresh your memory on all of this, think back to what you learned in Chapter 4 about the notion of extreme scores. You ultimately learned that the real value of a Z score was found in its universal applicability. You learned that a Z score of, let's say, 2.91 (either + or – 2.91) would be extreme, whether you were referring to dollars, ounces, pounds, miles per hour, or anything else—including levels of productivity, expressed as the number of units produced per day.

It should now be clear that the central task is actually a rather simple one. All we have to do is calculate the difference between our sample mean and the population mean (which is also the mean of the sampling distribution of sample means). Then we translate that difference into a ratio that expresses the difference in standard deviation units—or more correctly, standard error of the mean units. As you learned earlier, the standard error of the mean is simply the standard deviation of the sampling distribution of sample means.

> **LEARNING CHECK**
>
> *Question:* What is the central question surrounding a hypothesis test involving a single sample mean and a population mean?
>
> *Answer:* The central question is whether the difference between the two means is extreme—whether the difference is significant.

Calculation. In the example we're considering now, we know the values of the standard deviation of the population (σ = 31.55) and the sample size (n = 50). Therefore, the calculation of the standard error of the mean ($\sigma_{\bar{x}}$) is straightforward. As you know from the last chapter, it's simply a matter of dividing σ (31.55) by the square root of the sample size (the square root of 50, or 7.07). Here's the formula again, just as you encountered it before, along with its calculation in the present instance:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Armed with the value of the standard error of the mean ($\sigma_{\overline{X}} = 4.46$), we're now equipped to properly evaluate the difference between our sample mean and the assumed mean of the sampling distribution of sample means. First, we take note of the difference between the assumed mean ($\mu = 193.80$) and the sample mean ($\overline{X} = 202.94$). Subtracting the assumed mean from the sample mean, we discover that the difference equals 9.14 (i.e., the mean for the flextime workers is 9.14 units higher than the assumed mean of the population, which is also the assumed mean of the sampling distribution of sample means). From our earlier calculation, we determined that the standard error of the mean equals 4.46. To express our observed difference in terms of standard error units, we simply divide the observed difference (9.14) by the standard error of the mean (4.46). The calculation amounts to a conversion of a sample mean to a Z score, so the symbol Z now appears in the formula.

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

$$Z = \frac{202.94 - 193.80}{4.46}$$

$$Z = 2.05$$

We see that the difference of 9.14 (units produced per day) translates into a difference of 2.05 standard error units. The result (the difference divided by the standard error of the mean) is equivalent to a Z score (or a Z ratio) of +2.05. The logic behind the process we just went through may be summarized as follows:

1. Determine the difference between the two means—the sample mean minus the assumed population mean (which is also assumed to be the mean of the sampling distribution of sample means).
2. Calculate the standard error of the mean (divide the known σ by the square root of the sample size).
3. Convert the difference between the sample and population mean into a Z ratio by dividing the difference by the standard error of the mean.

Since the sampling distribution of sample means is known to approach a normal curve, we're in a position to evaluate whether or not the observed difference of 2.05 standard error units (or a Z of +2.05) is extreme. By now, you should already know the answer to that question: By most standards, a Z value of +2.05 is extreme. That's something you should already know at an intuitive level. After all, a Z value of +2.05 is more extreme than a value of +1.96, and only 5% of the sample means could be expected to fall beyond Z values of ±1.96. Since our calculated value of $Z = +2.05$ exceeds the value of +1.96 (see Figure 7-2), we know that it falls in a very extreme region of the curve (more specifically, a more extreme region in the upper portion of the curve). The difference is extreme enough that we'll reject our null hypothesis.

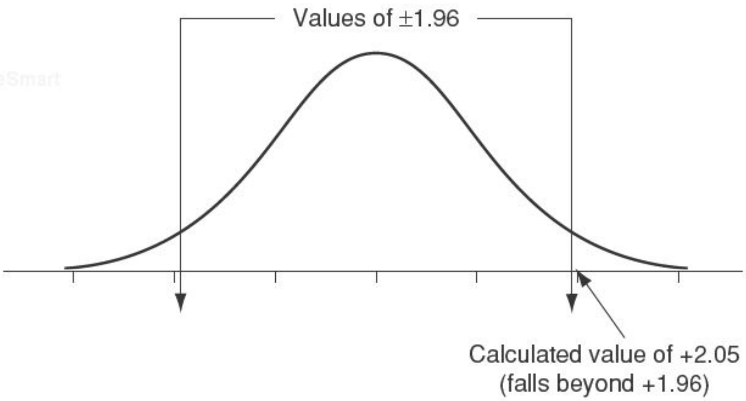


Figure 7-2 Location of Calculated Value of +2.05

Interpretation. By rejecting the null hypothesis, we're really rejecting the notion that the population of flextime workers would have a mean ($\overline{X}$) of 193.80. Accordingly, we're in a position to suggest that the productivity level of a population of flextime workers is different from the productivity level of a population of all workers (at least all workers over the past three years). But before we jump straight to that conclusion, though, let's think about how we arrived at our conclusion in the first place. Remember: It all goes back to whether an observed difference between the sample mean and a population mean is an extreme difference.

LEARNING CHECK

Question: In the example involving the flextime workers, what does it mean to reject the null hypothesis?

Answer: To reject the null hypothesis is to reject the idea that the sample mean came from a population having a mean ($\overline{X}$) of 193.80.

Levels of Significance, Critical Values, and the Critical Region

If you think about the approach that we took, we calculated our test statistic first, and then we turned to the issue of whether or not the difference was extreme. In truth, however, statisticians actually determine what constitutes an extreme value by setting a *level of significance* before they perform any calculations. Additionally, statisticians commonly set the level of significance (also known as the alpha level or α) at the .05 or .01 level. I'll eventually give you a definition of *level of significance,* and I'll also tell you why it's set in advance. For the moment, though, let's continue with our example.

To say that we're working at the .05 level of significance, for example, is to say that we're looking for a Z value or ratio (our final answer) that is so extreme it would occur by chance only 5% of the time, or less, if the null hypothesis is true. If we discover that our calculated test statistic, or Z ratio, is equal to or more extreme than ± 1.96, we can say that the value is extreme. We can say it is extreme because it's apt to occur by chance less than 5 times out of 100, assuming the null hypothesis is true. Remember: We'd expect 5% of the sample means (when converted to Z ratios) to fall at or beyond Z values of ± 1.96, assuming the null is true. In fact, we'd expect a Z value equal to or greater than $+1.96$ only 2.5 times out of 100, and we'd expect a Z value equal to or less than -1.96 only 2.5 times out of 100 (because the extreme 5% is evenly split between the two sides of the distribution).

The same sort of reasoning would apply if we were working at the .01 level of significance. In a case like that, we'd be focused on Z values of ± 2.58. Only 1% of sample means would be found at or beyond a Z value of ± 2.58, assuming the null is true. One-half of the extreme 1% (.5%) would be found on one side of the distribution and one-half of the extreme 1% (.5%) would be found on the other side.

LEARNING CHECK

Question: What levels of significance are commonly used by statisticians?

Answer: The .05 and .01 levels of significance.

To a statistician, a critical value like ± 1.96 or ± 2.58 is referred to as just that—the **critical value**. If our **calculated test statistic** (our result) meets or exceeds the critical value, then we can legitimately think of our observed difference as being extreme. If we can do that, we can reject the null hypothesis.

To truly appreciate the underlying logic, think back to the object of the exercise. It was to test a null hypothesis—for example, the null hypothesis that a sample mean came from a population with a certain mean value. To test the hypothesis, the difference between the sample mean and assumed mean of the sampling distribution of sample means (which is also the assumed mean of the population) was found, and the difference was evaluated. By determining a level of significance before the actual test, statisticians set the standard in advance. If the observed difference between the two means (expressed as a Z ratio) is large enough (and therefore meets the standard), the null hypothesis will be rejected. In rejecting the null hypothesis, the statistician is in a position to say that he/she has *significant* results.

In our flextime example, assuming we had predetermined our level of significance to be .05 (alpha, or $\alpha = .05$), we could say we have *rejected the null* *hypothesis at the .05 level of significance.* The same fundamental logic underlies all hypothesis-testing situations, so let's review the process, step by step. Take a look at the summary in Table 7-1, and focus carefully on each step of the process.

What if we had selected the .01 level of significance ($\alpha = .01$) for our test? If you think back to those familiar values you first encountered in Chapter 4, you'll realize that the critical value at the .01 level of significance would be ±2.58. Those values would enclose 99% of the area or cases under the normal curve. That information, in turn, tells us that only 1% of the area or cases would be found at or beyond ±2.58. In other words, at the .01 level of significance, we'd look for a calculated test statistic (Z) that is so extreme that it is apt to occur less than 1% of the time by chance, assuming the null hypothesis is true. If our calculated test statistic (Z) was equal to or beyond the critical value of ±2.58, we could reject our null hypothesis at the .01 level of significance.

As it turned out, of course, our calculated test statistic (Z) was +2.05—a value that would not meet or go beyond a critical value of ±2.58. Our calculated value (2.05) is close to the value of 2.58, but it doesn't meet or exceed 2.58. Therefore, we would *fail to reject the null hypothesis at the .01 level of significance*. Our results may have been significant at the .05 level of significance, but our results would lead us to a different conclusion at the .01 level of significance—we would *fail to reject the null hypothesis*.

There's obviously an important lesson in all of that—namely, that whether or not our results are statistically significant is largely a matter of the level of significance that we set in advance of our hypothesis test. Significant results at one level of significance may not be significant at a different level of significance. Now you can begin to appreciate why statisticians typically set the level of significance in advance. This procedure allows them to predetermine what will be necessary for them to consider their results as *significant*—an approach that reflects the presumed objectivity of the scientific pursuit.

Table 7-1 Summary of the Hypothesis Testing Process

Formulate the null hypothesis.	H_0: $\mu = 193.80$
Determine a level of significance.	$\alpha = .05$
Identify the critical value.	±1.96
Calculate the test statistic.	$Z = \dfrac{\overline{X} - \mu}{\sigma_{\overline{x}}}$
Evaluate the test statistic in light of the critical value.	Compare calculated Z to critical value ±1.96.
Make a decision about null hypothesis.	Reject or fail to reject the null hypothesis.

Now let me say a final word about critical values. The best way to understand the critical value is to regard it as a point that marks the beginning of what's referred to as the **critical region**. The critical region, in turn, is the portion of the sampling distribution (such as the sampling distribution of Z) that contains all the values that allow you to reject a null hypothesis. For that reason, we refer to the critical region as the **region of rejection**. If our calculated test statistic (Z) is equal to or falls beyond the critical value, it has fallen into the critical region—the region that allows us to reject the null hypothesis.

> LEARNING CHECK
>
> *Question:* How are the critical value and the critical region related?
>
> *Answer:* The critical value is the beginning of the critical region. If our calculated test statistic meets or exceeds the critical value, thereby falling into the critical region, we are in a position to reject the null hypothesis.

At the .05 level of significance, our critical value is ± 1.96. Thus, Z values of ± 1.96 are the values that begin the critical regions at the .05 level of significance. Any Z value (or Z ratio) that we calculate that meets or exceeds ± 1.96 is a value that falls within the critical region or the region of rejection (when working at the .05 level of significance). Similarly, ± 2.58 are the values that begin the critical regions when working at the .01 level of significance. A calculated test statistic (in this case, a Z ratio) that is equal to or more extreme than ± 2.58 would fall within the critical region at the .01 level of significance. *If our calculated value—our calculated test statistic—falls within the critical region, we reject the null hypothesis. If the calculated test statistic doesn't fall within the critical region, we fail to reject the null hypothesis.*

When we reject the null hypothesis, we're in a position to say that we have a *significant* finding—that's what the level of significance is all about. Another way of saying we have significant findings is to say that we have *statistically significant results*. In short, a statement that the results are significant is a statement that the calculated value of the test statistic falls within the critical region. In our flextime example, working at the .05 level of significance, we calculated a test statistic that fell within the critical region. Therefore, we rejected the null hypothesis; we had significant findings. We had reason to believe that the average productivity level of flextime workers really is different from the historical level of productivity.

But What If . . .

As we get ready to take this next step, let me urge you to first take a moment or two to think about what we've just covered. Take a moment to think about how we looked at a difference and evaluated that difference—ultimately coming to the conclusion that the difference was statistically significant at the .05 level of significance. The underlying logic and reasoning that we went through is fundamental to statistical analysis, so any time spent thinking about all of that will be well worth it.

Assuming you've reached a comfort level with the underlying reasoning, let me ask you to consider two more scenarios. Each scenario involves the same problem as before, but with some slight modifications:

> Scenario A: What if, for example, everything about the problem stayed the same (i.e., the same population mean and standard deviation, and the same number of cases in the sample), but the sample mean had been 184.51? Assuming we had worked at the .05 level of significance, what would we have concluded?
>
> Scenario B: What if the sample mean had been 199.53? Assuming we had worked at the .05 level of significance, what would we have concluded?

Let me suggest that you work through the problems presented in each scenario and spend some time thinking about the results.

Assuming you worked through the two problems represented by Scenarios A and B, you likely learned two very valuable lessons. Scenario A, for example, demonstrates that it's possible to end up with a negative Z value. Scenario B demonstrates that it's possible to end up with a difference between a sample and population mean that isn't significant (something we touched on earlier). Since those two scenarios expand the playing field in a noticeable way, let's take a closer look at each of them. As before, the assumption is made that you took the time to work through each of the scenarios.

As to a negative Z value under Scenario A, think about how you approached the difference between the two mean values. Assuming you used the formula presented earlier, you subtracted the mean of the population (the assumed mean of the sampling distribution of sample means) *from* the mean of the sample (see below).

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{x}}}$$

Since the mean of the sample in Scenario A was smaller than the mean of the population, the result was a negative difference (184.51 – 193.80 = –9.29). Carrying the negative value through the remainder of the calculations in the formula, you ended up with a negative Z value. As it turns out, the notion of a negative Z value should make a certain amount of intuitive sense. After all, you're considering a negative difference—negative in the sense that the mean of the sample is lower than the assumed mean. You would still reject the null hypothesis since the Z value of –2.08 is more extreme than the critical value of –1.96. It's just that you'd be looking at a significant difference in which the productivity level of the flextime workers was *lower* than the historical level of productivity.

As to the particulars of Scenario B, you were working with a sample mean of 199.53. As it turned out, the difference between the sample mean and the assumed mean, when expressed in terms of standard error units, wasn't that noticeable. Indeed, it equated to a Z value of 1.28. Since that Z value (1.28) didn't meet or exceed the critical value ± 1.96, you would fail to reject the null hypothesis. In other words, the sample mean was close enough to the assumed mean (at least in terms of standard error units), that it was reasonable to conclude that you could have obtained such a sample mean, just by chance. Therefore, you'd fail to reject the null hypothesis. You'd be inclined to say that there isn't a significant difference between the productivity levels of flextime workers and the historical level of productivity.

Now, just to recap the possibilities, we dealt with three different situations. In the first instance (the original problem), we found a significant difference—a difference in which the mean of the sample was higher than the historical mean of the population. In the second instance (Scenario A), we also found a significant difference—but it was a negative difference in the sense that the mean of the sample was lower than the historical mean of the population. In the third instance (Scenario B), the difference between the sample mean and the assumed mean was not significant.

But What If We're Wrong?

Now let's think about outcomes in a more general sense. Let's put aside the notion of a positive versus a negative difference and, instead, just think about two possible outcomes. One possible outcome was that we found a significant difference. In other words, we went through all of the calculations and the result was a calculated test statistic that met or exceeded the critical value. The other possibility, of course, was that we failed to find a significant difference. In other words, we went through all of the calculations and the result was a calculated test statistic that didn't meet or exceed the critical value.

As you now know, each of those outcomes would lead to a different conclusion, as follows:

> If we found a significant difference, we rejected the null hypothesis.
>
> If we didn't find a significant difference, we failed to reject the null hypothesis.

All of that is well and good, and the logic of hypothesis testing would be fairly easy to comprehend if that's all we had to consider—find a significant difference and reject the null or don't find a significant difference and fail to reject the null. There's just one problem. Regardless of the conclusion that we reached (i.e., we either rejected the null hypothesis or we failed to reject the null), we came to our conclusion on the basis of results obtained from only one sample. And it's the fact that we're relying on just one sample that's at the root of the problem, so to speak. To put the matter in the simplest of terms, it's quite possible that a different sample would have yielded different results.

As it turns out, that fact—the fact that different samples would likely yield different results—takes us another step down the road of statistical reasoning.

Type I Errors. Let's start with the assumption that the null hypothesis is true. In other words, the mean of the population of flextime workers is equal to the historical mean of all workers. Since this is fundamental to the notion of hypothesis testing, let me repeat the assumption: Assume that the null hypothesis is true. Don't ask how we know that the null hypothesis of no difference is true. Just assume that the null hypothesis is, in fact, true.

Now even if the null hypothesis is true, it's possible to obtain a sample mean that is extreme. For example, in selecting the sample of 50 productivity reports, we could, just by chance, select reports of the 50 most productive flextime workers. In a case like that, we'd eventually end up with a very high mean level of productivity for our sample of flextime workers. By the same token, we could, just by chance, select productivity reports for the 50 least productive flextime workers. In a case like that, we'd eventually end up with a very low mean level of productivity.

Now the likelihood of selecting a sample along those lines is fairly remote, but it could happen. And in either case, the point would be the same. We'd end up with an extreme mean as the result of what I like to call a *quirky* sample—a sample that doesn't really represent the population. What's more, that quirky sample would, in turn, lead to a sample mean that was quirky. Assuming that the sample mean was noticeably different from the assumed mean, we'd probably reject the null hypothesis. Unfortunately, we would have made an error.

All of that can be a confusing, so think it through again. Start with the assumption that the null is true. Then open your mind to the possibility that you could, just by chance, end up with a quirky sample. Then imagine a situation in which the quirky sample produced an extreme mean—in fact, a sample mean that was so extreme that you rejected the null hypothesis. Now think back to the fact that the null was true. You've obviously made a mistake. You've rejected the null when it was true.

To a statistician, this type of error is known as a **Type I error** (sometimes referred to as an *alpha error*). In short, a Type I error is a rejection of the null hypothesis when the null is, in fact, true. Unfortunately, we never know when we've made a Type I error. It all goes back to the notion of random sampling and the possibility of sampling error—the possibility that, just by chance, we were working with a sample that didn't really represent the population.

> **LEARNING CHECK**
>
> *Question:* What is the definition of a Type I error?
> *Answer:* A Type I error is the rejection of the null when, in fact, it is true.

Fine, you say. But what's the probability of making a Type I error? As it turns out, that's what the **level of significance** is all about. It's simply an expression of the probability of making a Type I or alpha error. As a matter of fact, that's why the level of significance is often referred to as the alpha level.

If you set your level of significance at .05 (often expressed as $\alpha = .05$), it's simply a statement that you're willing to tolerate a 5% chance of making a Type I error (rejecting the null when it's true). Similarly, the selection of the .01 level of significance when you set up a hypothesis test is a statement that you're willing to tolerate a 1% chance of committing a Type I or alpha error.

LEARNING CHECK

Question: What is the probability of making of a Type I error?
Answer: The probability of making a Type I error is equal to the level of significance (the alpha level).

Once again, a Type I error really derives from having selected a sample that, just by chance, doesn't really reflect reality. It results from bringing an extreme mean into the equation by accident. We know this can happen as a result of sampling error. Remember: We're always working with just one out of an infinite number of samples. When working at the .05 level of significance, for example, we can expect, just by chance, that 5 samples out of 100 would ultimately result in a rejection of the null, even though the null is true. In other words, we could go through the research exercise 100 times, each time selecting our sample and each time calculating a test statistic and each time arriving at a conclusion. In 5 of the 100 instances, though, we could end up with statistically significant results as a result of sampling error. *If* we're working with one of those samples is something we can never know. All we know is the probability that we've been working with one of those samples. All we know is the probability that we've committed a Type I or alpha error. Remember: That's what the level of significance is all about. It's the probability that we've committed a Type I error.

As we wrap up our discussion of Type I errors, let me give you a way to express your conclusion whenever you reject a null hypothesis, at least at the outset of your statistical education. It's a little more elaborate than the simple assertion *I reject the null hypothesis,* but the extra few words are important—at least in terms of helping you understand the fundamental logic involved in a hypothesis test.

Let's say, for example, that you're working at the .05 level of significance and you've found significant results. Consider the following as a way of expressing your conclusion: *I reject the null hypothesis, with the knowledge that 5 times out of 100 I could have committed a Type I error.*

I doubt that you'll see a statement like that in a scientific journal, to be sure, but rest assured of one thing—it's a statement that reflects the fundamental logic of what's involved in hypothesis testing. Remember: There's always a chance that you're working with an extreme sample (one of those quirky samples, as I like to call them). There's always a chance that you've rejected the null when, in fact, it's true.

My suggestion is that you get in the habit of phrasing your conclusion in a complete fashion at the outset of your statistical experience, at least when you find yourself in the position of rejecting a null hypothesis. Whenever you reject a null hypothesis, remind yourself that you're rejecting the null with the knowledge that there's a known probability of making a Type I error. It's the sort of thing that will constantly remind you of the logical underpinnings of the decision-making process.

So much for Type I errors. As you might suspect, there's also something known as a **Type II (or beta) error**. That's what we'll cover now.

Type II Errors. To understand what a Type II error is all about, start with the assumption that our null hypothesis is actually a false hypothesis. Returning to the flextime worker example, let's assume that the null hypothesis is false—the mean level of productivity for flextime workers *is* significantly different from the historical mean level of productivity, even though the null says that they are equal. Now even though there may be a difference (in other words, even though the null may be false) it's possible to select a sample that doesn't pick up that difference. Once again, that's something that could happen just by chance.

If we're dealing with one of those instances—an instance in which there's a significant difference, but we've failed to detect it—we've committed a Type II error. Simply put, a Type II error occurs when we fail to reject a false null hypothesis. Here's the logic again. The null hypothesis is false (i.e., there is a difference), but we didn't discover the difference. Because we didn't discover the difference, we failed to reject the null. The result is that we let the null stand. We failed to reject it. In truth, however, the null was false, and we should have rejected it. If we've failed to reject a false null hypothesis, we've committed a Type II error.

LEARNING CHECK

Question: What is the definition of a Type II error?
Answer: It's the failure to reject the null when it is false.

At this point in your statistical education, it's not critical that you worry about how a Type II error could occur or how to determine the probability of making a Type II error. We'll deal with Type II errors in greater depth when we reach Chapter 9. At that point, we'll take a closer look at both Type I and Type II errors, and we'll also consider Type II errors in the context of power and effect size—two concepts that are particularly relevant in experimental research designs. We'll also take up the topic of alternative or research hypotheses. All of that, however, can wait. Right now the idea is to solidify your thinking on what we've just covered.

A Final Word About Phrasing Your Conclusions. As we close out this section, let me once again urge you to be very specific in how you should phrase your conclusions when testing hypotheses. After all, it makes little sense to work through a research problem, getting all of the calculations just right, only to make a mistake when state your conclusions.

If you reject a null hypothesis, I urge you to phrase your conclusion with some reference to the level of significance. For example, (and assuming you're working at the .05 level of significance), simply state that you reject the null hypothesis, with the knowledge that there is a 5% chance of having committed a Type I error. If, on the other hand, you fail to reject the null hypothesis, all you have to do is simply state that you fail to reject the null hypothesis. You don't have to state anything else. Just end the sentence right there—*I fail to reject the null hypothesis*.

On this last point (i.e., use of the phrase, *I fail to reject the null hypothesis*), I find that students often ask why they can't simply *accept* the null hypothesis. It may be my conservative nature, but here's the way I see it. A simple acceptance of the null hypothesis always strikes me as closing the door on further research. It is as though you have accepted the null, announced that the case is closed, and that is that. In saying that you *fail to reject the null,* however, it is as though you've left the door open for further research, so to speak. In that regard, I'm often reminded of the words of Popper:

> The game of science is, in principle, without end. He who decides one day that scientific statements do not call for any further test, and that they can be regarded as finally verified, retires from the game. (Popper, 1961, p. 53)

At this point, I suggest that you take a moment or two to catch your breath. Rather than plowing ahead, let me suggest that you spend a little time going over the material that we've covered so far. Concentrate on the fundamental logic first. Get familiar with all the central concepts. Once you've done that, take a step back and consider how far you've just traveled. You've learned how to conduct a hypothesis test using a single sample mean when the population standard deviation σ is known. You've also learned what it means to reject or fail to reject a null hypothesis. If you're comfortable with all of that, we can move forward to hypothesis tests using a single sample mean when the standard deviation of the population σ is unknown.

Single Sample Test With σ Unknown

Let's say that the average number of cases processed last month by social workers throughout the state is reported as 23.12 per worker. As members of the agency's management team in a regional office, we want to know how the caseworkers in the regional office compare to those throughout the state. Let's assume that we want to save time and effort, so we identify 30 caseworkers in our region and examine their work records from the previous month. The examination reveals that the workers in our sample processed a mean ($\overline{X}$) of 24.74 cases, with a standard deviation for the sample (s) equal to 4.16.

At this point, you should recognize the problem as one that's remarkably similar to those we covered in the previous section, with one minor exception. In the present example, we don't know the value of the standard deviation of the population (σ). Instead, what we have is the standard deviation for our sample (s). Nonetheless, the structure of the problem is identical to what we covered earlier. We want to determine if there is a significant difference between productivity of caseworkers throughout the state and that of the caseworkers in our region. As we move ahead with the problem, let's assume that we've decided to work at the .05 level of significance.

To deal with the fact that we don't know the value of σ, we'll make two minor changes. First, we'll rely on the family of t distributions to obtain our critical value (instead of relying on Z values). Second, we'll have to estimate the standard error of the mean (instead of calculating it in a direct fashion). By now you should recognize those as the same changes we made when we moved from the construction of a confidence interval for the mean with σ known to the construction of a confidence interval for the mean when σ is unknown.

LEARNING CHECK

Question: In a single sample hypothesis test, what is the difference in the procedures when sigma is unknown as opposed to when sigma is known?

Answer: When sigma is unknown, the standard error of the mean is estimated and t is used. When sigma is known, the standard error of the mean is calculated in a direct fashion and Z is used.

In the material about confidence intervals, you learned quite a bit about the family of t distributions. In the same context, you also learned how to estimate the standard error of the mean. Just to refresh your memory, here's the formula we use to estimate the standard error of the mean.

$$s_{\overline{X}} = \frac{s}{\sqrt{n}}$$

Now we have everything we need to tackle the problem.

Applying the Test

The formula we'll use in this case is the same as the one we used in our previous (flextime) example, with the exception of the two minor changes

I mentioned earlier. You'll note that the formula now reflects a t value (as opposed to Z), and we use the estimate (instead of the calculated value) of the standard error of the mean.

Calculation. For the sake of clarity, let's start with the formula for estimating standard error of the mean.

$$s_{\overline{X}} = \frac{s}{\sqrt{n}}$$
$$s_{\overline{X}} = \frac{4.16}{\sqrt{30}}$$
$$s_{\overline{X}} = \frac{4.16}{5.48}$$
$$s_{\overline{X}} = 0.76$$

Now we use the formula for calculating the test statistic. Note that now we are calculating t instead of Z.

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}}$$
$$t = \frac{24.74 - 23.12}{0.76}$$
$$t = \frac{1.62}{0.76}$$
$$t = 2.13$$

As you learned in the last chapter, we'll have to take note of the relevant degrees of freedom in using the family of t distributions. As before, our degrees of freedom will be calculated as $n - 1$. Once we determine the appropriate number of degrees of freedom (in this case, $30 - 1 = 29$), we take note of the corresponding row in Appendix B. Since we're not constructing a confidence interval, we don't have to go through any mental conversion (1 minus the level of significance) to locate the appropriate column. Instead, we can focus directly on the appropriate column—in this case, the one labeled .05 level of significance. Once we find the intersection of the appropriate degrees of freedom row (29) and the appropriate level of significance column (.05), we take note of the t value of 2.045, which we can round to 2.05. That t value—2.05—becomes our critical value. As before, it is this value that our calculated t value (or t ratio) must meet or exceed if we're to reject the null hypothesis. We now have all the information we need to come to a conclusion and interpretation.

Interpretation. With a calculated test statistic at hand (a t ratio of 2.13), and knowledge of the critical value at the .05 level of significance (2.05), we're in a position to formulate a conclusion and interpretation. Since our calculated *t* value exceeds the critical value, we're in a position to reject the null hypothesis. We can reject the null hypothesis with the knowledge that there is a 5% chance or probability of having made a Type I error.

Though we never know whether or not we actually committed a Type I, or alpha, error, we always know the probability of having done so. In this case, we know it is .05, or 5%. We know that because that's what the level of significance is all about—it is simply a statement of the probability of making a Type I error.

Some Variations on a Theme

Now let's consider some more examples. We can use the same situation as before, comparing the performance of the caseworkers in our region to the population of caseworkers throughout the state, but let's alter the specifics in various ways. This is one of the best ways to school yourself on the logic of hypothesis testing, from the point of setting up the hypothesis all the way through the point of conclusion and interpretation.

Let's assume that much of the problem remains the same: the same mean ($\overline{X}$ = 24.74), the same standard deviation of the sample (4.16 cases per month), and the same population mean (23.12 cases per month). As to the sample size (*n*), we'll vary that a bit. If the sample size changes, we'll have to change our estimate of the standard error of the mean, but that's easily dealt with. Finally, let's vary the level of significance. As I said before, all these variations, taken together, should give you a solid grounding in all that's involved in deriving a conclusion and interpretation of the hypothesis test.

> **Situation:** First let's assume that we're working with a sample of 16 caseworkers instead of 30, and that everything else about the problem remains the same. How would the conclusion change, if at all?
>
> **Commentary:** The change in the number of cases (*n*) results in two immediate changes. The number of degrees of freedom changes to 15 ($n - 1 = 15$), and the estimate of the standard error of the mean changes from 0.76 to 1.04. With the change in degrees of freedom, the critical value of *t* becomes 2.13. The larger estimate of the standard error of the mean results in a smaller calculated *t*—it is now 1.56. The calculated value of 1.56 does not equal or exceed the critical value of 2.13. Therefore, we would fail to reject the null hypothesis.
>
> **Situation:** Let's assume that we're working with a sample of 61 caseworkers instead of 30, and that we're working at the .01 level of significance. Everything else about the problem remains the same. How would the conclusion change, if at all?
>
> **Commentary:** As above, the change in the number of cases (61) results in two immediate changes. The number of degrees of freedom changes to 60 ($n - 1 = 60$), and the estimate of the standard error of the mean becomes 0.53. With degrees of freedom = 60 and our level of

significance set at .01, the critical value is 2.66. The smaller estimate of the standard error of the mean results in a larger calculated *t*—it is now 3.06. The calculated value of 3.06 exceeds the critical value of 2.66. Therefore, we would reject the null hypothesis at the .01 level of significance. We reject the null with the knowledge that there's a 1% chance or probability of having made a Type I error.

Situation: Let's assume that we're working with a sample of 20 caseworkers instead of 30, and that we're working at the .01 level of significance. Everything else about the problem remains the same. How would the conclusion change, if at all?

Commentary: As above, the change in the number of cases (20) results in two immediate changes. The number of degrees of freedom changes to 19 ($n - 1 = 19$), and the estimate of the standard error of the mean becomes 0.93. With degrees of freedom = 19 and our level of significance set at .01, the critical value is 2.86. The change in the estimate of the standard error of the mean results in a change of the final calculated *t*—it is now 1.74. The calculated value of 1.74 does not meet or exceed the critical value of 2.66. Therefore, we would fail to reject the null hypothesis.

And so it goes. Different answers and different conclusions—all dependent on a couple of factors (namely, the number of cases and the level of significance).

Chapter Summary

As we bring this chapter to a close, take a moment or two to reflect again on what we've covered. As I mentioned at the outset, the chapter is as much about hypothesis testing as it is about a particular application. Indeed, if you were to get just one thing out of the chapter, I would hope it would be that—the logic of hypothesis testing.

As to the application—a single sample test—by now you should have a fairly solid understanding about how it can be used. For example, you should now know what's involved when you want to test a hypothesis involving a single sample. You should know that the underlying logic remains the same, whether σ is known or unknown. Indeed, you should know by now that the variations in the two approaches trace squarely back to the differences you encountered when constructing confidence intervals for the mean (with σ known and σ unknown).

Perhaps most important, you should have digested the fundamentals of hypothesis testing logic along the way. The new concepts—null hypothesis, critical value and critical region, level of significance, and Type I and II errors—are concepts you will encounter over and over again. Specific statistical applications will change from situation to situation, but the fundamental logic will remain the same. As I'm fond of telling my students at this point in their statistical journey, the logic of hypothesis testing is a little bit like Mozart's music—or rap music, for that matter: It's the same darned thing over and over again.

Some Other Things You Should Know

Let me bring two matters to your attention at this point. One has to do with how the logic you just encountered extends to other types of research situations. The other has to do with a possible source of confusion that you may encounter down the road.

As to the first of these matters, let me mention that the logic you encountered in this chapter extends to hypotheses involving proportions as well as means. For example, a simple procedure will allow you to determine if a proportion (or percentage value) observed in a sample differs significantly from a known or assumed proportion in a population. The procedure is directly analogous to the research situations covered in this chapter, with one exception—the focus is on proportions, rather than means. For a discussion of this procedure, consult Utts and Heckard (2002).

As to the possible confusion that may arise down the road, let me ask you to focus on the fundamental difference between the material you encountered in this chapter and the material presented in the previous chapter. My experience tells me that students often confuse the material, probably because all of the material falls into the category of inferential statistics and all the material is tied to the use of Z or t. Let me offer a simple way out of this possible confusion.

As you think about the material we covered in Chapter 6, think about the purpose of the procedures that we explored. There, we constructed confidence intervals. We didn't test hypotheses; we didn't even formulate hypotheses. We simply constructed confidence intervals. Our goal was to make statements about population parameters, based on sample statistics. We were developing estimates.

In the present chapter, however, our goal was different. In this chapter, we set out to test hypotheses. Yes, we calculated Z and we calculated t. We used the Table of Areas Under the Normal Curve and we used the table for the distribution of t (Family of t Distributions). Those are the same elements that came into play when we were constructing confidence intervals. But there was a major difference, and it had to do with our ultimate purpose. Remember: We construct a confidence interval because we want to estimate the value of a population parameter. Testing a hypothesis represents a very different goal.

Key Terms

calculated test statistic
critical region
critical value
hypothesis
level of significance
null hypothesis
region of rejection
Type I error
Type II error

Chapter Problems

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

General Thought Questions

1. The level of significance is the probability of making a(n) ________ error.
2. Rejecting a null hypothesis when it is true represents a(n) ________ error.
3. Failing to reject a null hypothesis when it is false represents a(n) ________ error.
4. The ________________ is represented by the symbol H_0.
5. The alpha level is also known as the level of ____________.
6. Another name for the region of rejection is the ________________.
7. If your calculated test statistic does not meet or exceed the critical value, you would ________ the null hypothesis.
8. The levels of significance most commonly used by statisticians are the _____ and ____ levels.
9. When our calculated test statistic falls within the ________, we reject the ______________.

Application Questions/Problems: Hypothesis Test Based on a Single Sample $\overline{X}$ With σ Known

1. The police department of a major city reports that the mean number of auto thefts per neighborhood per year is (μ) 6.88 with a standard deviation (σ) of 1.19. As the mayor of a suburban community just outside the major city, you're curious as to how the auto theft rate in your community compares. You determine that the mean ($\overline{X}$) number of auto thefts per neighborhood per year for a random sample of 15 neighborhoods in your community is 8.13. Assume that you're working at the .05 level of significance.
   a. State an appropriate null hypothesis.
   b. What is the value of the calculated test statistic (Z)?
   c. State your conclusion.
2. Reports indicate that graduating seniors in a local high school have an average (μ) reading comprehension score of 72.55 with a standard deviation (σ) of 12.62. As an instructor in a GED program that provides alternative educational opportunities for students, you're curious how seniors in your program compare. Selecting a sample of 25 students from your program and administering the same reading comprehension test, you discover a sample mean ($\overline{X}$) of 79.53. Assume that you're working at the .05 level of significance.

a. State an appropriate null hypothesis.
b. What is the value of the calculated test statistic (Z)?
c. State your conclusion.

3. Students participating in a drug education program are given a drug awareness test at the beginning of the program. The mean score (μ) for the population of 526 students is 61, with a standard deviation (σ) of 12. As program director, you're curious as to how parents of the students perform on the drug awareness test and whether or not they are significantly different from the students. Selecting a random sample of 50 parents, you administer the test and discover a mean drug awareness score ($\overline{X}$) of 56. Assume that you're working at the .05 level of significance.
a. State an appropriate null hypothesis.
b. What is the value of the calculated test statistic (Z)?
c. State your conclusion.

4. The mean (μ) educational level for adults in a community is reported as 10.45 years of school completed with a standard deviation (σ) of 3.8. Responses to a questionnaire by a sample of 40 adult community residents indicate a mean educational level ($\overline{X}$) of 11.45. Assume that you're working at the .05 level of significance.
a. State an appropriate null hypothesis.

b. What is the value of the calculated test statistic (Z)?
c. State your conclusion.

5. The historical mean level of production workers at an industrial plant is shown to be 155 units produced per day, with a standard deviation of 15. Following the introduction of a new flextime worker option, a sample of productivity reports for 100 flextime workers reveals a sample mean ($\overline{X}$) of 160. Assume that you're working at the .05 level of significance.
a. State an appropriate null hypothesis.
b. What is the value of the calculated test statistic (Z)?
c. State your conclusion.

6. A standardized test, designed to measure the mathematical skill level of seventh graders is said to have a mean score (μ) = 75 with a standard deviation (σ) = 10. As the principal of a private school, you're curious how seventh graders, in your school compare. Selecting a sample of 25 students from your school and administering the mathematical skill test, you discover a sample mean ($\overline{X}$) of 79. Assume that you're working at the .05 level of significance.
a. State an appropriate null hypothesis.
b. What is the value of the calculated test statistic (Z)?
c. State your conclusion.

Application Questions/Problems: Hypothesis Test Based on a Single Sample Mean With σ Unknown

1. The mean (μ) level of absenteeism rate for the local school district is reported as 8.45 days per year, per student. The mean rate ($\overline{X}$) for a sample of 30 students enrolled in a vocational training program is reported as 6.79 days per year with a standard deviation (s) of 2.56 days. Assume that you're working at the .05 level of significance.

- **a.** State an appropriate null hypothesis.
- **b.** What is the value of the calculated test statistic (t)?
- **c.** Identify the critical value.
- **d.** State your conclusion.

2. The national mean (μ) absentee rate for workers working for the Old Mill Store Company is reported as 8.25 days per year. The mean rate ($\overline{X}$) for a sample of 14 workers working at your Old Mill Store franchise is reported as 7.53 days per year with a standard deviation (s) of 2.72 days. Assume that you are working at the .05 level of significance.

- **a.** State an appropriate null hypothesis.
- **b.** What is the value of the calculated test statistic (t)?
- **c.** Identify the critical value.
- **d.** State your conclusion.

3. Information collected at a local university indicates that students are working, on average (a value for μ) 15.23 hours per week while in school. Information collected from a random sample of 25 fraternity members, however, reveals a mean of 12.34 hours per week with a standard deviation (s) of 2.50 hours. Assume that you're working at the .05 level of significance.

- **a.** State an appropriate null hypothesis.
- **b.** What is the value of the calculated test statistic (t)?
- **c.** Identify the critical value.
- **d.** State your conclusion.

4. Information collected at a local university indicates that business majors enroll for an average (μ) of 10.65 credit hours per semester. As the Dean of the School of Communication, you are interested in how journalism majors compare. Taking a sample of enrollment records on 31 journalism majors, you find a mean ($\overline{X}$) credit hour enrollment of 12.22 hours, with a standard deviation (s) of 3.26. Assume that you're working at the .05 level of significance.

- **a.** State an appropriate null hypothesis.
- **b.** What is the value of the calculated test statistic (t)?
- **c.** Identify the critical value.
- **d.** State your conclusion.

5. A recent news report indicates that the mean (μ) number of years that first-time drug offenders are sentenced is 12.16. A sample of 25 court records from your county indicates a mean number of years = 11.24, with a standard deviation of 3.11. Assume that you're working at the .05 level of significance.
   - **a.** State an appropriate null hypothesis.
   - **b.** What is the value of the calculated test statistic (t)?
   - **c.** Identify the critical value.
   - **d.** State your conclusion.
6. A recent news report asserts that the weekly mean (μ) number of drinks (both mixed drinks and beer) consumed by college students is 12.56 drinks. Data from a sample of 30 students enrolled at your university indicate a weekly consumption level ($\overline{X}$) of 11.21 drinks, with a standard deviation (s) of 3.88. Assume that you're working at the .05 level of significance.
   - **a.** State an appropriate null hypothesis.
   - **b.** What is the value of the calculated test statistic (t)?
   - **c.** Identify the critical value.
   - **d.** State your conclusion.