PART II One- and Two-Sample Difference Tests

The four chapters in this section form the core of statistics for psychology. First, hypothesis testing is introduced. Hypothesis testing is the procedure that statisticians use to make decisions that are objective and data-based, not subjective and emotion-based.

Then the most commonly used tool in the psychologist's toolbox, the difference test, is introduced in several varieties. Single-sample difference tests are used to answer questions such as the following: Is the average GPA in this *sample* of students different from the mean GPA for the entire *population* of students? Two-sample difference tests are used to answer questions like those posed by classic experiments that have one experimental group and one control group. In this section, we'll learn about two, the independent-samples *t* test and the paired-samples *t* test. In the next section, we'll cover ANOVA, tests used in more complex situations when experiments include three or more groups.

Introduction to Hypothesis Testing

6

LEARNING OBJECTIVES

- Explain how hypothesis testing works.
- List the six steps to be followed in completing a hypothesis test.
- Explain and complete a single-sample *z* test.
- Explain the decisions that can be made in hypothesis testing.

CHAPTER OVERVIEW

This chapter makes the transition from descriptive statistics to inferential statistics. In inferential statistics, *specific* information from a sample is used to draw a *general* conclusion about a population. For example, this allows a researcher to study one group of people with depression and to draw a conclusion that applies to people with depression in general. This transition to inferential statistics starts with the logic that statisticians use to reach decisions, a process called hypothesis testing.

After this prelude, the chapter takes a pragmatic turn with coverage of the single-sample *z* test, which we'll use to demonstrate how hypothesis testing works. Psychology researchers rely on statistics to help them make thoughtful, data-driven decisions. The single-sample *z* test allows researchers to determine if a sample mean is statistically significantly different from a population mean or some other specified value.

Completing a hypothesis test is a lot like baking. Before you can put the cake in the oven, you must go to the store to buy the ingredients, turn on the oven, measure out the ingredients, and mix them together in the right order. This chapter introduces a six-step procedure that can be used for all hypothesis tests to make sure that the right steps happen in the right order. We even offer a mnemonic—Tom and Harry despise crabby infants—to keep everything straight.

The chapter ends with an exploration of the different types of correct and incorrect conclusions that can occur. The incorrect conclusions have names: Type I error and Type II error. For a single-sample *z* test, a Type I error occurs when one erroneously concludes that a sample mean differs from a population mean. Type II error is the opposite—it occurs when the researcher indicates there is no evidence that the sample mean differs from the population mean and the sample really does differ.

6.1 The Logic of Hypothesis Testing

A **hypothesis** is a proposed explanation for observed facts. If, for example, a psychologist noted that people living in sunny climates were happy and people living in cloudy climates were sad, this might lead to the hypothesis that a lack of sun leads to depression.

Science involves gathering observations, generating hypotheses to explain them, and then testing the hypotheses to see if the predictions they make hold true. In statistics, **hypothesis testing** is the procedure in which data from a *sample* are used to evaluate a hypothesis about a *population*.

Here's how hypothesis testing works. A cognitive psychologist might hypothesize that the mean IQ in a population is 100. If she gathered a representative sample of people from that population and found that the mean IQ of her sample was 100 or close to 100, there would be little reason to question the hypothesis. However, if the observed mean in the sample were far from the expected mean of 100, then she would have to wonder if the population mean were really 100.

Here's another example that shows how humans already use hypothesis testing in everyday life. Amie has a boyfriend, Allen, and she believes (hypothesizes) that Allen loves her. Based on this hypothesis, she expects that Allen will behave in certain ways. If he loves her, then he should want to spend time with her, do nice things for her, hold her hand, and so on. And if Allen does these things, then there's no reason for Amie to question his love for her. But, if Allen doesn't do these things—if the observed behavior doesn't match the expected behavior—then she'll to start wondering if he truly loves her.

Ten Facts About Hypotheses and Hypothesis Testing

The love example shows how humans use hypothesis testing intuitively. Now, let's see how scientists have formalized the process with 10 facts about hypothesis testing (Table 6.1):

1. A hypothesis is a statement about a *population*, not a sample. This is because a researcher wants to draw a conclusion about the larger population and not about the specific sample being studied.
2. There will be two hypotheses—one called the null hypothesis, H_0, and the other called the alternative hypothesis, H_1.
3. The two hypotheses must be *all-inclusive,* which means they cover all possible outcomes. For example, if one hypothesis said that all cars in the world were white and the other that all cars were black, those two hypotheses wouldn't be all-inclusive because red and green and yellow cars also exist. However, if one hypothesis stated, "All cars are white" and the other hypothesis, "Not all cars are white," then the two hypotheses would be all-inclusive.
4. The two hypotheses must be *mutually exclusive,* which means that only one hypothesis at a time can be true. For example, a coin can land on heads or on tails, but not on both sides at once. Heads and tails are mutually exclusive outcomes.
5. The null hypothesis is a negative statement. The **null hypothesis** says that in the population the explanatory variable does *not* have an impact on the outcome variable. For example, if a researcher were testing a technique to improve

TABLE 6.1 Ten Facts About Hypotheses and Hypothesis Testing

1.	A hypothesis is a statement about a *population,* not a sample.
2.	In hypothesis testing, there are two hypotheses, the null hypothesis (H_0) and the alternative hypothesis (H_1).
3.	The two hypotheses must be all-inclusive.
4.	The two hypotheses must be mutually exclusive.
5.	The null hypothesis is a negative statement and makes a specific prediction.
6.	The alternative hypothesis, which is what the researcher believes is true, isn't a specific prediction.
7.	One can't prove a negative statement.
8.	But, one can *disprove* a negative statement. And all it takes is one example.
9.	If the null hypothesis is disproved (rejected), then the researcher has to accept the mutually exclusive alternative hypothesis.
10.	If the researcher fails to disprove the null hypothesis, he or she can't say that the null hypothesis is true. The best that can be said is that one hasn't found enough evidence to reject it.

intelligence, the null hypothesis would state that the technique does *not* improve intelligence. Because the null hypothesis makes a specific prediction, it would say that the technique has zero impact and does not improve intelligence *at all.*

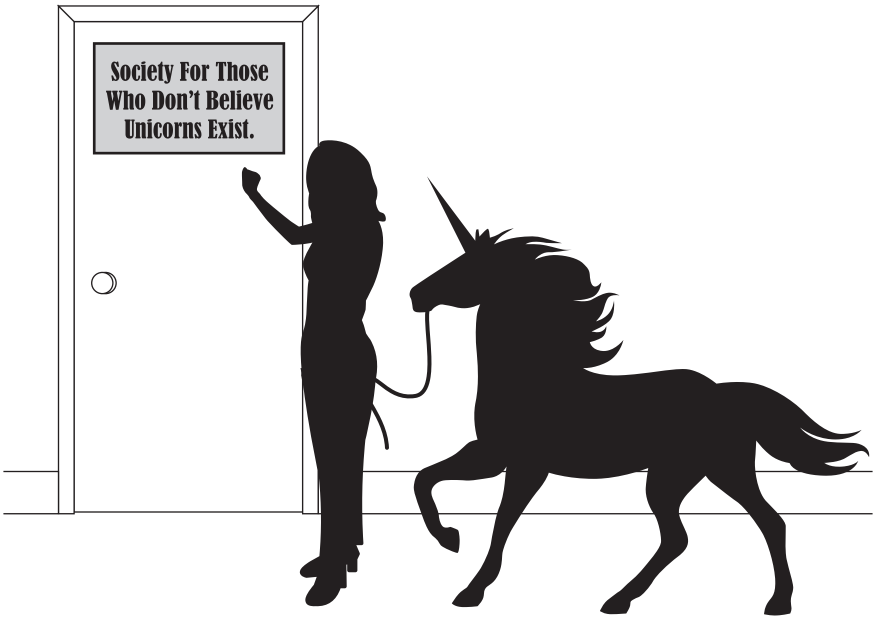
6. The **alternative hypothesis** is what the researcher believes is true; it states that, in the population, the explanatory variable has an impact on the outcome variable. For the researcher testing the intelligence-improving technique, the alternative hypothesis would state that the technique has *some* impact on intelligence. Notice that the alternative hypothesis doesn't specify how much impact. Unlike the null hypothesis, the alternative hypothesis doesn't make a specific prediction.

7. A negative statement, like the null hypothesis, can't be proven true. Suppose an adult wants to prove the negative statement "There are no unicorns" to a child who believes in unicorns. No matter what evidence the adult offers (e.g., "I looked all through Ohio and didn't find one"), the child will counter that the adult wasn't looking in the right place, or that the unicorns heard the adult coming and fled. A negative can't be proven.

8. However, a negative can be disproven. It just takes one example. If the child ever walks up leading a unicorn on a leash, the adult can no longer claim, "There are no unicorns." Because the null hypothesis makes a specific prediction, it is used to predict how a researcher's experiment will turn out. If the experiment doesn't turn out as the null hypothesis predicted, the researcher disproves, or "rejects," the null hypothesis. The null hypothesis can be nullified.

9. If the null hypothesis is rejected, then one has to accept the mutually exclusive alternative hypothesis as true. This happens because the null hypothesis and the alternative hypothesis are all-inclusive and mutually exclusive. If one hypothesis is not true, then the other hypothesis has to be true. As the researcher usually believes that the alternative hypothesis is true, the objective of a study is almost always to reject the null hypothesis.

10. However, if a researcher fails to disprove the null hypothesis, he or she can't say that the null hypothesis is true. Remember, a negative statement, like the null hypothesis, can't be proven true. The best the researcher can say is that he or she hasn't found enough evidence to reject the null hypothesis. This is different from saying the null hypothesis is true, just as a not guilty verdict in a trial is different from saying that the defendant is innocent. Not guilty just means that there wasn't enough evidence to persuade the jury that the defendant was guilty.



It is impossible to prove a negative hypothesis, but one positive example will disprove it.

Practice Problems 6.1

Review Your Knowledge

6.01 How do the null hypothesis and the alternative hypothesis differ?

Apply Your Knowledge

6.02 Explain this statement: "Hypothesis testing involves comparing what is observed to happen in an experiment to what is expected to happen."

6.2 Hypothesis Testing in Action

Let's put the 10 facts about hypothesis testing to work. Let's suppose a psychologist, Dr. Pell, wonders whether children who are adopted differ in intelligence from non-adopted children. To explore this, she obtains a random sample of 72 children from the population of adopted children in the United States and gives each of them an IQ test. She finds that the mean IQ of these 72 children is 104.5. Knowing that the average IQ for children in the United States is 100 with a standard deviation of 15, can she conclude that adopted kids differ in intelligence from the general population?

It is tempting to say that the answer is obvious. 104.5 is a bigger number than 100, so it is true that adopted kids differ from the general population of kids. But, that's not how statisticians think.

Remember sampling error, introduced in Chapter 5? Because of sampling error, researchers don't expect a sample mean to be exactly the same as the population mean. With a population mean for IQ of 100, Dr. Pell can expect the sample mean to be close to 100, but not exactly 100.

So, here's the question for Dr. Pell to ask: Is 104.5 close enough to 100 that sampling error can explain it? If so, then there's no reason to think adopted kids differ from the average IQ. However, if sampling error fails as a reasonable explanation for the difference, then she can conclude that adopted kids differ in intelligence.

The Six Steps of Hypothesis Testing

Learning how to complete a statistical test is like learning how to cook—it's better to follow a recipe. That's why we're going to use a six-step "recipe" for hypothesis testing. Here are the six steps:

Step 1 Test: Pick the right statistical test.

Step 2 Assumptions: Check the assumptions to make sure it is OK to do the test.

Step 3 Hypotheses: List the null and alternative hypotheses.

Step 4 Decision rule: Find the critical value of the statistic that determines when to reject the null hypothesis.

Step 5 Calculation: Calculate the value of the test statistic.

Step 6 Interpretation: State in plain language what the results mean.

Here's a mnemonic to help remember the six steps in order: "Tom and Harry despise crabby infants." The first letters of the six words in the mnemonic stand for "**T**est," "**A**ssumptions," "**H**ypotheses," "**D**ecision rule," "**C**alculation," and "**I**nterpretation." First, let's walk through the steps, then follow Dr. Pell as she applies them to the adoption IQ study.

A mnemonic to remember the six steps of hypothesis testing (**t**est, **a**ssumptions, **h**ypotheses, **d**ecision rule, **c**alculation, **i**nterpretation): "Tom and Harry despise crabby infants."

Picking the right test requires thought.

Step 1 Pick a Test

The first step in hypothesis testing is picking which test to use. There are hundreds of statistical tests, each designed for a specific purpose. Choosing the correct test depends on a variety of factors such as the question being asked, the type of study being done, and the level of measurement of the data. In Chapter 16, there are flowcharts that guide one through choosing the correct hypothesis test. Feel free to peek ahead and take a look at them.

Step 2 Check the Assumptions

All statistical tests have assumptions, conditions that need to be met before a test is completed. If the assumptions aren't met, then researchers can't be sure what the results of the test mean.

Here's a nonstatistics explanation about the role of assumptions. In this day and age, athletes are tested for performance-enhancing substances. The drug test depends on a number of assumptions. For the test to be meaningful, it is assumed that the sample being tested is the athlete's, that the sample was stored at the right temperature after it was taken, that no one has tampered with the sample, and that the machine being used to test it is correctly calibrated.

Imagine that an athlete wins a race, provides a urine sample, and it is tested on an incorrectly calibrated machine. The assumption that the machine is calibrated correctly has been violated. Suppose the results indicate that the urine does not test positive for performance-enhancing substances. Is that true? Maybe it is, maybe it isn't. When an assumption is violated, it is still physically possible to complete the test. But, one should not because it is impossible to interpret the results.

As hypothesis tests are covered, their assumptions will be listed to make sure one knows the conditions that must be met to proceed with the test. There are two types of assumptions: *not robust* and *robust.* A **nonrobust assumption** has to be met for the test to proceed. If a nonrobust assumption is violated, a researcher should stop proceeding with the planned statistical test. A **robust assumption** can be violated, to some degree, and the test can still be completed and interpreted.

Here is a way to think about the difference between the two assumptions. People can be described as being in "robust" health or not. Imagine that Carl is not in robust health—he has a compromised immune system. In fact, his health is so bad that he is in an isolation ward in the hospital. One day he gets a visitor who has the flu and who sneezes while in the room. Carl, whose health has just been violated, is likely to get the flu. In contrast, Bethany is in robust health. One day her roommate, who has the flu, sneezes directly in her face. Bethany, with a strong immune system, is able to fight off this violation and stay healthy. However, there is a limit to what Bethany's immune system can handle. If one roommate with the flu sneezes in her face, by mistake Bethany uses the toothbrush of another roommate with the flu, and it turns out that Bethany's significant other—whom she spends a lot of time kissing—also has the flu, well, that may be too many violations for her immune system to handle. Even a robust assumption will break if it is violated too much.

Step 3 List the Hypotheses

Step 3 involves listing the null hypothesis and the alternative hypothesis. Hypotheses can be for *two-tailed tests* (also called nondirectional tests) or *one-tailed*

Before one can proceed with a test, one needs to make sure that its assumptions have been met.

tests (directional tests). A **two-tailed hypothesis test** doesn't indicate whether the explanatory variable (adoption in our example) has a positive or negative impact on the outcome variable (IQ), just that it has an impact. An advantage of a two-tailed test is that it allows a researcher to test for an effect in either direction, that children who are adopted end up as more intelligent than average or as less intelligent than average. Two-tailed tests are the norm and are used much more frequently than one-tailed tests.

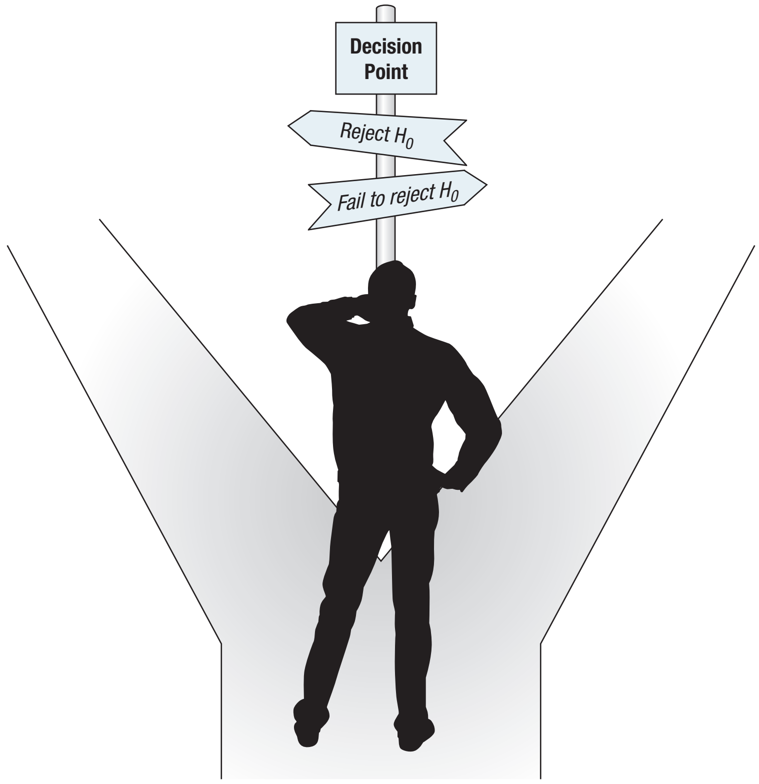
The null and alternative hypotheses need to be specified for every statistical test.

When, in advance of collecting any data, a researcher has an expectation about the direction of the impact of the explanatory variable, a **one-tailed hypothesis test** is called for. One-tailed tests predict that the results will turn out in a certain direction. An advantage to using one-tailed tests is that it is easier to reject the null hypothesis and be forced to accept the alternative hypothesis.

With the current example, it seems reasonable to think a one-tailed test is called for because it is already known that the mean IQ of this sample of adopted children is above average. But the question—do adopted children differ in intelligence from the average—was formulated before any data were collected. The question doesn't specify a direction for the difference, so a two-tailed test is called for. Changing the question after looking at the data is not how statisticians operate.

Step 4 Set the Decision Rule

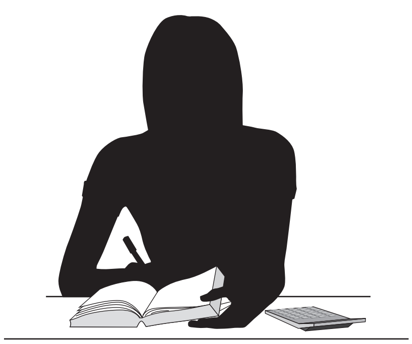
Setting the decision rule involves finding the *critical value* of the test statistic. The **critical value** is the value that the test statistic must meet or exceed in order to reject the null hypothesis. What the critical value is depends on a number of



Setting the decision rule in advance means that one knows what to do when a choice point is reached.

factors, such as how willing the researcher is to draw the wrong conclusion and how many cases there are.

Just as it is not fair to change from a two-tailed test to a one-tailed test after looking at the results, the decision rule is made in advance. That way, if the results are just short of the point where the null hypothesis is rejected, the researcher won't be tempted to slide the critical value over a bit to get the results he or she desires.



Calculating the value for a test statistic is the fifth step in hypothesis testing.

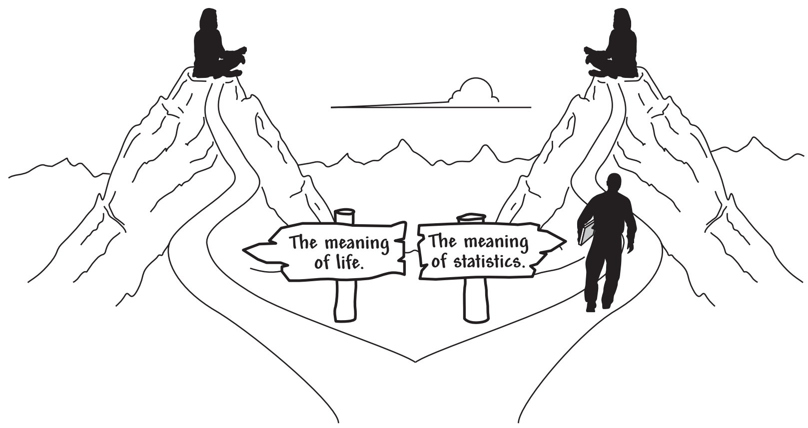
Step 5 Calculate the Test Statistic

Calculating the test statistic is the most straightforward of the six steps for hypothesis testing. Plug the right numbers into a formula, and push the right buttons in the right order on the calculator: that's how to calculate the test statistic.

Step 6 Interpret the Results

Interpreting the results is the reason statistical tests are done. In Step 6, the researcher explains, in plain language, what the results are and what they mean. Interpretation is a human skill.

Interpretation involves answering questions about the results. For example, in this chapter, we'll ask whether the null hypothesis was rejected. In future chapters, more interpretation questions will be added and the questions will change slightly from hypothesis test to hypothesis test. When there are multiple questions, they build on each other, so a researcher gains a greater understanding of the results. It is possible to stop after any question with enough information to offer an interpretation, but the more questions answered, the better a researcher will understand the results.



Interpretation, explaining what the results mean, is the goal of statistics.

There are four parts to writing an interpretation:

1. Recap why the study was done.
2. Provide the main factual results, for example, the mean scores for the control group and the experimental group.
3. Explain what the results mean.
4. Make suggestions for future research.

This four-part guide to writing an interpretation is presented in Table 6.2. It may look like interpretations will be long and detailed, but short and clear is better. For many studies, a good interpretation can be accomplished in one paragraph of four or five sentences.

TABLE 6.2 Template for Writing a Four-Point Interpretation Paragraph

1.	Recap the study. What was done? Why?
2.	Present the main results factually. For example, what were the mean scores for the control and experimental groups? Present the results of the hypothesis test in APA format.
3.	Explain what the results mean.
4.	Make suggestions for future research. What were the strengths and/or weaknesses of this study? What should be done in the next study?

A thorough interpretation can often be accomplished in one paragraph. In an interpretation, the results are presented both objectively (e.g., group means) and subjectively (the researcher's explanation).

The Single-Sample *z* Test

Now that we've covered the general logic behind hypothesis testing, let's move on to the specific steps of the single-sample *z* test.

Step 1 Pick a Test. The single-sample *z* test is the test to use to see whether adopted children differ in intelligence from the general population. Dr. Pell has selected it because the single-sample *z* test is used to compare a sample mean to a population mean when the standard deviation of the population is known. (It is known that the population standard deviation, σ, for IQ is 15.)

Step 2 Check the Assumptions. The assumptions for the single-sample *z* test are listed in Table 6.3. The table also notes whether the assumptions are robust or not.

TABLE 6.3 Assumptions for the Single-Sample *z* Test

	Assumption	Robustness
1. Random sample	The sample is a random sample from the population.	Robust
2. Independence of observations	The cases in the sample are independent of each other.	Not robust
3. Normality	The dependent variable is normally distributed in the population.	Robust

If a nonrobust assumption is violated, one needs to find a different statistical test, one with different assumptions, to use.

- **Random sample:** The first assumption, that the sample is a random sample from the population, is not violated. In the IQ example, the sample was a random sample from the population of adopted children. This is a robust assumption, so even if the sample were not a random one, we could still proceed with the test. One just needs to be careful about the population to which one generalizes the results.
- **Independence of observations:** The second assumption is that the observations within the group are independent. This assumption means that the scores of cases in the sample aren't influenced by other cases in the sample. In the IQ example, the participants were randomly sampled, each case was in the sample only once, and each case was tested individually, so this assumption was not violated. This is not a robust assumption, so if the cases were not independent, we would not be able to proceed with the test.
- **Normality:** The third assumption is that the dependent variable is normally distributed in the population. Intelligence is one of the many variables that psychologists assume to be normally distributed, so that assumption is not violated. This is also a robust assumption. So if it were violated, as long as the violation is not too great, we would still be able to proceed with the test.

With no assumptions violated, Dr. Pell can proceed with the planned test.

A Common Question

Q Suppose a teacher has developed a new way of teaching math and then tests the new method with three different groups, each with five students being taught together. When he uses the final exam to evaluate his new method, should he treat the data as 15 scores, one from each student, or three scores, one from each group?

A To avoid violating the independence of observations assumption, this should be treated as three scores, each score the mean of the five students taught together in a group. However, few researchers are this rigorous. Researchers worry more about violating this assumption by including the same participant in a study twice.

Step 3 List the Hypotheses. With a two-tailed test, it is easier to generate the null hypothesis first and the alternative hypothesis second. The null hypothesis is going to be (1) a statement about the population, (2) a negative statement, and (3) a specific statement. All of these conditions are met by

$$H_0: \mu_{\text{AdoptedChildren}} = 100$$

Dr. Pell's null hypothesis says that the population of adopted children has a mean IQ of 100.

- Because it is about μ, the population mean, it is a statement about a population.
- It is a negative statement because it says that the intelligence of adopted children is *not* different from the mean IQ for the population of children in general.
- It is a specific statement because it says the population mean is exactly 100.

Next, Dr. Pell needs to state the alternative hypothesis. The alternative hypothesis has to be mutually exclusive to the null hypothesis, and the two hypotheses together have to be all-inclusive. Further, the alternative hypothesis is not going to make a specific prediction. The alternative hypothesis is just going to say that the null hypothesis is wrong. This means that the alternative hypothesis states the population mean for adopted children is something other than 100:

$$H_1: \mu_{\text{AdoptedChildren}} \neq 100$$

Step 4 Set the Decision Rule. Hypothesis testing works by assuming that the null hypothesis is true. Assume that the null hypothesis, $\mu = 100$, is true and imagine taking hundreds and hundreds of repeated random samples of size 72 from the population of adopted children. For each sample, calculate a mean and then make a sampling distribution for all the means from all the samples. Thanks to the central limit theorem, three things are known about this sampling distribution: (1) the sampling distribution of the mean will be centered at 100, (2) it will have a normal shape, and (3) its standard deviation, called the standard error of the mean, can be calculated using Equation 5.1. For the adoption IQ example, this would be

$$\begin{aligned}\sigma_M &= \frac{\sigma}{\sqrt{N}}\\ &= \frac{15}{\sqrt{72}}\\ &= \frac{15}{8.4853}\\ &= 1.7678\\ &= 1.77\end{aligned}$$

The sampling distribution of the mean that should occur if the null hypothesis is true is shown in **Figure 6.1**. Note that it has a normal shape, is centered around the population mean of 100, and that IQ scores on the *X*-axis are marked off by the size of the standard error of the mean, 1.77.

Now, divide the sampling distribution into two parts (**Figure 6.2**). The middle section is called the **common zone,** because it is the section in which sample means commonly fall. The two extreme sections form the **rare zone** because it is the part of the sampling distribution in which it is rare that a sample mean falls.

Next, place the observed sample mean, 104.5, in the sampling distribution of the mean.

- If the observed sample mean falls in the common zone, there's no reason to question the null hypothesis. A common result, what is expected to happen if the null hypothesis is true, did happen.

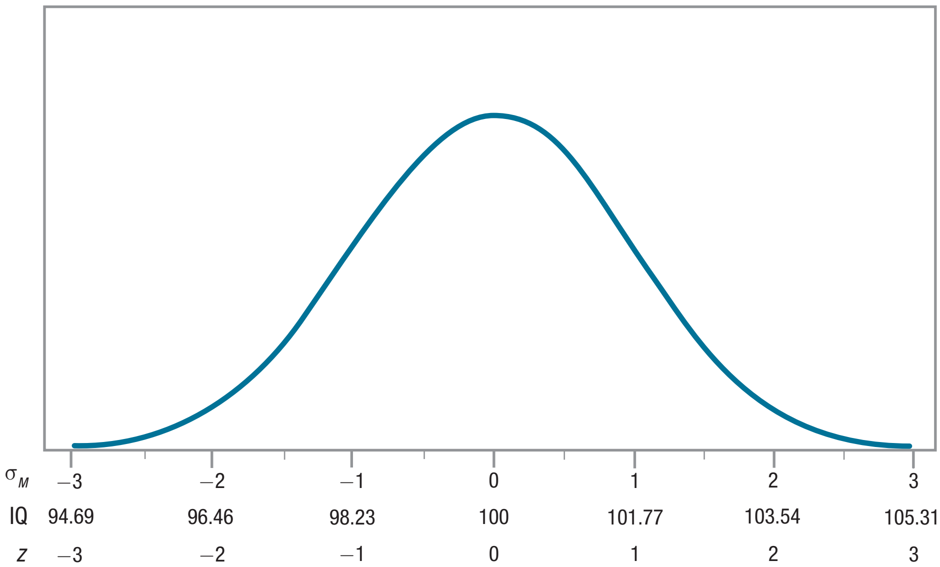


Figure 6.1 Sampling Distribution of the Mean This distribution shows what the sampling distribution of the mean would look like for repeated random samples of size 72 from a population with a mean IQ of 100. Note that it is normally distributed and centered around the population mean. The standard error of the mean, σ_M, is 1.77 and is being used to mark off IQ scores by standard deviation units on the *X*-axis.

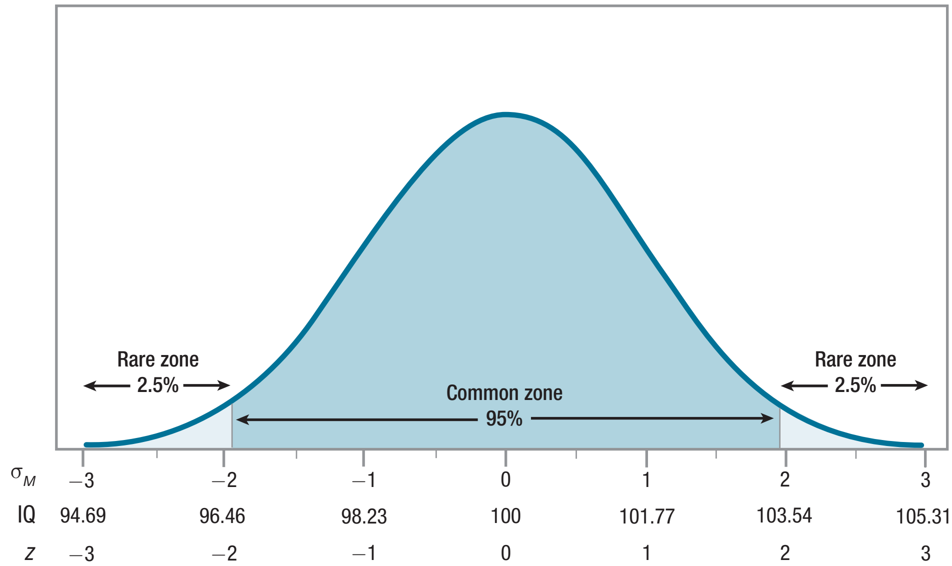


Figure 6.2 Sampling Distribution of the Mean with the Common Zone and Rare Zone Marked for IQ Data When the alpha level is set at .05 and the hypotheses are nondirectional, the common zone captures the middle 95% of the sampling distribution, and the rare zone captures the extreme 5%. The border between the rare zone and common zone occurs at *z* scores of −1.96 and 1.96, respectively, because they are the *z* values that separate the middle 95% of a normal distribution from the extreme 5%.

- If the observed sample mean falls in the rare zone, then an unusual event has happened. In this situation, by the logic of hypothesis testing, the null hypothesis is rejected and the researcher is forced to accept the alternative hypothesis as true. As the alternative hypothesis is what the researcher believes to be true, he or she is happy to be "forced" to accept it.

The line drawn between the common zone and the rare zone is called, as was noted above, the critical value. The critical value of *z*, abbreviated z_{cv}, depends on how small the researcher wants the rare zone to be. The convention most commonly used in statistics is to say that something that happens 95% of the time is common and something that happens 5% of the time or less is rare. This means that the common zone is the middle 95% of the sampling distribution, and the rare zone is the extreme 5% of the sampling distribution. Notice that the rare zone is two-tailed, with 2.5% at the very top (the right-hand side) of the sampling distribution and 2.5% at the very bottom (the left-hand side) of the distribution. This is why the test is called a two-tailed test—the rare zone falls in both sides (tails) of the sampling distribution.

In Chapter 4, it was found that the *z* scores of ±1.96 marked off the middle 95% of a normal distribution from the extreme 5%. Now, the *z* scores of ±1.96 will be used as the critical values, z_{cv}, for the single-sample *z* test. Figure 6.2 shows how the critical values divide the sampling distribution of the expected outcomes into common and rare zones. Here is the decision rule:

- If the *z* score calculated in Step 5 (the next step) is less than or equal to −1.96 *or* if the *z* score is greater than or equal to 1.96, the researcher will reject the null hypothesis.
  - Written mathematically, this is: if $z \leq -1.96$ *or* if $z \geq 1.96$, then reject H_0.

- If the z score calculated in Step 5 is greater than -1.96 *and* less than 1.96, then the researcher will fail to reject the null hypothesis.
  - Written mathematically: if $-1.96 < z < 1.96$, then fail to reject H_0.
- Notice that the critical value itself is part of the rare zone. If $z = -1.96$ or $z = 1.96$, then the researcher will reject the null hypothesis.

The size of the rare zone, expressed as a probability, is the *alpha level* of the test. The **alpha level,** or **alpha,** is the probability that a result will fall in the rare zone and the null hypothesis will be rejected when the null hypothesis is really true. Alpha levels are also called **significance levels.**

When statisticians discuss alpha levels, they use proportions, not percentages. So in this example, alpha is set at .05. Alpha is abbreviated with a lowercase Greek letter, α, so one would write, "$\alpha = .05$." This means that the researcher considers a rare event something that happens no more than 5% of the time.

The critical value that is the dividing line between the rare zone and the common zone depends on (1) whether the researcher is doing a one-tailed or a two-tailed test and (2) what alpha level is selected. The most commonly used options for a single-sample z test are shown in Table 6.4.

TABLE 6.4 Critical Values of *z*, One-Tailed and Two-Tailed, for Commonly Used Alpha Levels for Single-Sample *z* Tests

	Critical Values of *z*			
df	$\alpha = .10$, two-tailed *or* $\alpha = .05$, one-tailed	$\alpha = .05$, two-tailed *or* $\alpha = .025$, one-tailed	$\alpha = .02$, two-tailed *or* $\alpha = .01$, one-tailed	$\alpha = .01$, two-tailed *or* $\alpha = .005$, one-tailed
	$z_{cv} = 1.65$	$z_{cv} = 1.96$	$z_{cv} = 2.33$	$z_{cv} = 2.58$

Note that as the alpha level gets smaller, the critical value of z moves farther away from zero, making the rare zone smaller.

Step 5 Calculate the Test Statistic. The test statistic to be calculated is a z value. So, Step 5 involves turning the observed sample mean of 104.50 into a z score. The formula for calculating the z value for a single-sample test is shown in Equation 6.1.

Equation 6.1 Formula for the Single-Sample *z* Test

$$z = \frac{M - \mu}{\sigma_M}$$

where z = the z score
M = the sample mean
μ = the population mean
σ_M = the standard error of the mean (Equation 5.1)

Earlier, using Equation 5.1, the standard error of the mean for the adoption IQ example was calculated as $\sigma_M = 1.77$. With that value, and with $M = 104.5$ and

$\mu = 100$, Dr. Pell can calculate the value of z for the single-sample z test for the adoption IQ study:

$$z = \frac{M - \mu}{\sigma_M}$$
$$= \frac{104.50 - 100}{1.77}$$
$$= \frac{4.5000}{1.77}$$
$$= 2.5424$$
$$= 2.54$$

Having calculated $z = 2.54$, this step is over.

Step 6 Interpret the Results. In interpreting the results, plain language is used to explain what the results mean. In this chapter, we'll start with the most basic interpretative question, "Was the null hypothesis rejected?"

This is addressed by comparing the observed value of the test statistic, 2.54, to the critical value, ± 1.96. Which of the following statements generated in Step 4 is true?

- Is either $2.54 \leq -1.96$, or is $2.54 \geq 1.96$?
- Is $-1.96 < 2.54 < 1.96$?

The first statement is true as 2.54 is greater than or equal to 1.96. This means that the results fall in the rare zone and the null hypothesis is rejected. (To help visualize the results falling in the rare zone, see Figure 6.3.) By rejecting the null hypothesis, Dr. Pell must accept the alternative hypothesis and conclude that the population mean is something other than 100.

It is possible to go a step beyond just saying that the population mean of IQ for adopted children is something other than 100. It is possible to comment on the direction of the difference by comparing the sample mean to the population mean. The sample mean of 104.5 was above the general population mean of 100, so Dr. Pell

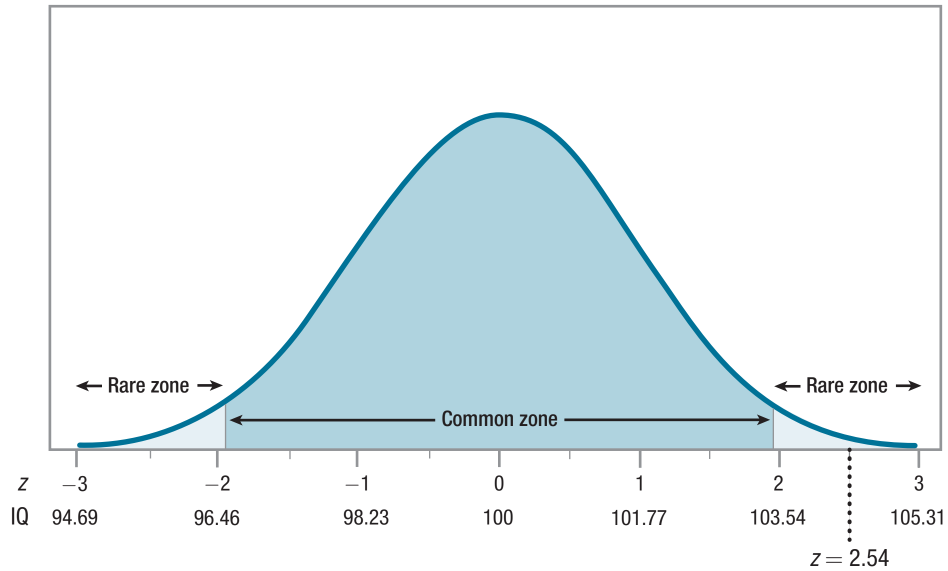


Figure 6.3 Sampling Distribution of the Mean, Marked with Observed z Score, for IQ Data The observed mean of 104.5 has been converted to a z score of 2.54. Note that it falls in the rare zone, which means the null hypothesis is rejected, and the alternative hypothesis has to be accepted.

can conclude that the population of adopted children has an average intelligence that is higher than 100.

The terms "statistically significant" and "not statistically significant" are commonly used in reporting results. **Statistically significant** means that a researcher concludes the observed sample results are more different from the null-hypothesized population value than would be expected by chance. With the adoption IQ example, Dr. Pell could state that the observed mean of 104.5 is "statistically significant" or "statistically different from 100."

It is also common to report the results in APA format. APA format indicates what test was done, how many cases there were, what the value of the test statistic was, what alpha level was selected, and whether the null hypothesis was rejected. In APA format, Dr. Pell would report the results for the adoption IQ study as

$$z\,(N = 72) = 2.54,\ p < .05$$

- z tells what test was done, a z test.
- $N = 72$ gives the sample size.
- 2.54 is the value of the test statistic that was calculated.
- .05 refers to the alpha level. (Because it is associated in APA format with the letter p, for probability, it is often referred to as the ***p* value.**)
- $p < .05$ informs the reader that the null hypothesis was rejected.

A Common Question

Q What does $p < .05$ mean?

A $p < .05$ means that the observed result is a rare result as the probability of it happening is less than .05 if the null hypothesis is true. If a researcher fails to reject the null hypothesis, he or she would write $p > .05$. This means that the result is a common occurrence—it happens more than 5% of the time—when the null hypothesis is true.

Here is what the psychologist, Dr. Pell, wrote for her interpretation. Note how she follows the four-point interpretation template from Table 6.2: (1) indicating what was done, (2) providing some facts (the sample and population means as well as the z test results), (3) telling what the results mean, and (4) making a suggestion for future research:

> In this study, the intelligence of adopted children was compared to the IQ of the general population of children in the United States. The mean IQ of a random sample of 72 adopted children in the United States was 104.5. Using a single-sample z test, their mean of 104.5 was statistically significantly different from the population mean of 100 ($z\,(N = 72) = 2.54,\ p < .05$). This study shows that adopted children have a higher average IQ than children in general. As adoptive parents are carefully screened before they are allowed to adopt, future research may want to explore the role that this plays in the higher intelligence of their children.

Worked Example 6.1

Let's practice with hypothesis testing and the single-sample z test, this time with an example where the null hypothesis is not rejected. Imagine a psychic who claimed he could "read" blood pressure. A public health researcher, Dr. Levine, tested the man's claim by asking him to select people with abnormal blood pressure. The psychic picked out 81 people. Dr. Levine took their blood pressures and the average systolic blood pressure for the sample, M, was 127. From previous research, Dr. Levine knew that the population mean, μ, for systolic blood pressure was 124 with a population standard deviation, σ, of 18.

The sample mean in this study, 127, is three points higher than the population mean. Is this different enough from normal blood pressure to support the psychic's claim that he can read blood pressure? Did he find people with higher than average blood pressure? Or, can the deviation of 127 from 124 be explained by sampling error?

Step 1 Pick a Test. Comparing a sample mean to a population mean when the population standard deviation is known calls for a single-sample z test.

Step 2 Check the Assumptions. Table 6.3 lists the assumptions for the single-sample z test:

- The sample is not a random sample from the population of people the psychic considers to have high blood pressure, so the first assumption is violated. This is a robust assumption, however, so it can be violated and the test still completed. Dr. Levine will need to be careful about the population to which he generalizes the results.
- Each participant takes part in the study only once. There's no evidence that the 81 observations influence the measurement of each other's blood pressure. The second assumption, independence of observations, is not violated.
- Eighty-one cases is a large enough sample to graph its frequency distribution. If the distribution looks normal-ish, it seems reasonable to assume that blood pressure is normally distributed in the larger population. In this scenario, the third assumption, normality, is not violated.

With no nonrobust assumptions violated, Dr. Levine can proceed with the planned test.

Step 3 List the Hypotheses. The null hypothesis states that the psychic is *not* able to read blood pressure. As a result, the mean blood pressure of the people he picks should be no different from the mean blood pressure of those with normal pressure. This is stated mathematically as

$$H_0: \mu_{\text{PsychicSelected}} = 124$$

The null hypothesis and the alternative hypothesis, together, have to be all-inclusive and mutually exclusive. The alternative hypothesis will state that the mean blood pressure in the population of people the psychic selects is different from the mean normal systolic blood pressure of 124:

$$H_1: \mu_{\text{PsychicSelected}} \neq 124$$

Step 4 Set the Decision Rule. Dr. Levine wants to do a two-tailed test and has set alpha at .05. According to Table 6.4, the critical value of z is ± 1.96. If the value of z calculated in the next step is less than or equal to -1.96 *or* greater than or equal to 1.96, the researcher will reject the null hypothesis. If z is greater than -1.96 *and* less than 1.96, he will fail to reject the null hypothesis. Figure 6.4 displays the rare and common zones for this decision.

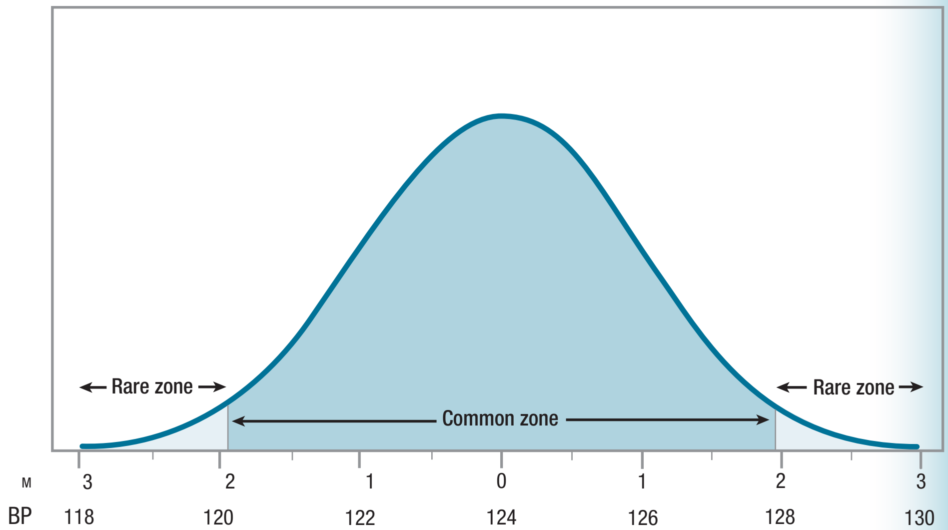


Figure 6.4 Sampling Distribution of the Mean, Marked with the Rare Zone and Common Zone, for Psychic Blood Pressure Data This is the sampling distribution of the mean for the blood pressure data with a population mean of 124 and a standard error of the mean of 2.00.

Step 5 Calculate the Test Statistic. In order to calculate the z value, Dr. Levine will first use Equation 5.1 to calculate the standard error of the mean:

$$\begin{aligned}\sigma_M &= \frac{\sigma}{\sqrt{N}}\\ &= \frac{18}{\sqrt{81}}\\ &= \frac{18}{9.0000}\\ &= 2.0000\\ &= 2.00\end{aligned}$$

Now that the standard error of the mean is known, Dr. Levine can use Equation 6.1 to calculate the z value:

$$\begin{aligned}z &= \frac{M - \mu}{\sigma_M}\\ &= \frac{127 - 124}{2.00}\\ &= \frac{3.0000}{2.00}\\ &= 1.5000\\ &= 1.50\end{aligned}$$

Step 6 Interpret the Results. Which of the following two statements is true?

- Is either $1.50 \leq -1.96$, or is $1.50 \geq 1.96$?
- Or, is $-1.96 < 1.50 < 1.96$?

The second statement is true as 1.50 falls between −1.96 and 1.96. Dr. Levine has failed to reject the null hypothesis. Figure 6.5 shows how the observed value of the test statistic, 1.50, falls in the common zone.

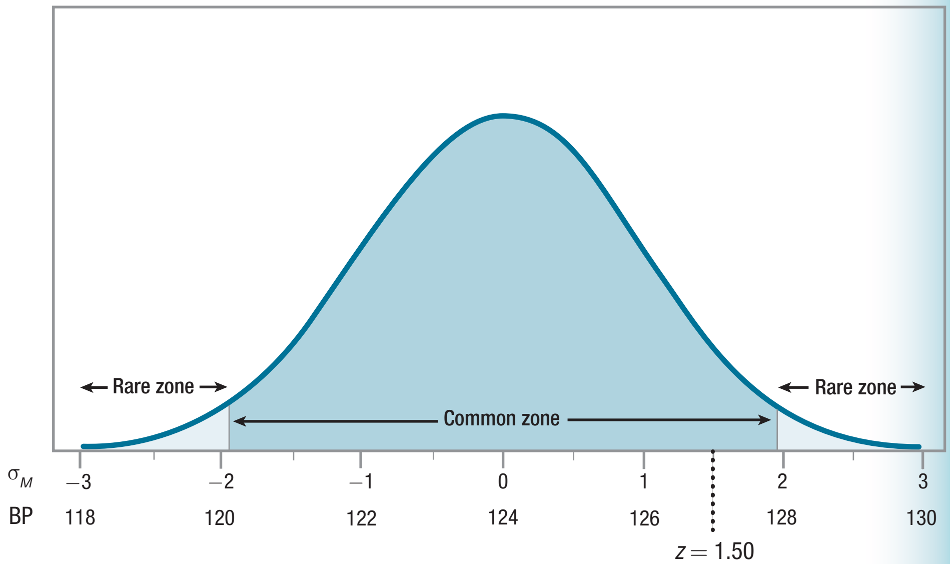


Figure 6.5 Sampling Distribution of the Mean for the Blood Pressure Data, Marked with Observed Value of z The sample mean of 127 is equivalent to a z value of 1.50. Note that this falls in the common zone, meaning that one fails to reject the null hypothesis.

The results are not statistically significant. Not enough evidence exists to conclude that this sample mean, 127, is statistically different from 124. Because there is not sufficient evidence of a difference, the direction of the difference doesn't matter.

In APA format, a researcher indicates failure to reject the null hypothesis by writing "$p > .05$." (.05 is used because alpha was set at .05.) This indicates that the value of the test statistic fell in the common zone as it is a value that happens more than 5% of the time when the null hypothesis is true. In APA format for these results, Dr. Levine would write

$$z\,(N = 81) = 1.50,\ p > .05$$

Here is what Dr. Levine wrote for an interpretation. Note that he follows the template in Table 6.2: (1) telling what was done, (2) providing some facts,

(3) explaining what the results mean, and (4) making a suggestion for future research. Again, the interpretation takes just a paragraph:

> This study was set up to determine if a psychic could "read" blood pressure as he claimed. The psychic selected 81 people whom he believed had abnormal blood pressure. Their systolic blood pressures were measured ($M = 127$) and compared to the population mean ($\mu = 124$) using a single-sample z test. No statistically significant difference was found ($z\,(N = 81) = 1.50$, $p > .05$), indicating there was no evidence that this group of people selected for having high blood pressure had above-average blood pressure. Based on these data, there is not enough evidence to suggest that this psychic has any ability to read blood pressure. Though the results from this study seem conclusive, if one wanted to test this psychic's ability again, it would be advisable to use a larger sample size.

Practice Problems 6.2

Review Your Knowledge

6.03 What are the six steps of hypothesis testing?

Apply Your Knowledge

6.04 If $N = 55$ and $\sigma = 12$, what is σ_M?

6.05 If $M = 19.40$, μ is 22.80, and $\sigma_M = 4.60$, what is z?

6.06 A researcher believes the population mean is 20 and the population standard deviation is 4. He takes a random sample of 64 cases from the population and calculates $M = 20.75$. (a) Do the calculations for a single-sample z test and (b) report the results in APA format.

6.3 Type I Error, Type II Error, Beta, and Power

Imagine a study in which a clinical psychologist, Dr. Cipriani, is investigating if a new treatment for depression works. One way that her study would be successful is if the treatment were an effective one and the data she collected showed that. But, her study would also be successful if the treatment did not help depression and her data showed that outcome.

Those are "good" outcomes; there are also two "bad" outcomes. Suppose the treatment really is an effective one, but her sample does not show this. As a result of such an outcome, this effective treatment for depression might never be "discovered." Another bad outcome would occur if the treatment were in reality an ineffective one, but for some odd reason it worked on her subjects. As a result, this ineffective treatment would be offered to people who are depressed and they would not get better.

Table 6.5 shows the four possible outcomes for hypothesis tests. The two columns represent the reality in the population. The column on the left says the null hypothesis is really true. In Dr. Cipriani's terminology, this means treatment has no impact. The column on the right says the treatment does make a difference; the null hypothesis is not true. The two rows represent the conclusions based on the sample. The top row says that the results fell in the common zone, meaning that the researcher will fail to

TABLE 6.5 The Four Possible Outcomes for a Hypothesis Test

	H_0 is really true	H_0 is really not true
We fail to reject H_0	**A** Results fall in common zone. Correctly say, "Fail to reject H_0"	**B** Results fall in common zone. Erroneously say, "Fail to reject H_0"
We reject H_0	**C** Results fall in rare zone. Erroneously say, "Reject H_0"	**D** Results fall in rare zone. Correctly say, "Reject H_0"

Cells A and D are correct decisions, while B and C are erroneous decisions.

reject the null hypothesis. (Dr. Cipriani will conclude that her new treatment does not work.) The bottom row says that the results fell in the rare zone, so the researcher will reject the null hypothesis. (Dr. Cipriani will conclude that her treatment works.)

The two scenarios in which a study would be considered a success are Outcome A and Outcome D. In Outcome A, in the terminology of hypothesis testing, a researcher correctly fails to reject the null hypothesis. For Dr. Cipriani, this outcome means that treatment is ineffective and the evidence of her sample supports this. In Outcome D, the null hypothesis is correctly rejected. For Dr. Cipriani, the treatment does work and she finds evidence that it does.

The two other outcomes, Outcome B and Outcome C, are bad outcomes for a hypothesis test because the conclusions would be wrong. In Outcome B, the researcher fails to reject the null hypothesis when it should be rejected. If this happened to Dr. Cipriani, she would say that there is insufficient evidence to indicate the treatment works and she would be wrong as the treatment really does work. In Outcome C, a researcher erroneously rejects the null hypothesis. If Dr. Cipriani concluded that treatment works, but it really doesn't, she would have erroneously rejected the null hypothesis.

A problem with hypothesis testing is that because the decision relies on probability, a researcher can't be sure if the conclusion about the null hypothesis is right (Outcomes A or D) or wrong (Outcomes B or C). Luckily, a researcher can calculate the probability that the conclusion is in error and know how *probable* it is that the conclusion is true. With hypothesis testing, one can't be sure the conclusion is true, but one can have a known degree of certainty.

Type I Error

Routinely, researchers decide they are willing to make a Type I error 5% of the time.

Let's start the exploration of errors in hypothesis testing with Outcome C, a wrong decision. When a researcher rejects the null hypothesis and shouldn't have, this is called a **Type I error.** In the depression treatment example, a Type I error occurs if Dr. Cipriani concludes that the treatment makes a difference when it really doesn't. If that happened, psychologists would end up prescribing a treatment that doesn't help.

Routinely, researchers decide they are willing to make a Type I error 5% of the time. Five percent is an arbitrarily chosen value, but over the years it has become the standard value. One doesn't need to use 5%. If the consequences of making a Type I error seem particularly costly, one might be willing only to risk a 1% chance of this mistake. For example, imagine using a test to select technicians for a nuclear

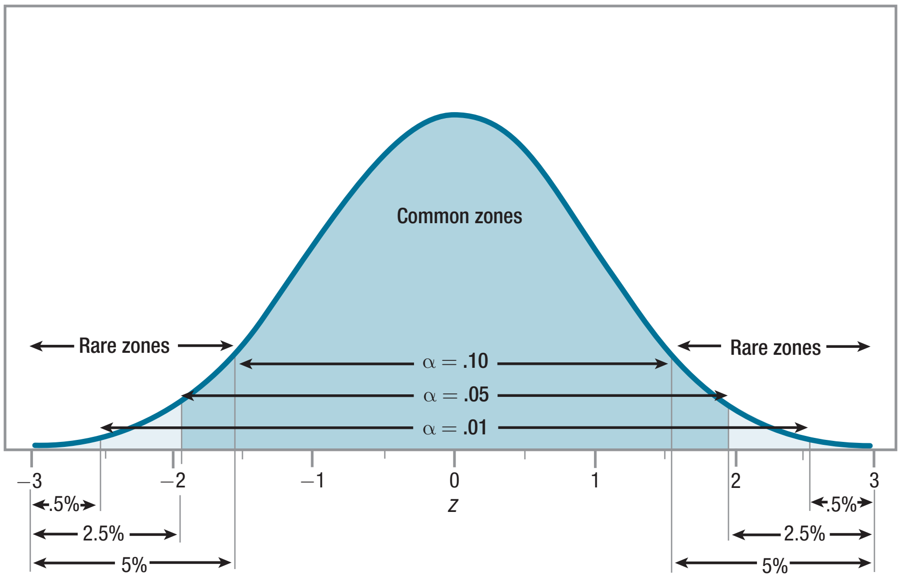


Figure 6.6 The Size of the Rare Zone and the Common Zone for Different Alpha Levels As the alpha level decreases from .10 to .05 to .01, the common zone increases in size and the rare zone shrinks. As alpha gets smaller, it is more difficult to reject the null hypothesis.

power plant. The potential catastrophe if the wrong person is hired is so great that the cut-off score on the test should be very high. On the other hand, if making a Type I error doesn't seem consequential—for example, if one is hiring people for a task that doesn't have serious consequences for failure—then one might be willing to live with a greater chance of this error.

The likelihood of Type I error is determined by setting the alpha level. As alpha gets larger, so does the rare zone. A larger rare zone makes it easier to reject the null hypothesis. Figure 6.6 shows the common and rare zones for a single-sample *z* test for three common alpha levels ($\alpha = .10$, $\alpha = .05$, and $\alpha = .01$). Note that as alpha decreases, so does the size of the rare zone, making it more difficult to reject the null hypothesis.

Setting alpha at .05 gives a modest chance of Type I error and a reasonable chance of being able to reject the null hypothesis.

Making errors is never a good idea, so why isn't alpha always set at .01 or even lower? The reason is simple. When alpha is set low, it is more difficult to reject the null hypothesis because the rare zone is smaller. Rejecting the null hypothesis is almost always the goal of a research study, so researchers would be working against themselves if they made it too hard to reject the null hypothesis. When setting an alpha level, a researcher tries to balance two competing objectives: (1) avoiding Type I error and (2) being able to reject the null hypothesis. The more likely one is to avoid Type I error, the less likely one is to reject the null hypothesis. The solution is a compromise. Setting alpha at .05 gives a modest chance of Type I error and a reasonable chance of being able to reject the null hypothesis.

Type II Error and Beta

There's another type of error possible, *Type II error.* This is Outcome B in Table 6.5. A **Type II error** occurs when the null hypothesis should be rejected, but it isn't. If Dr. Cipriani's new treatment for depression really was effective, but she didn't find evidence that it was effective, that would be a Type II error.

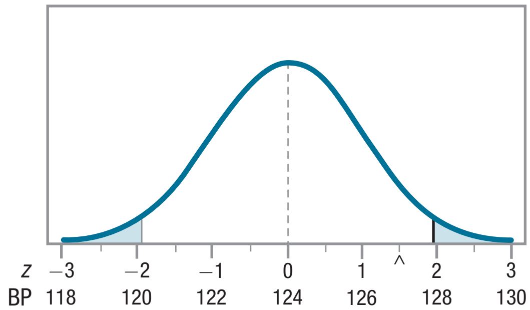


Figure 6.7 Sampling Distribution of the Mean for Blood Pressure if $\mu = 124$ The sampling distribution of the mean is centered at 124 and is normally distributed. The rare zones, where $z \leq -1.96$ and $z \geq 1.96$, are shaded in.

The probability of making a Type II error is known as **beta,** abbreviated with the lowercase Greek letter β. The consequences of making a Type II error can be as serious as the consequences of making a Type I error. But, when statisticians attend to beta, they commonly set it at .20. This gives a 20% chance for this error, not a 5% chance as is usually set for alpha.

A researcher only needs to worry about beta when failing to reject the null hypothesis. Let's use the psychic blood pressure example from Worked Example 6.1—where the null hypothesis wasn't rejected—to examine how Type II error may occur.

In that study, Dr. Levine studied a sample of 81 people whom a psychic believed had abnormal blood pressure. Their mean blood pressure was 127. The population mean was 124, with $\sigma = 18$. Dr. Levine tested the null hypothesis (H_0: $\mu = 124$) vs. the alternative hypothesis (H_1: $\mu \neq 124$) using a two-tailed single-sample z test with $\alpha = .05$. The critical value of z was ± 1.96. After calculating $\sigma_M = 2.00$, he calculated $z = 1.50$. As z fell in the common zone, he failed to reject the null hypothesis. There was insufficient evidence to say that the psychic could discern people with abnormal blood pressure.

The null hypothesis was tested by assuming it was true and building a sampling distribution of the mean centered around $\mu = 124$. In Figure 6.7, note these four things about the sampling distribution:

1. The midpoint—the vertical line in the middle—is marked on the X-axis both with a blood pressure of 124 and a z score of 0.

2. The other blood pressures marked on the X-axis are two points apart because $\sigma_M = 2$. The X-axis is also marked at these spots with z scores from -3 to 3.

3. The rare zone, the area less than or equal to a z value of -1.96 and greater than or equal to a z value of 1.96, is shaded. If a result falls in the rare zone, the null hypothesis is rejected.

4. A caret, ^, marks the spot where the sample mean, 127, falls. This is equivalent to a z score of 1.50 and falls in the common zone. The null hypothesis is not rejected.

Now, imagine that the null hypothesis is not true. Imagine that the population mean is 127 instead of 124. Why pick 127? Because that was the sample mean and is the only objective evidence that we have for what the population mean might be for people the psychic says have high blood pressure.

With this in mind, look at Figure 6.8. The top panel (A) is a repeat of the sampling distribution shown in Figure 6.7, the distribution centered around the mean of 124, the mean hypothesized by the null hypothesis. The bottom panel (B), the dotted line

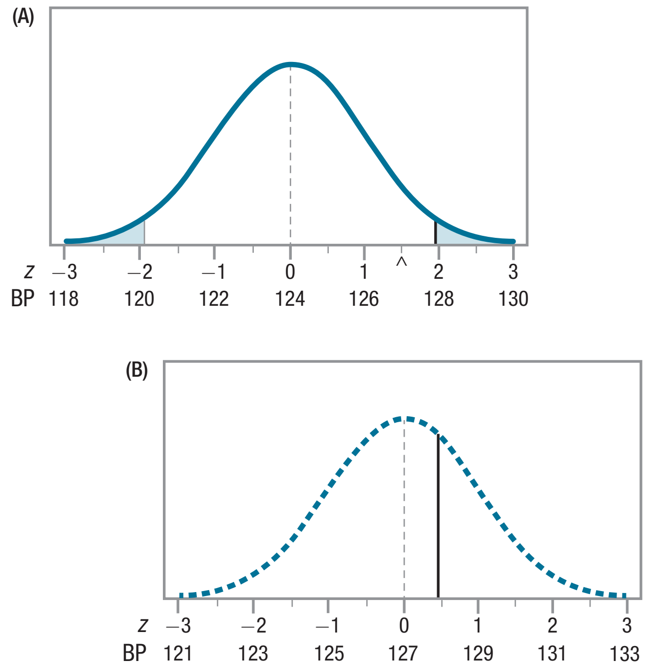


Figure 6.8 Sampling Distribution if $\mu = 127$ The top panel (A), centered at 124, is the same as the sampling distribution in Figure 6.7. The bottom panel (B) is centered at a blood pressure of 127 and shows what the sampling distribution would look like if that were the population mean. The vertical line in the bottom panel is an extension of the critical value, $z = 1.96$, from the top panel.

distribution, is new. This is what the sampling distribution would look like if $\mu = 127$. Note these four characteristics about the sampling distribution in the bottom panel:

1. The dotted line sampling distribution has exactly the same shape as the one in the top panel; it is just shifted to the right so that the midpoint, represented by a dotted vertical line, occurs at a blood pressure of 127.
2. This midpoint, 127, is right under the spot marked by a caret in the top panel. That caret represents where the sample mean, 127, fell.
3. The other points on the *X*-axis are marked off by blood pressure scores ranging from 121 to 133, and *z* scores ranging from −3 to 3.
4. There is a solid vertical line in the lower panel of Figure 6.8. The vertical line is drawn at the same point as the *z* score of 1.96 was in the top panel. This vertical line marks the point that was one of the critical values of *z* in the top panel. In the top panel, scores that fell to the left of this line fell in the common zone and scores that fell on or to the right of the line fell in the rare zone.

Figure 6.9 takes Figure 6.8 and hatches in the area to the left of the vertical line with /// in the bottom panel (B). This area, which is more than 50% of the area under the curve, indicates the likelihood of a Type II error. How so? Well, if μ were really 127, then the null hypothesis that the population mean is 124 should be rejected! However, looking at the top panel (A) to make this decision, a researcher wouldn't reject the null hypothesis after obtaining a mean that falls in the area hatched like ///.

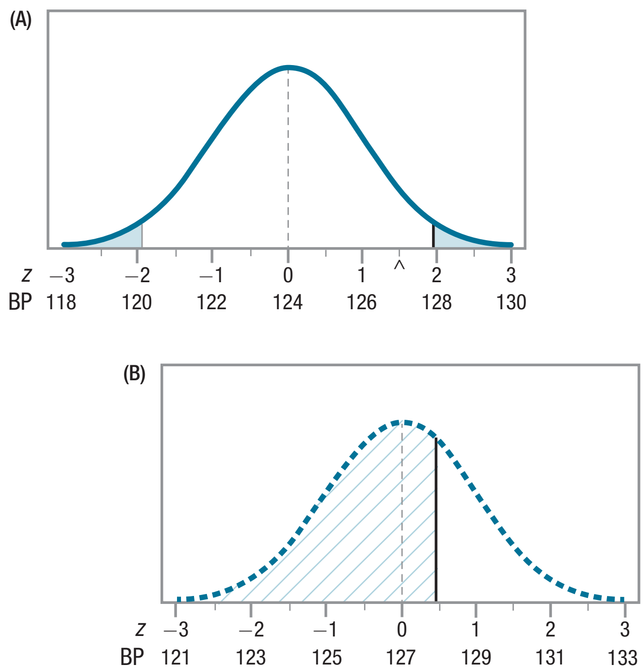


Figure 6.9 The Probability of Type II Error (Beta) if $\mu = 127$ This figure is the same as Figure 6.8, but the area to the left of the solid vertical line in the bottom panel (B) has been hatched in. Note that the solid vertical line is directly below the critical value of *z*, 1.96, in the top panel (A). Any sample mean that falls in the hatched area of the bottom panel would fall in the common zone of the top panel, and the null hypothesis would not be rejected. But, as the bottom distribution has a population mean of 127, not 124, the null hypothesis of $\mu = 124$ should be rejected. Note that more than half of the distribution is shaded in, meaning there is a large probability of Type II error if we hypothesize $\mu = 124$ but the population mean is really 127.

(Remember, the sample means will be spread around the population mean of 127 because of sampling error.) Whenever a sample mean falls in this hatched-in area, a Type II error is committed and a researcher fails to reject a null hypothesis that should be rejected.

Under these circumstances, the researcher would commit a Type II error fairly frequently if the population mean were really 127, but the null hypothesis claimed it was 124. The researcher would commit a Type II error less frequently if the population mean were greater than 127. Right now, the difference between the null hypothesis value (124) and the value that may be the population value (127) is fairly small, only 3 points. This suggests that if the psychic can read blood pressure, he has only a small ability to do so. If the effect were larger, say, the dotted line distribution shifted to a midpoint of 130, the size of the effect, the psychic's ability to read blood pressure, would be larger. The bottom panel (C) of Figure 6.10 shows the sampling distribution made with a dashed line. This is the sampling distribution that would exist if the population mean were 130. Notice how much smaller the hatched-in error area is in this sampling distribution, indicating a much smaller probability of Type II error.

The point is this: the probability of Type II error depends on the size of the effect, the difference between what the null hypothesis says is the population parameter and what the population parameter really is. When the difference between these

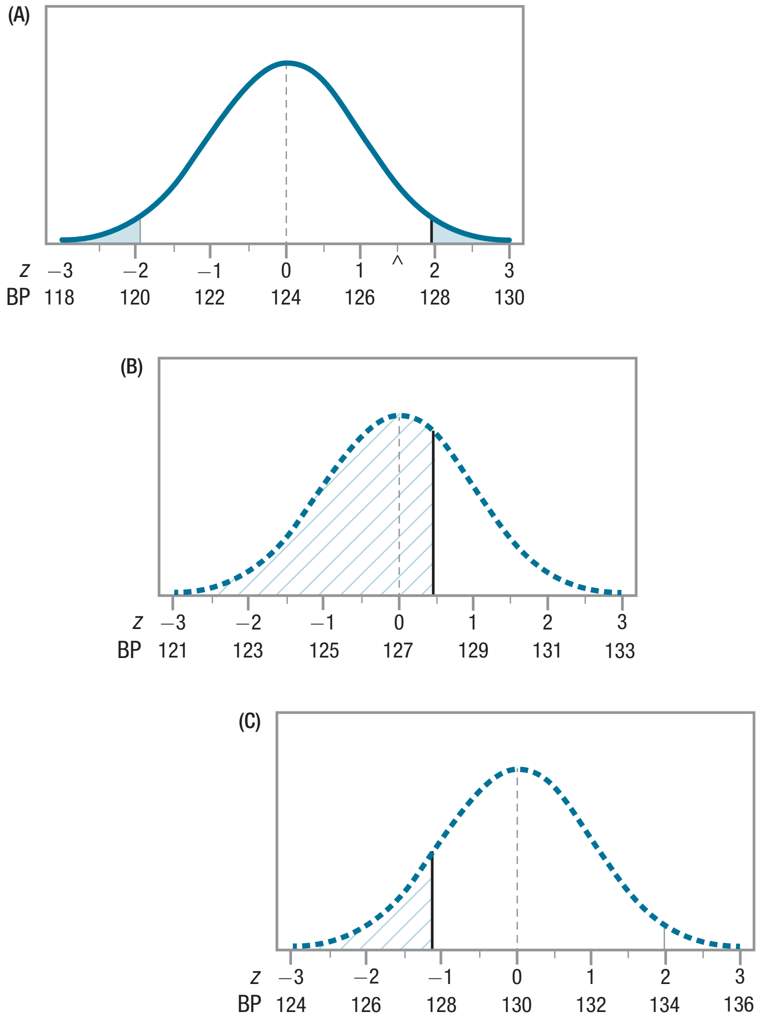


Figure 6.10 The Impact of Effect Size on the Probability of Type II Error The top panel (A) in this figure shows the sampling distribution if $\mu = 124$, the middle panel (B) if $\mu = 127$, and the bottom panel (C) if $\mu = 130$. The shaded portion in the middle panel reflects beta, the probability of Type II error if it is hypothesized that the population mean is 124 but it really is 127. Note that the shaded portion in the bottom distribution, which represents beta if $\mu = 130$, is smaller. When the effect is bigger, the probability of Type II error decreases.

values is small, the probability of a Type II error is large. Type II error can still occur when the effect size is large, but it is less likely.

When one is starting out in statistics, it is hard to remember the differences between Type I error and Type II error. Table 6.6 summarizes the differences between the two.

Power

There's one more concept to introduce before this chapter draws to a close and that is *power*. **Power** refers to the probability of rejecting the null hypothesis when the null hypothesis should be rejected. Because the goal of most studies is to reject the null hypothesis and be forced to accept the alternative hypothesis (which is what the researcher really believes is true), researchers want power to be as high as possible.

The area that is hatched \\\ in the bottom panel (C) in Figure 6.11 demonstrates the power of the single-sample z test in Dr. Levine's psychic blood pressure study. If the population mean is really 127, whenever a sample mean falls in this hatched-in

TABLE 6.6 How to Choose: Type I Error vs. Type II Error

	Type I Error	Type II Error
Definition	One rejects the null hypothesis when one should have failed to reject the null hypothesis.	One fails to reject the null hypothesis when the null hypothesis should have been rejected.
Erroneous conclusion reached	Reject the null hypothesis.	Fail to reject the null hypothesis.
Example of erroneous conclusion	Conclude that a treatment makes a difference in the outcome, but it really doesn't.	Conclude there's no evidence a treatment makes a difference in outcome, but it really does.
Probability of this error	α	β
Need to worry about making this error	If the null hypothesis is rejected.	If one fails to reject the null hypothesis.

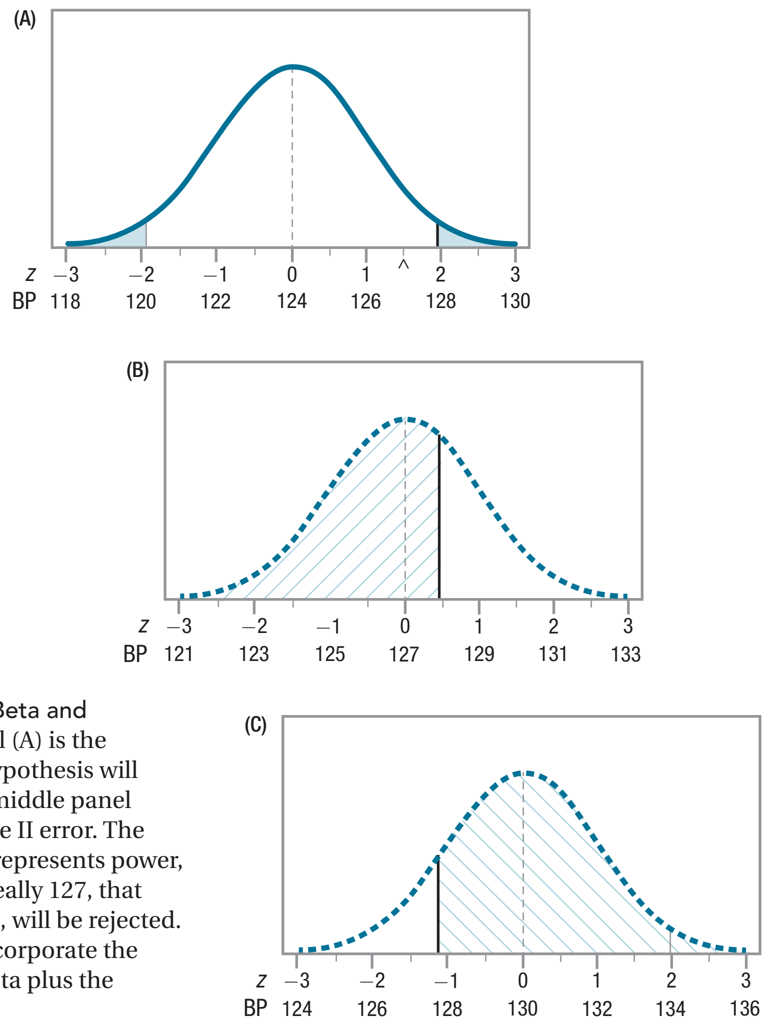


Figure 6.11 The Relationship Between Beta and Power The shaded area on the top panel (A) is the rare zone. If a result falls there, the null hypothesis will be rejected. The area hatched /// on the middle panel (B) represents beta, the probability of Type II error. The area shaded \\\ on the bottom panel (C) represents power, the likelihood if the population mean is really 127, that the null hypothesis, which claims it is 124, will be rejected. Note that the areas shaded /// and \\\ incorporate the whole distribution. (The probability of beta plus the probability of power equals 1.)

area, the researcher will reject the null hypothesis. Why? Because a population mean in this hatched-in area falls in the shaded-in rare zone shown in the top panel (A).

Look at the area hatched in /// in the middle panel (B) of Figure 6.11 and the area hatched in \\\ in the bottom panel (C) of Figure 6.11. The hatched-in area in the middle panel represents beta, the probability of Type II error, while the hatched-in area in the bottom panel represents power. Together, these two hatched-in areas shade in the entire dotted-line curve. This shows that beta and power are flip sides of the same coin. When the null hypothesis should be rejected, beta plus power equals 100% of the area under the dotted-line curve.

Statisticians refer to beta and power using probabilities, not percentages. They would say: $\beta + \text{power} = 1.00$. If either beta or power is known, the other can be figured out. If beta is set at .20, then power is .80. If power is .95, then beta is .05.

Type I error and Type II error are errors of the conclusion, not errors of the facts on which conclusions are based.

There's more to be learned about Type I and Type II errors, beta, and power in later chapters. For now, here's a final point. If a Type I error or a Type II error occurs, this doesn't mean the results of the study are wrong. After all, the sample mean *is* the mean that was found for the sample. What a Type I or Type II error means is that the *conclusion* drawn from the sample mean about the population mean is wrong. Type I error and Type II error are errors of the conclusion, not errors of the facts on which conclusions are based.

Practice Problems 6.3

Review Your Knowledge

6.07 When does Type II error occur?

6.08 What is power?

Apply Your Knowledge

6.09 What is a correct conclusion in hypothesis testing?

6.10 What is an incorrect conclusion in hypothesis testing?

Application Demonstration

To see hypothesis testing in action, let's leave numbers behind and turn to a simplified version of one way the sex of a baby is identified before it is born. Around the fifth month of pregnancy, an ultrasound can be used to identify whether a fetus is a boy or a girl. The test relies on hypothesis testing to decide "boy" or "girl," and it is not 100% accurate. Both Type I errors and Type II errors are made. Here's how it works.

The null hypothesis, which is set up to be rejected, says that the fetus is a girl. Null hypotheses state a negative, and the negative statement here is that the fetus has no penis. Together, the null hypothesis and the alternative hypothesis have to be all-inclusive and mutually exclusive, so the alternative hypothesis is that the fetus has a penis.

The sonographer, then, starts with the hypothesis that the fetus is a girl. If the sonographer sees a penis, the null hypothesis is rejected and the alternative hypothesis is accepted, that the fetus is a boy.

Of course, mistakes are made. Suppose the fetus is a girl and a bit of her umbilical cord is artfully arranged so that it masquerades as a penis. The sonographer would say “Boy” and be wrong. This is a Type I error, where the null hypothesis is wrongly rejected.

A Type II error can be made as well. Suppose the fetus is a shy boy who arranges his legs or hands to block the view. The sonographer doesn’t get a clear view of the area, says “Girl,” and would be wrong. This is a Type II error, wrongly failing to reject the null hypothesis.

However, in this case, a smart sonographer doesn’t say “Girl.” Instead, the sonographer, failing to see male genitalia, indicates that there’s not enough evidence to conclude the fetus is a boy. That is, there’s not enough evidence to reject the hypothesis that it’s a girl. Absence of evidence is not evidence of absence.

DIY

Grab a handful of coins, say, 25, and let’s investigate Type I error. Type I error occurs when one erroneously rejects the null hypothesis. In this exercise, the null hypothesis is going to be that a coin is a fair coin. This means that each coin has a heads and a tails, and there is a 50-50 chance on each toss that a heads will turn up.

For each coin in your handful, you will determine if it is a fair coin by tossing it, by observing its behavior. Imagine you toss the coin once and it turns out to be heads. Does that give you any information whether it is a fair coin or not? No, it doesn’t. A fair coin could certainly turn up heads on the first toss. So, you toss it again and again it turns up heads. Could a fair coin turn up heads 2 times in a row? Yes. There are four outcomes (HH, HT, TH, and TT), one of which is two heads, so two heads in a row should happen 25% of the time. How about three heads in a row? That will happen 12.5% of the time with a fair coin. Four heads? 6.25% of the time. Finally, if there are five heads in a row, we will step over the 5% rule. Five heads in a row, or five tails in a row, will happen only 3.125% of the time with a fair coin. It could happen, but it is a rare event, and when a rare event occurs—one that is unlikely to occur if the null hypothesis is true—we reject the null hypothesis.

So, here’s what you do. Test one coin at a time from your handful. Toss each coin 5 times in a row. If the result is a mixture of heads and tails, which is what you’d expect if the test were fair, there is not sufficient evidence to reject the null hypothesis. But, if it turns up five heads in a row or five tails in a row, you have to reject the null hypothesis and conclude the coin is not fair. If that happens to you, then pick up the coin and examine it. Does it have a heads and a tails? Toss it 5 more times. Did it turn up five heads or five tails again? This time, it probably behaved more like a fair coin.

There are two lessons here. First, occasionally, about 3% of the time, a fair coin will turn up heads 5 times in a row. If that happens, by the rules of hypothesis testing, we would declare the coin unfair. That conclusion would be wrong; it would be a Type I error. If you stopped there, if you did not replicate the experiment, you’d never know that you made a mistake. And that is a danger of hypothesis testing—your conclusion can be wrong and you don’t know it. So here’s the second lesson: replicate. When the same results occur 2 times in a row, you have a lot more faith in them.

SUMMARY

Explain how hypothesis testing works.

- Hypothesis testing uses data from a sample to evaluate a hypothesis about a population. If what is observed in the sample is close to what was expected based on the hypothesis, then there is no reason to question the hypothesis.
- The null hypothesis is paired with a mutually exclusive alternative hypothesis, which is what the researcher believes is true. The researcher's goal is to disprove (reject) the null hypothesis and be "forced" to accept the alternative hypothesis. If the null hypothesis is not rejected, then the researcher doesn't say it was proven, but states that there wasn't enough evidence to reject it. This is like a "not guilty" verdict because there wasn't enough evidence to make a convincing case.

List the six steps to be followed in completing a hypothesis testing.

- To complete a hypothesis test, a researcher (1) picks an appropriate test, (2) checks its assumptions to make sure it can be used, (3) lists the null and alternative hypotheses, (4) sets the decision rule, (5) calculates the value of the test statistic, and (6) interprets the results.

Explain and complete a single-sample *z* test.

- A single-sample z test is used to compare a sample mean to a population mean, or a specified value, when the population standard deviation is known. The decision rule says that if the deviation of the sample mean from the specified value can't be explained by sampling error, then the null hypothesis is rejected.

Explain the decisions that can be made in hypothesis testing.

- The conclusion from a hypothesis test may or may not be correct. It's a correct decision (1) if the null hypothesis should be rejected and it is, or (2) if the null hypothesis should not be rejected and it is not. The potential incorrect decisions are (3) the null hypothesis should not be rejected and it is (Type I error), or (4) the null hypothesis should be rejected and it is not (Type II error).
- Error can't always be avoided, but the probability of one occurring can be determined. The probability of a Type I error, alpha, is usually set, so the error occurs no more than 5% of the time. The probability of Type II error is called beta. Power is the probability of making a correct decision "1." That is, power is the probability of rejecting the null hypothesis when it should be rejected. Since the goal of research is usually to reject the null hypothesis, researchers want power to be as high as possible.

KEY TERMS

alpha or alpha level – the probability of making a Type I error; the probability that a result will fall in the rare zone and the null hypothesis will be rejected when the null hypothesis is true; often called significance level; abbreviated α; usually set at .05 or 5%.

alternative hypothesis – abbreviated H_1; a statement that the explanatory variable has an effect on the outcome variable in the population; usually, a statement of what the researcher believes to be true.

beta – the probability of making a Type II error; abbreviated β.

common zone – the section of the sampling distribution of a test statistic in which the observed outcome should fall if the null hypothesis is true; typically, 95% of the sampling distribution.

critical value – the value of the test statistic that forms the boundary between the rare zone and the common zone of sampling distribution of the test statistic.

hypothesis – a proposed explanation for observed facts; a statement or prediction about a population value.

hypothesis testing – a statistical procedure in which data from a sample are used to evaluate a hypothesis about a population.

nonrobust assumption – an assumption for a statistical test that must be met in order to proceed with the test.

null hypothesis – abbreviated H_0; a statement that in the population the explanatory variable has no impact on the outcome variable.

one-tailed hypothesis test – hypothesis that predicts the explanatory variable has an impact on the outcome variable in a specific direction.

***p* value** – the probability of Type I error; the same as alpha level or significance level.

power – the probability of rejecting the null hypothesis when the null hypothesis should be rejected.

rare zone – the section of the sampling distribution of a test statistic in which it is unlikely an observed outcome will fall if the null hypothesis is true; typically, 5% of the sampling distribution.

robust assumption – an assumption for a statistical test that can be violated to some degree and it is still OK to proceed with the test.

significance level – the probability of Type I error; the same as alpha level or *p* value.

statistically significant – when a researcher concludes that the observed sample results are different from the null-hypothesized population value.

two-tailed hypothesis test – hypothesis that predicts the explanatory variable has an impact on the outcome variable but doesn't predict the direction of the impact.

Type I error – the error that occurs when the null hypothesis is true but is rejected; $p(\text{Type I error}) = \alpha$.

Type II error – the error that occurs when we fail to reject the null hypothesis but should have rejected it; $p(\text{Type II error}) = \beta$.

CHAPTER EXERCISES

Answers to the odd-numbered exercises appear at the back of the book.

Review Your Knowledge

6.01 A hypothesis is a proposed ____ for observed ____.

6.02 The procedure by which the observation of a ____ is used to evaluate a hypothesis about a ____ is called ____.

6.03 If what is observed in a sample is close to what is expected if the hypothesis is true, there is little reason to question the ____.

6.04 A hypothesis is a statement about a ____ not a ____.

6.05 The null hypothesis is abbreviated as ____ and the alternative hypothesis as ____.

6.06 The null and alternative hypotheses must be ____ and ____.

6.07 The null hypothesis is a ____ prediction and a ____ statement.

6.08 The null hypothesis says that the explanatory variable *does/does not* have an impact on the outcome variable.

6.09 The hypothesis a researcher believes is really true is the ____ hypothesis.

6.10 One ____ prove that a negative statement is true.

6.11 It takes just one example to ____ a negative statement.

6.12 When the null hypothesis is rejected, the researcher is forced to accept the ____.

6.13 Because the null and alternative hypotheses are mutually exclusive, if one is not true, then the other is ____.

6.14 If one fails to disprove the null hypothesis, one can't say it has been ____.

6.15 One shouldn't expect the sample mean to be exactly the same as the population mean because of ____.

6.16 The mnemonic to remember the six steps of hypothesis testing is ____.

6.17 The six steps of hypothesis testing, in order, are ____.

6.18 In the first step of hypothesis testing, one picks a ____.

6.19 If the ____ of a hypothesis test aren't met, one can't be sure what the results mean.

6.20 If a robust assumption is violated, one ____ proceed with the test.

6.21 A two-tailed test has ____ hypotheses.

6.22 A two-tailed test allows one to test for a positive or a negative effect of the ____ on the ____.

6.23 It is easier to reject the null hypothesis with a ____-tailed test than a ____-tailed test.

6.24 Once the data are collected, it is *OK/not OK* to change from a two-tailed test to a one-tailed test.

6.25 The decision rule involves finding the ____ of the test statistic.

6.26 When the value of the test statistic meets or exceeds the critical value of the test statistic, one ____ the null hypothesis.

6.27 Explaining, in plain language, what the results of a statistical test mean is called ____.

6.28 A single-sample z test is used to compare a ____ mean to a population ____.

6.29 In order to use a single-sample z test, one must know the ____ standard deviation.

6.30 The random sample assumption for a single-sample z test says that the ____ is a random sample from the ____.

6.31 Independence of observations within a group means that the cases don't ____ each other.

6.32 The normality assumption states that the ____ is normally distributed in the ____.

6.33 The null hypothesis for a single-sample z test says that the ____ is a specific value.

6.34 The alternative hypothesis for a single-sample z test says that the population mean is not what the ____ indicated it was.

6.35 The common zone of the sampling distribution of the mean is centered around a z score of ____.

6.36 Sample means will commonly fall in the ____ of the sampling distribution of the mean.

6.37 It is rare that a sample mean will fall in the ____ of the sampling distribution of the mean.

6.38 If the observed mean falls in the common zone, then what was expected to happen if the ____ is true did happen.

6.39 If the sample mean falls in the rare zone, then this is a ____ event if the null hypothesis is true.

6.40 Statisticians say that something that happens more than ____% of the time is common and less than or equal to ____% of the time is rare.

6.41 The z scores that are the critical values for a two-tailed, single-sample z test with alpha set at .05 are ____ and ____.

6.42 If ____ $\leq$ ____, reject the null hypothesis.

6.43 If ____ $\geq$ ____, reject the null hypothesis.

6.44 The alpha level is the probability that an outcome that is ____ to occur if the null hypothesis is true does occur.

6.45 ____ is the abbreviation for alpha.

6.46 If alpha equals ____, then a rare event is something that happens at most only 5% of the time.

6.47 The numerator in calculating a single-sample z test is the difference between the ____ and the ____.

6.48 The denominator in calculating a single-sample z test is ____.

6.49 The first question to be addressed in an interpretation is whether one ____ the null hypothesis.

6.50 If one rejects the null hypothesis, one can decide the direction of the difference by comparing the ____ to the ____.

6.51 If the result of a single-sample z test is statistically significant, that means the sample mean is ____ from the population mean.

6.52 APA format indicates what statistical test was done, how many ____ there were, what the value of the ____ was, what ____ was selected, and whether the null hypothesis was ____.

6.53 ____, in APA format, means the null hypothesis was rejected.

6.54 ____, in APA format, means the null hypothesis was not rejected.

6.55 If one fails to reject the null hypothesis for a single-sample z test, one *does/does not* need to be concerned about the direction of the difference between the sample mean and the population mean.

6.56 If one fails to reject the null hypothesis, one says there is ____ evidence to conclude that the independent variable affects the dependent variable.

6.57 It is a correct conclusion in hypothesis testing if one rejects the null hypothesis and the null hypothesis should ____.

6.58 It is an incorrect conclusion in hypothesis testing if one ____ the null hypothesis and it should have been rejected.

6.59 With hypothesis testing, one *can / can't* be sure that the conclusion about the null hypothesis is correct.

6.60 Type ____ error occurs when one rejects the null hypothesis but shouldn't have.

6.61 If $\alpha = .05$, then the probability of making a Type I error is ____.

6.62 If the cost of making a Type I error is high, one might set alpha at ____.

6.63 Compared to $\alpha = .01$, $\alpha = .10$ has a ____ rare zone, making it ____ to reject the null hypothesis.

6.64 When alpha is set low, say, at .01, the chance of being able to reject the null hypothesis is *larger/smaller*.

6.65 In hypothesis testing, one wants to keep the probability of Type I error ____ and still have a reasonable chance to ____ the null hypothesis.

6.66 The term for the error that occurs when the null hypothesis should be rejected but isn't is ____.

6.67 The probability of Type I error is usually set at ____.

6.68 The probability of Type II error is commonly set at ____.

6.69 If one fails to reject the null hypothesis, one needs to worry about ____ error but not ____ error.

6.70 As the size of the effect increases, the probability of Type II error ____.

6.71 Power is the probability of rejecting the null hypothesis when ____.

6.72 $1.00 =$ ____ $+$ power.

6.73 If one rejects the null hypothesis, one needs to worry about ____ error but not ____ error.

6.74 If one makes a Type I error or a Type II error, then the conclusion about the ____ is wrong.

Apply Your Knowledge

Select the right statistical test.

6.75 A scientific supply company has developed a new breed of lab rat, which it claims weighs the same as the classic white rat. The population mean (and standard deviation) for the classic white rat is 485 grams (50 grams). A researcher obtained a sample of 76 of the new breed of rats, weighed them, and found $M = 515$ grams. What test should he do to see if the company's claim is true?

6.76 The mean vacancy rate for apartment rentals in the United States is 10%, with a standard deviation of 4.6. An urban studies major obtained a sample of 15 rustbelt cities and found that the mean vacancy rate was 13.3%. What statistical test should she use to see if the mean vacancy rate for these cities differs from the U.S. average?

Check the assumptions.

6.77 A researcher has a first-grade readiness test that is administered to kindergarten students and scored at the interval level. The population mean is 60, with a standard deviation of 10. He has administered it, individually, to a random sample of 58 kindergarten students in a city, $M = 66$, and wants to use a single-sample z test, two-tailed with $\alpha = .05$, in order to see whether first-grade readiness in this city differs from the national level. Check the assumptions and decide if it is OK for him to proceed with the single-sample z test.

6.78 A Veterans Administration researcher has developed a test that is meant to predict combat soldiers' vulnerability to developing post-traumatic stress disorder (PTSD). She has developed the test so that $\mu = 45$ and $\sigma = 15$. She is curious if victims of violent crime are as much at risk for PTSD as combat veterans. So, she obtains a sample of 122 recent victims of violent crime and administers the test to each one of them. She's planning to use a single-sample z test, two-tailed and with alpha set at .05, to compare the sample mean (42.8) to the population mean. Check the assumptions and decide if it is OK to proceed with the single-sample z test.

List the hypotheses.

6.79 List the null and alternative hypotheses for Exercise 6.77.

6.80 List the null and alternative hypotheses for Exercise 6.78.

State the decision rule.

6.81 State the decision rule for Exercise 6.77.

6.82 State the decision rule for Exercise 6.78.

Calculate σ_M.

6.83 Calculate σ_M using the data from Exercise 6.77.

6.84 Calculate σ_M using the data from Exercise 6.78.

Calculate z.

6.85 Calculate z for $M = 100$, $\mu = 120$, and $\sigma_M = 17.5$.

6.86 Calculate z for $M = 97$, $\mu = 85$, and $\sigma_M = 4.5$.

Calculate σ_M. Use it to calculate z.

6.87 Use the following information to calculate (a) σ_M and (b) z. $M = 12$, $\mu = 10$, $\sigma = 5$, and $N = 28$.

6.88 Use the following information to calculate (a) σ_M and (b) z. $M = 15$, $\mu = 21$, $\sigma = 1.5$, and $N = 63$.

Determine if the null hypothesis was rejected and use APA format.

6.89 Given $N = 23$ and $z = 2.37$, (a) decide if the null hypothesis was rejected, and (b) report the results in APA format. Use $\alpha = .05$, two-tailed.

6.90 Given $N = 87$ and $z = -1.96$, (a) decide if the null hypothesis was rejected, and (b) report the results in APA format. Use $\alpha = .05$, two-tailed.

Determine if the difference was statistically significant and the direction of the difference.

6.91 $M = 16$, $\mu = 11$, and the results were reported in APA format as $p < .05$. (a) Was there a statistically significant difference between the sample mean and the population mean? (b) What was the direction of the difference?

6.92 $M = 20$, $\mu = 24$, and the results were reported in APA format as $p > .05$. (a) Was there a statistically significant difference between the sample mean and the population mean? (b) What was the direction of the difference?

Given the results, interpret them. Be sure to tell what was done in the study, give some facts, and indicate what the results mean.

6.93 A researcher obtained a sample of 123 American women who said they wanted to lose weight. She weighed each of them and found $M = 178$ pounds. The mean weight for women in the United States is 164 pounds and the population standard deviation is known. The researcher used a single-sample z test and found $z\,(N = 123) = 3.68$, $p < .05$. Interpret her results to see if women who want to lose weight differ from the general population in terms of weight.

6.94 In the population of children in a school district, the mean number of days tardy per year is 2.8. A sociologist obtained a sample of children from single-parent families and found the mean number of days tardy was 3.2. He used a single-sample z test to analyze the data, finding $z\ (N = 28) = 1.68$, $p > .05$. Interpret the results to see if coming from a single-parent family is related to tardiness.

Type I vs. Type II error

6.95 A journalist was comparing the horsepower of a sample of contemporary American cars to the population value of horsepower for American cars of the 1970s. He concluded that there was a statistically significant difference, such that contemporary cars had more horsepower. Unfortunately, his conclusion was in error. (a) What type of error did the journalist make? (b) What conclusion should he have reached?

6.96 A Verizon researcher compared the number of text messages sent by a sample of teenage boys to the population mean for all Verizon users. She found no evidence to conclude that there was a difference. Unfortunately, her conclusion was in error. (a) What type of error did she make? (b) What conclusion should she have reached?

Given β or power, calculate the other.

6.97 If $\beta = .75$, power = ____.

6.98 If power = .90, β = ____.

Doing a complete statistical test

6.99 A dietitian wondered if being on a diet was related to sodium intake. She knew that the mean daily sodium intake in the United States was 3,400 mg, with a standard deviation of 270. She obtained a random sample of 172 dieting Americans and found, for sodium consumption, $M = 2{,}900$ mg. Complete all six steps of hypothesis testing.

6.100 At a large state university, the population data show that the average number of times that students meet with their academic advisors is 4.2 with $\sigma = 1.8$. The dean of student activities at this university wondered what the relation was between being involved in student clubs and organizations (such as the band, student government, or the ski club) and being involved academically in the university. She assumed that meeting with an academic advisor indicated students took their academics seriously. From the students who had been involved in student clubs and organizations for all four years of college, she obtained a random sample of 76 students, interviewed them individually, and found $M = 4.8$ for the number of times they met with their academic advisors. Complete all six steps of hypothesis testing.

Expand Your Knowledge

For Exercises 6.101 to 6.104, decide which option has a higher likelihood of being able to reject the null hypothesis.

6.101 (a) $\beta = .60$; (b) power = .60

6.102 (a) $\beta = .30$; (b) power = .50

6.103 (a) $\mu_0 = 10$, $\mu_1 = 15$; (b) $\mu_0 = 10$, $\mu_1 = 20$

6.104 (a) $\mu_0 = -17$, $\mu_1 = -23$; (b) $\mu_0 = -17$; $\mu_1 = -18$

For Exercises 6.105 to 6.108, label each conclusion as "correct" or "incorrect." If incorrect, label it as a "Type I error" or a "Type II error." Use the following scenario. A biochemist has developed a test to determine if food is contaminated or not. The test operates with the null hypothesis that a food is not contaminated.

6.105 Noncontaminated food is tested and it is called "noncontaminated."

6.106 Noncontaminated food is tested and it is called "contaminated."

6.107 Contaminated food is tested and it is called "noncontaminated."

6.108 Contaminated food is tested and it is called "contaminated."

6.109 A researcher is testing the null hypothesis that $\mu = 35$. She is doing a two-tailed test, has set alpha at .05, and has 121 cases in her sample. What is beta?

6.110 A researcher is testing the null hypothesis that $\mu = 120$, with $\sigma_M = 20$. In reality, the population mean is 140. (a) Draw a figure like Figure 6.12. Designate in the figure the areas representing Type II error and power. (b) Does one need to worry about Type I error?

For 6.111 and 6.112, determine how many tails there are.

6.111 A health science researcher believes that athletes have a lower resting heart rate than the general population. He knows, for the general population, that $\mu = 78$ and $\sigma = 9$. He is planning to do a study in which he obtains a random sample of athletes, measures their resting heart rates, and uses a single-sample *z* test to see if the mean heart rate for the sample of athletes is lower than the general population mean. (a) Should he do a one-tailed test or a two-tailed test? (b) Write the null and alternative hypotheses.

6.112 A lightbulb manufacturer believes that her compact fluorescent bulbs last longer than incandescent bulbs. She knows that the mean number of hours is 2,350 hours for a 60-watt incandescent bulb, with a standard deviation of 130 hours (these are population values). She gets a random sample of her compact fluorescent bulbs and measures the mean number of hours they last. She is going to use a single-sample *z* test to compare the sample mean to the population mean. (a) Should she do a one-tailed test or a two-tailed test? (b) Write the null and alternative hypotheses.

6.113 The population of random, two-digit numbers ranges from 00 to 99, has $\mu = 49.50$ and $\sigma = 28.87$. A statistician takes a random sample from this population and wants to see if his sample is representative. He figures that if it is representative, the sample mean will be close to 49.50. He plans to use a single-sample *z* test to test the sample against the population. Can this test be used?

SPSS

Sorry. SPSS doesn't do single-sample *z* tests. In Chapter 7, we'll learn about single-sample *t* tests, and SPSS will return!