CHAPTER SIX

INFERENC E FOR PROPORTIONS

6.1 Inference for a Single Proportion
6.2 Comparing Two Proportions

Statistical inference draws conclusions about a population or process based on sample data. It also provides a statement, expressed in terms of probability, of how much confidence we can place in our conclusions. Although there are many specific techniques for inference, there are only a few general types of statistical inference. This chapter introduces the two most common types: confidence intervals and tests of significance.
Our study of these two types of inference begins with inference about proportions. We frequently collect data on categorical variables, such as whether or not a person is employed, the brand name of a cell phone, or the country where a college student studies abroad. In these settings, our data consist of counts or of percents obtained from counts, and our goal is to say something about the corresponding population proportions. We begin in Section 6.1 with inference about a single population proportion. Section 6.2 concerns methods for comparing two proportions.

### 6.1 Inference for a Single Proportion

What percent of college students favor allowing concealed weapons on campus? What proportion of a company’s sales records have an incorrect sales tax classification? What proportion of likely voters approve of the president’s conduct in office? Often we want to know about the proportion \( p \) of some characteristic in a large population.

Suppose a market research firm interviews a random sample of 2000 adults. The result: 66% think bottled water is cleaner than tap water. That’s the truth about the 2000 people in the sample. What is the truth about the almost 235 million American adults who make up the population? Because the sample was chosen at random, it’s reasonable to think that these 2000 people represent the entire population fairly well. So the market researchers turn the fact that 66% of the sample think bottled water is cleaner into an estimate that 66% of all American adults think this way.

That’s a basic move in statistics: **use a fact about a sample to estimate the truth about the whole population.** We call this **statistical inference** because we infer conclusions about the wider population from data on selected individuals.

To think about inference, we must keep straight whether a number describes a sample or a population. Here is the vocabulary we use.

---

**PARAMETERS AND STATISTICS**

A **parameter** is a number that describes the **population**. A parameter is a fixed number, but in practice we do not know its value.

A **statistic** is a number that describes a **sample**. The value of a statistic is known when we have taken a sample, but it can change from sample to sample. We often use a statistic to estimate an unknown parameter.

For this market research study, the **parameter** is the proportion (we call it \( p \)) of adults who think bottled water is cleaner than tap water. The value of \( p \) is unknown, so the firm uses a **statistic** to estimate it, specifically the proportion of adults in the sample who think bottled water is cleaner than tap water.

---

**USE YOUR KNOWLEDGE**

**6.1 Sexual harassment of college students.** A survey of 2036 undergraduate college students aged 18 to 24 reports that 62% of college
students say they have encountered some type of sexual harassment while at college. Describe the sample and the population for this setting.

**6.2 Web polls.** If you connect to the Web site zdaily.com/polls.htm/, you will be given the opportunity to give your opinion about a different question of human interest each day. Can you apply the ideas about populations and samples that we have just discussed to this poll? Explain why or why not.

### Sample proportions

In the previous chapter, we discussed the situation where we draw a random sample of size \( n \) from a population and record the count \( X \) of “successes.” In this chapter, we focus on a closely related statistic that is the estimator of the population parameter \( p \). It is called the **sample proportion**:

\[
\hat{p} = \frac{\text{count of successes in sample}}{\text{size of sample}} = \frac{X}{n}
\]

The market research firm used the sample proportion of 66% to estimate the proportion of adults who think bottled water is cleaner. Here is another example.

**EXAMPLE 6.1**

**Adults and video games.** A DFC Intelligence report estimates that the global video game market will grow from $66 billion in 2010 to $81 billion in 2016. What proportion of adults in the United States play these games? A Pew survey, conducted by Princeton Survey Research International, reports that over half of American adults aged 18 and over play video games. The Pew survey used a nationally representative sample of 2054 adults. Of the total, 1063 adults said that they played video games. Here, \( p \) is the proportion of adults in the U.S. population who play video games and

\[
\hat{p} = \frac{X}{n} = \frac{1063}{2054} = 0.51753
\]

is the sample proportion. We estimate that 52% of adults play video games.

**USE YOUR KNOWLEDGE**

**6.3 Bank acquisitions.** The American Bankers Association Community Bank Competitiveness Survey had responses from 760 community banks. Of these, 283 reported that they expected to acquire another bank within five years.

(a) What is the sample size \( n \) for this survey?
322  CHAPTER 6  Inference for Proportions

(b) What is the count $X$ of community banks that expect to acquire another bank?

(c) Find the sample proportion $\hat{p}$.

6.4 How often do they play? In the Pew survey described in Example 6.1, those who played video games were asked how often they played. In this subpopulation, 223 adults said that they played every day or almost every day.

(a) What is the sample size $n$ for the subpopulation of U.S. adults who play video games? (Hint: Look at Example 6.1.)

(b) What is the count $X$ of those who said that they played every day or almost every day?

(c) Find the sample proportion $\hat{p}$.

While $\hat{p} = 0.518$ in Example 6.1 provides an estimate for the proportion of adults who play video games, we typically want to also know how reliable this estimate is. A second random sample of 2054 adults would have different people in it. It is almost certain that there would not be exactly 1063 positive responses. That is, the value of the statistic $\hat{p}$ varies from sample to sample.

This basic fact is called sampling variability: the value of a statistic varies in repeated random sampling. Could it happen that this second random sample of 2054 adults finds that only 38% play video games? If the variation from sample to sample is too great, then we can’t trust the results of any one sample.

All of statistical inference is based on one idea: to see how trustworthy a procedure is, ask what would happen if we repeated it many times. In terms of Example 6.1, this means we want to study the distribution of the statistic $\hat{p}$ when multiple SRSs of size $n = 2054$ are drawn. This distribution reveals the sampling variability and allows us to determine how unusual a second sample with $\hat{p} = 0.38$ would be.

Before addressing this specific example, let’s apply what we learned in Chapter 5 to describe the distribution of $\hat{p}$ in an SRS of size $n$. Recall that if the population is much larger than the sample (at least 10 times as large), the distribution of the count $X$ has approximately the binomial distribution $B(n, p)$.\(^5\)

Since the sample proportion $\hat{p}$ is just the count $X$ divided by the sample size $n$, the distribution of $\hat{p}$ is related to that of $X$. This does not mean that the proportion $\hat{p}$ has a binomial distribution. The count $X$ takes whole-number values between 0 and $n$, but a proportion is always a number between 0 and 1. We can, however, do probability calculations about $\hat{p}$ by restating them in terms of the count $X$ and using binomial methods. Here’s an example:

EXAMPLE 6.2

**Buying clothes online.** A survey by the Consumer Reports National Research Center revealed that 85% of all respondents were very or completely satisfied with their online clothes-shopping experience.\(^6\) It was also reported,
6.1 Inference for a Single Proportion

however, that people over the age of 40 were generally more satisfied than younger respondents.

You decide to take a nationwide random sample of 2500 college students and ask if they agree or disagree that “I am very or completely satisfied with my online clothes-shopping experience.” Suppose that 60% of all college students would agree if asked this question. In other words, assume that we know that the population parameter \( p = 0.60 \). For a sample of \( n = 2500 \) students, what is the probability that the sample proportion who agree is at least 58%?

Since the population of college students is much larger than 10 times the sample, the count \( X \) who agree has the binomial distribution \( B(2500, 0.60) \). The sample proportion \( \hat{p} = X/2500 \) does not have a binomial distribution, because it is not a count. However, we can translate any question about a sample proportion \( \hat{p} \) into a question about the count \( X \). Because 58% of 2500 is 1450,

\[
P(\hat{p} \geq 0.58) = P(X \geq 1450) = \sum_{X=1450}^{2500} P(X = k)
\]

This is a rather elaborate calculation. We must add more than 1000 binomial probabilities. Software tells us that \( P(\hat{p} \geq 0.58) = 0.9802 \). Because some software packages cannot handle an \( n \) as large as 2500, we need another way to do this calculation.

As a first step, find the mean and standard deviation of a sample proportion. We know how to find the mean and standard deviation of a sample count, so apply the rules for the mean and variance of a constant times a random variable. Here is the result.

**MEAN AND STANDARD DEVIATION OF A SAMPLE PROPORTION**

Let \( \hat{p} \) be the sample proportion of successes in an SRS of size \( n \) drawn from a large population having population proportion \( p \) of successes. The mean and standard deviation of \( \hat{p} \) are

\[
\mu_{\hat{p}} = p \\
\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}
\]

The formula for \( \sigma_{\hat{p}} \) is exactly correct in the binomial setting. It is approximately correct for an SRS from a large population. We will use it when the population is at least 10 times as large as the sample.

**EXAMPLE 6.3**

The mean and the standard deviation. In reference to Example 6.2 where the population proportion \( p = 0.60 \), the mean and standard deviation of the proportion of the survey respondents who are very satisfied with their online
clothes-shopping experience are

\[ \mu_p = p = 0.60 \]
\[ \sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.60)(0.40)}{2500}} = 0.0098 \]

**USE YOUR KNOWLEDGE**

**6.5 Change the sample size.** Suppose that in Example 6.2 a total of only 1250 students are sampled. Find the mean and standard deviation of the sample proportion. Compare these values with those of Example 6.3. How does this halving of n change the mean and standard deviation?

**6.6 Find the mean and the standard deviation.** If we toss a fair coin 100 times, the number of heads is a random variable that is binomial.

(a) Find the mean and the standard deviation of the sample proportion.

(b) Is your answer to part (a) the same as the mean and the standard deviation of the sample count? Explain your answer.

**Sampling distributions**

Now that we’ve found the mean and standard deviation of \( \hat{p} \), let’s take a closer look at its distribution. Since we want to know what would happen if we took many samples, here’s an approach to answer that question:

- Take a large number of samples from the same population.
- Calculate the sample proportion \( \hat{p} \) for each sample.
- Make a histogram of the values of \( \hat{p} \).
- Examine the distribution displayed in the histogram for shape, center, and spread, as well as outliers or other deviations.

In practice it is too expensive to take many samples from a large population such as all adult U.S. residents. But we can imitate many samples by using random digits. Using random digits from a table or computer software to imitate chance behavior is called **simulation**.

**EXAMPLE 6.4**

**Simulate a random sample.** Recall Example 6.2 (page xxx) where \( p = 0.60 \). We will simulate drawing a random sample of size 100 from the population of college students. Of course, we would not sample in practice if we already knew that \( p = 0.60 \). We are sampling here to understand how sampling behaves.

We can imitate the population by a table of random digits, with each entry standing for a person. Six of the 10 digits (say 0 to 5) stand for people who are very satisfied with online clothes shopping. The remaining four digits, 6 to 9, stand for those who are not. Because all digits in a random number
table are equally likely, this assignment produces a population proportion of students equal to \( p = 0.60 \). We then imitate an SRS of 100 students from the population by taking 100 consecutive digits from Table B. The statistic \( \hat{p} \) is the proportion of 0s to 5s in the sample.

Here are the first 100 entries in Table B, with digits 0 to 5 highlighted:

\[
\begin{array}{cccccccc}
19223 & 95034 & 05756 & 28713 & 96409 & 12531 & 42544 & 82853 \\
73676 & 47150 & 99400 & 01927 & 27754 & 42648 & 82425 & 36290 \\
45467 & 71709 & 77558 & 00095 \\
\end{array}
\]

There are 64 digits between 0 and 5, so \( \hat{p} = \frac{64}{100} = 0.64 \). A second SRS based on the second 100 entries in Table B gives a different result, \( \hat{p} = 0.55 \). The two sample results are different, and neither is equal to the true population value \( p = 0.60 \). That’s sampling variability.

Simulation is a powerful tool for studying chance. Now that we see how simulation works, it is faster to abandon Table B and to use a computer programmed to generate random numbers.

**EXAMPLE 6.5**

Take many random samples. Figure 6.1 illustrates the process of choosing many samples and finding the sample proportion \( \hat{p} \) for each one. Follow the flow of the figure from the population at the left, to choosing an SRS and finding the \( \hat{p} \) for this sample, to collecting together the \( \hat{p} \)'s from many samples. The histogram at the right of the figure shows the distribution of the values of \( \hat{p} \) from 1000 separate SRSs of size 100 drawn from a population with \( p = 0.60 \).

In Example 6.2, we discuss a sample of 2500 students, not 100. Figure 6.2 is parallel to Figure 6.1. It shows the process of choosing 1000 SRSs, each of size 2500, from a population in which the true proportion is \( p = 0.60 \). The 1000 values of \( \hat{p} \) from these samples form the histogram at the right of

![Image](image-url)
CHAPTER 6 Inference for Proportions

the figure. Figures 6.1 and 6.2 are drawn on the same scale. Comparing them shows what happens when we increase the size of our samples from 100 to 2500. These histograms display the sampling distribution of the statistic \( \hat{p} \) for two sample sizes.

![Sampling Distribution Diagram](image)

**FIGURE 6.2** The distribution of sample proportions for 1000 SRSs of size 2500 drawn from the same population as in Figure 6.1. The two histograms have the same scale. The statistic from the larger sample is less variable.

**SAMPLING DISTRIBUTION**

The sampling distribution of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population.

Strictly speaking, the sampling distribution is the pattern that would emerge if we looked at all possible samples of a particular size from the population. A distribution obtained from a fixed number of samples, like the 1000 SRSs in Figure 6.1, is only an approximation to the sampling distribution. We will see in Chapter 7 that probability theory, the mathematics of chance behavior, can sometimes describe sampling distributions exactly. The interpretation of a sampling distribution is the same, however, whether we obtain it by simulation or by the mathematics of probability.

We can use the tools of data analysis to describe any distribution. Let’s apply those tools to Figures 6.1 and 6.2.

- **Shape:** The histograms look Normal. Figure 6.3 is a Normal quantile plot of the values of \( \hat{p} \) for our samples of size 100. It confirms that the distribution in Figure 6.1 is close to Normal. The 1000 values for samples of size 2500 in Figure 6.2 are even closer to Normal. The Normal curves drawn through the histograms describe the overall shape quite well.

- **Center:** In both cases, the values of the sample proportion \( \hat{p} \) vary from sample to sample, but the values are centered at 0.60. We already knew this to be the case because the mean of \( \hat{p} \) is \( p \). Some samples have a \( \hat{p} \) less
6.1 Inference for a Single Proportion

FIGURE 6.3 Normal quantile plot of the sample proportions in Figure 6.1. The distribution is close to Normal except for some granularity due to the fact that sample proportions from a sample size of 100 can take only values that are multiples of 0.01. Because a plot of 1000 points is hard to read, this plot presents only every 10th value.

than 0.60 and some greater, but there is no tendency to be always low or always high. That is, \( \hat{p} \) has no bias as an estimator of \( p \). This is true for both large and small samples.

- **Spread:** The values of \( \hat{p} \) from samples of size 2500 are much less spread out than the values from samples of size 100. Less spread means a smaller standard deviation. Earlier we showed that the standard deviation of \( \hat{p} = \sqrt{p(1-p)/n} \).

Although these results describe just two sets of simulations, they reflect facts that are true whenever we use random sampling.

USE YOUR KNOWLEDGE

6.7 Effect of sample size on the sampling distribution. You are planning a study and are considering taking an SRS of either 200 or 400 observations. Explain how the sampling distribution would differ for these two scenarios.

Normal approximation for a single proportion

Using simulation, we’ve shown that the sampling distribution of a sample proportion \( \hat{p} \) is close to Normal. We also know that the distribution of \( \hat{p} \) is that of a binomial count divided by the sample size \( n \). This seems at first to be a contradiction. To clear up the matter, look at Figure 6.4. This is a probability histogram of the exact distribution of the proportion of contented shoppers \( \hat{p} \), based on the binomial distribution \( B(2500, 0.60) \). There are hundreds of narrow bars, one for each of the 2501 possible values of \( \hat{p} \). Most have probabilities too small to show in a graph. *The probability histogram looks very Normal!*
CHAPTER 6  Inference for Proportions

FIGURE 6.4  Probability histogram of the sample proportion \( \hat{p} \) based on a binomial count with \( n = 2500 \) and \( p = 0.6 \). The distribution is very close to Normal.

NORMAL APPROXIMATION FOR A SINGLE PROPORTION

Draw an SRS of size \( n \) from a large population having population proportion \( p \) of successes. Let \( X \) be the count of successes in the sample and \( \hat{p} = X/n \) be the sample proportion of successes. When \( n \) is large, the sampling distribution of \( \hat{p} \) is approximately Normal:

\[
\hat{p} \text{ is approximately } N \left( p, \sqrt{\frac{p(1-p)}{n}} \right)
\]

As a rule of thumb, we will use this approximation for values of \( n \) and \( p \) that satisfy \( np \geq 10 \) and \( n(1-p) \geq 10 \).

This Normal approximation is easy to remember because it says that \( \hat{p} \) is Normal, with its usual mean and standard deviation. Whether or not you should use this Normal approximation should depend on how accurate your calculations need to be. For most statistical purposes great accuracy is not required. Our “rule of thumb” for use of this Normal approximation reflects this judgment.

The accuracy of the Normal approximation improves as the sample size \( n \) increases. It is most accurate for any fixed \( n \) when \( p \) is close to \( 1/2 \) and least accurate when \( p \) is near 0 or 1. You can compare binomial distributions with their Normal approximations by using the Normal Approximation to Binomial applet. This applet allows you to change \( n \) or \( p \) while watching the effect on the binomial probability histogram and the Normal curve that approximates it.

EXAMPLE 6.6

Compare the Normal approximation with the exact calculation. Let’s compare the Normal approximation for the calculation of Example 6.2 with
6.1 Inference for a Single Proportion

the exact calculation from software. We want to calculate $P(\hat{p} \geq 0.58)$ when the sample size is $n = 2500$ and the population proportion is $p = 0.60$. Example 6.3 shows that

$$\mu_\hat{p} = p = 0.60$$

$$\sigma_\hat{p} = \sqrt{\frac{p(1 - p)}{n}} = 0.0098$$

Act as if $\hat{p}$ were Normal with mean 0.60 and standard deviation 0.0098. The approximate probability, as illustrated in Figure 6.5, is

$$P(\hat{p} \geq 0.58) = P\left( \frac{\hat{p} - 0.60}{0.0098} \geq \frac{0.58 - 0.60}{0.0098} \right)$$

$$= P(Z \geq -2.04) = 0.9793$$

That is, about 98% of all samples of size $n = 2500$ from the population of college students have a sample proportion that is at least 0.58. Because the sample was large, this Normal approximation is quite accurate. It misses the software value 0.9802 by only 0.0009.

**FIGURE 6.5** The Normal probability calculation for Example 6.6.

---

**USE YOUR KNOWLEDGE**

6.8 Use the Normal approximation. Suppose we toss a fair coin 100 times. Use the Normal approximation to find the probability that the sample proportion is

(a) between 0.3 and 0.7.

(b) between 0.4 and 0.65.

---

**Statistical confidence**

Now that we know the sampling distribution of $\hat{p}$, we can proceed with inference. We will consider only inference procedures based on the Normal
FIGURE 6.6 Distribution of the sample proportion for Example 6.2. The statistic \( \hat{p} \) lies within \( \pm 0.0196 \) points of \( p \) in 95% of all samples. This also means that \( p \) is within \( \pm 0.0196 \) points of \( \hat{p} \) in those samples.

\[
\text{Probability} = 0.95
\]

\[
\text{Density curve of } \hat{p} - 0.0196 \quad \hat{p} \quad \text{unknown} \quad \hat{p} + 0.0196
\]

approximation.\(^7\) In other words, we will assume that the sample size \( n \) is sufficiently large so that \( \hat{p} \) has approximately the Normal distribution with mean \( \mu_\hat{p} = p \) and standard deviation \( \sigma_\hat{p} = \sqrt{p(1 - p)/n} \).

Consider the setting of Example 6.2 and this line of thought, which is illustrated by Figure 6.6:

- Given an SRS of \( n = 2500 \) students and \( p = 0.60 \), the distribution of \( \hat{p} \) is approximately Normal with mean \( \mu_\hat{p} = 0.60 \) and standard deviation \( \sigma_\hat{p} = \sqrt{0.60(1 - 0.60)/2500} = 0.0098 \) (see Example 6.3).
- The 68–95–99.7 rule says that the probability is about 0.95 that \( \hat{p} \) will be within 0.0196 (that is, two standard deviations of \( \hat{p} \)) of the population proportion \( p \).
- To say that \( \hat{p} \) lies within 0.0196 of \( p \) is the same as saying that \( p \) is within 0.0196 of \( \hat{p} \).
- So about 95% of all samples will contain the population proportion \( p \) in the interval from \( \hat{p} - 0.0196 \) to \( \hat{p} + 0.0196 \).

We have simply restated a fact about the sampling distribution of \( \hat{p} \). The language of statistical inference uses this fact about what would happen in the long run to express our confidence in the results of any one sample. We cannot know whether our sample is one of the roughly 95% for which the interval catches \( p \) or one of the unlucky 5% that does not catch \( p \). The statement that we are 95% confident is shorthand for saying, “We arrived at these numbers by a method that gives correct results 95% of the time.”

### Confidence intervals

The interval between the values \( \hat{p} \pm 0.0196 \) is called an approximate 95% confidence interval for \( p \). Like most confidence intervals we will discuss, this one has the form

\[
\text{estimate} \pm \text{margin of error}
\]
6.1 Inference for a Single Proportion

FIGURE 6.7 Twenty-five samples from the same population gave these 95% confidence intervals. In the long run, 95% of all samples give an interval that covers $p$. The sampling distribution of $\hat{p}$ is shown at the top.

The estimate ($\hat{p}$ in this case) is our guess for the value of the unknown parameter based on the sample. The margin of error (0.0196 here) reflects how accurate we believe our guess is, based on the variability of the estimate, and how confident we are that the procedure will catch the true population proportion $p$.

Figure 6.7 illustrates the behavior of 95% confidence intervals in repeated sampling. The center of each interval is at $\hat{p}$ and therefore varies from sample to sample. The approximate sampling distribution of $\hat{p}$ appears at the top of the figure to show the long-term pattern of this variation. The approximate 95% confidence intervals, $\hat{p} \pm 0.0196$, from 25 SRSs of size $n = 2500$ appear below the sampling distribution. The center $\hat{p}$ of each interval is marked by a dot. The arrows on either side of the dot span the confidence interval. All except 1 of the 25 intervals cover the true value of $p$. In a very large number of samples, approximately 95% of the confidence intervals would contain $p$.

Statisticians have constructed confidence intervals for many different parameters based on a variety of designs for data collection. We will meet a number of these in the following chapters. In all these situations, there are two common and important aspects of a confidence interval:

1. It is an interval of the form $(a, b)$, where $a$ and $b$ are numbers computed from the data.

2. It has a property called a confidence level that gives the probability of producing an interval that contains the unknown parameter.

Users can choose the confidence level, but 95% is the standard for most situations. Occasionally, 90% or 99% is used. We will use $C$ to stand for the
confidence level in decimal form. For example, a 95% confidence level corresponds to $C = 0.95$.

**CONFIDENCE INTERVAL**

A level $C$ confidence interval for a parameter is an interval computed from sample data by a method that has probability $C$ of producing an interval containing the true value of the parameter.

**Large-sample confidence interval for a single proportion**

We will now construct a level $C$ confidence interval for the proportion $p$ of a population when the data are an SRS of size $n$ and $n$ is sufficiently large so $\hat{p}$ has approximately the Normal distribution.

First, note that the standard deviation $\sigma_{\hat{p}}$ depends upon the unknown parameter $p$. To estimate this standard deviation using the data, we replace $p$ in the interval formula by the sample proportion $\hat{p}$. We use the term standard error for the standard deviation of a statistic that is estimated from data.

**STANDARD ERROR**

When the standard deviation of a statistic is estimated from the data, the result is called the standard error of the statistic. The standard error of the sample proportion is

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Second, recall that our construction of an approximate 95% confidence interval for the population proportion began by noting that any Normal distribution has probability about 0.95 within ±2 standard deviations of its mean. To construct a level $C$ confidence interval we first catch the central $C$ area under a Normal curve. That is, we must find the number $z^*$ such that any Normal distribution has probability $C$ within ±$z^*$ standard deviations of its mean. Because all Normal distributions have the same standardized form, we can obtain everything we need from the standard Normal curve. Figure 6.8 shows how $C$ and $z^*$ are related. Values of $z^*$ for many choices of $C$ appear in the row labeled $z^*$ at the bottom of Table D. Here are the most important entries from that row:

<table>
<thead>
<tr>
<th>$z^*$</th>
<th>1.645</th>
<th>1.960</th>
<th>2.576</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>90%</td>
<td>95%</td>
<td>99%</td>
</tr>
</tbody>
</table>

For values of $C$ not in Table D, use Table A to find $z^*$.

As Figure 6.8 reminds us, any Normal curve has probability $C$ between the point $z^*$ standard deviations below the mean and the point $z^*$ standard deviations above the mean. The sample proportion $\hat{p}$ has the approximate Normal distribution with mean $p$ and standard deviation $\sigma_{\hat{p}}$, so there is probability $C$
6.1 Inference for a Single Proportion

FIGURE 6.8 To construct a level \( C \) confidence interval, we must find the number \( z^* \). This is how they are related. The area between \(-z^*\) and \( z^* \) under the standard Normal curve is \( C \).

that \( \hat{p} \) lies between

\[ p - z^*\sigma_{\hat{p}} \quad \text{and} \quad p + z^*\sigma_{\hat{p}} \]

This statement about \( \hat{p} \)'s location relative to the unknown population proportion \( p \) is exactly the same as saying that \( p \) lies between

\[ \hat{p} - z^*\sigma_{\hat{p}} \quad \text{and} \quad \hat{p} + z^*\sigma_{\hat{p}} \]

If we replace the standard deviation of \( \hat{p} \) with the standard error, there is approximate probability \( C \) that the interval \( \hat{p} \pm z^*SE_{\hat{p}} \) contains \( p \). This interval is our confidence interval. The estimate of the unknown \( p \) is \( \hat{p} \), and the margin of error is \( z^*SE_{\hat{p}} \).

LARGE-SAMPLE CONFIDENCE INTERVAL FOR A POPULATION PROPORTION

Choose an SRS of size \( n \) from a large population with an unknown proportion \( p \) of successes. The sample proportion is

\[ \hat{p} = \frac{X}{n} \]

where \( X \) is the number of successes. The standard error of \( \hat{p} \) is

\[ SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]

and the margin of error for confidence level \( C \) is

\[ m = z^*SE_{\hat{p}} \]

where the critical value \( z^* \) is the value for the standard Normal density curve with area \( C \) between \(-z^*\) and \( z^* \). An approximate level \( C \) confidence interval for \( p \) is

\[ \hat{p} \pm m \]

Use this interval for 90%, 95%, or 99% confidence when the number of successes and the number of failures are both at least 15.
EXAMPLE 6.7

Inference for adults and video games. The sample survey in Example 6.1 found that 1063 of a sample of 2054 adults reported that they played video games. In that example we calculated \( \hat{p} = 0.5175 \). The standard error is

\[
SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{0.5175(1 - 0.5175)}{2054}} = 0.011026
\]

The critical value for 95% confidence is \( z^* = 1.96 \), so the margin of error is

\[
m = 1.96SE_{\hat{p}} = (1.96)(0.011026) = 0.021610
\]

The confidence interval is

\[
\hat{p} \pm m = 0.52 \pm 0.02
\]

We conclude that we are 95% confident that between 50% and 54% of adults play video games. We do not know if the true parameter \( p \) is 51% or 53%, but we are 95% confident that \( p \) is somewhere between 50% and 54%. Alternatively, we can say that based on this sample, 52% of adults play video games with a 95% margin of error of 2%.

In performing these calculations we have kept a large number of digits for our intermediate calculations. However, when reporting the results we prefer to use rounded values—for example, “52% with a margin of error of 2%.” In this way we focus attention on the important parts that we have found. There is no additional information to be gained by reporting “0.5175 with a margin of error of 0.021610.”

Remember that the margin of error in any confidence interval includes only random sampling error. If people do not respond honestly to the questions asked, for example, your estimate is likely to miss by more than the margin of error.

Because the calculations for statistical inference for a single proportion are relatively straightforward, we often do them with a calculator or in a spreadsheet. Figure 6.9 gives output from Minitab and SAS for the data on adults and video games in Example 6.1 (page xxx). As usual, the output reports more digits than are useful. When you use software, be sure to think about how many digits are meaningful for your purposes. Do not clutter your report with information that is not meaningful. SAS gives the standard error next to the label ASE, which stands for “asymptotic standard error.” The SAS output also includes an alternative interval based on an “exact” method.

We recommend the large-sample confidence interval for 90%, 95%, and 99% confidence whenever the number of successes and the number of failures are both at least 15. For smaller sample sizes, we recommend exact methods that use the binomial distribution. These are available as the default or as options in many statistical software packages.

USE YOUR KNOWLEDGE

6.9 Bank acquisitions. Refer to Exercise 6.3 (page xxx).

(a) Find \( SE_{\hat{p}} \), the standard error of \( \hat{p} \).
6.1 Inference for a Single Proportion

(b) Give the 95% confidence interval for $p$ in the form of estimate plus or minus the margin of error.

(c) Give the confidence interval as an interval of percents.

6.10 How often do they play? Refer to Exercise 6.4 (page xxx).

(a) Find $SE_{\hat{p}}$, the standard error of $\hat{p}$.

(b) Give the 95% confidence interval for $p$ in the form of estimate plus or minus the margin of error.

(c) Give the confidence interval as an interval of percents.

Significance tests

The confidence interval is appropriate when our goal is to estimate population parameters. The second common type of inference is directed at a quite different goal: to assess the evidence provided by the data in favor of some claim about the population parameters.

A significance test is a formal procedure for comparing observed data with a hypothesis whose truth we want to assess. The hypothesis is a statement about the population parameters. The results of a test are expressed in terms of a probability that measures how well the data and the hypothesis agree. We use the following example to illustrate these concepts.
EXAMPLE 6.8

College expectations. Each year the Cooperative Institutional Research Program (CIRP) Freshman Survey is administered to first-time incoming students at hundreds of colleges and universities. In 2011, 203,967 freshmen were polled. Of the respondents, 67.5% thought they had a very good chance of having at least a B average in college. You decide to see if the view of incoming freshmen at your large university is more optimistic than this. You draw an SRS of \(n = 200\) students and find \(X = 147\) of them have this view. Can we conclude from this survey that incoming freshmen at your university have a more optimistic view of their grade point average than the general freshman population?

One way to answer this is to compute the probability of obtaining a sample proportion as large or larger than the observed \(\hat{p} = 147/200 = 73.5\%\) assuming that, in fact, the population proportion at your university is 67.5%. Software tells us that \(P(X \geq 147) = 0.04\). Because this probability is relatively small, we conclude that observing a sample proportion of 73.5% is surprising when the true proportion is 67.5%. The data provide evidence for us to conclude that the incoming freshmen at your university are more optimistic than the overall freshman population.

What are the key steps in this example?

- We started with a question about the population proportion. In this case the population was the incoming freshmen at your university. We wanted to see if this population had a proportion compatible with that of the overall freshman population. In other words, we wanted to compare the university population proportion with 0.675.
- Next we compared the data, \(\hat{p} = 0.735\), with the value that comes from the question, \(p = 0.675\).
- The result of the comparison is the probability 0.04.

The 0.04 probability is relatively small. Something that happens with probability 0.04 occurs 4 times out of 100. In this case we have two possible explanations:

1. We have observed something that is quite unusual, or
2. The assumption that underlies the calculation, \(p = 0.675\), is not true.

Because this probability is small, we prefer the second conclusion: the incoming freshmen at your university are more optimistic than the overall freshman population.

Stating hypotheses

In Example 6.8, we asked whether the observed sample proportion is reasonable if, in fact, the underlying true proportion is 0.675. To answer this, we begin by supposing that the statement following the “if” in the previous sentence is true. In other words, we suppose that the true proportion is 0.675. We then ask
whether the data provide evidence against the supposition we have made. If so, we have evidence in favor of an effect (the proportion is larger) that we are seeking. The first step in a test of significance is to state a claim that we will try to find evidence against.

**NULL HYPOTHESIS**
The statement being tested in a test of significance is called the null hypothesis. The test of significance is designed to assess the strength of the evidence against the null hypothesis. Usually the null hypothesis is a statement of “no effect” or “no difference.”

We abbreviate “null hypothesis” as $H_0$. A null hypothesis is a statement about the population parameters. For example, the null hypothesis for Example 6.8 is

$$H_0: \text{There is no difference between the population proportion at your university and that of the overall freshman population, } p = 0.675.$$  

Note that the null hypothesis refers to the population proportion for all freshmen from your university, including those for whom we do not have data.

It is convenient also to give a name to the statement we hope or suspect is true instead of $H_0$. This is called the alternative hypothesis and is abbreviated as $H_a$. In Example 6.8, the alternative hypothesis states that the population proportion is larger than 0.675 (more optimistic). We write this as

$$H_a: \text{The population proportion at your university is larger, } p > 0.675.$$  

**USE YOUR KNOWLEDGE**

**6.11 Food court survey.** The food court closest to your dormitory has been redesigned. A survey is planned to determine whether or not students think that the new design is an improvement. Sampled students will respond “Yes” if they think it is an improvement and “No” otherwise. The redesign will be considered a success if at least half the students view
the redesign favorably. State the null and alternative hypotheses you would use to examine whether the redesign is a success.

6.12 More on the food court survey. Refer to the previous exercise. Suppose that the food court staff had input into the redesign and expect that 3 of every 4 students will view the redesign favorably. State the null and alternative hypotheses you would use to examine whether or not student opinions are different from staff expectations.

Test statistics

We will learn the form of significance tests in a number of common situations. Here are some principles that apply to most tests and that help in understanding these tests:

- The test is based on a statistic that estimates the parameter that appears in the hypotheses. Usually this is the same estimate we would use in a confidence interval for the parameter. When $H_0$ is true, we expect the estimate to take a value near the parameter value specified by $H_0$.
- Values of the estimate far from the parameter value specified by $H_0$ give evidence against $H_0$. The alternative hypothesis determines which directions count against $H_0$.
- To assess how far the estimate is from the parameter, standardize the estimate. In many common situations the test statistic has the form

$$z = \frac{\text{estimate} - \text{hypothesized value}}{\text{standard deviation of the estimate}}$$

A test statistic measures compatibility between the null hypothesis and the data. We use it for the probability calculation that we need for our test of significance. It is a random variable with a distribution that we know.

In Example 6.8, an SRS of $n = 200$ students resulted in a sample proportion $\hat{p} = 147/200 = 0.735$. We can standardize this estimate based on the Normal approximation (page xxx) to get the observed test statistic. For these data it is

$$z = \frac{0.735 - 0.675}{\sqrt{0.675(1 - 0.675)/200}} = 1.812$$

This means we have observed a sample estimate that is slightly more than 1.8 standard deviations away from the hypothesized value of the parameter. We will now use facts about the Normal distribution to assess how unusual an observation this far away from the hypothesized value is.

P-values

If all test statistics were Normal, we could base our conclusions on the value of the $z$ test statistic. In fact, the Supreme Court of the United States has said that “two or three standard deviations” ($z = 2$ or 3) is its criterion for rejecting $H_0$, and this is the criterion used in most applications involving the law. Because not all test statistics are Normal, we translate the value of test statistics into a common language, the language of probability.
A test of significance finds the probability of getting an outcome as extreme or more extreme than the actually observed outcome. “Extreme” means “far from what we would expect if $H_0$ were true.” The direction or directions that count as “far from what we would expect” are determined by $H_a$ and $H_0$.

**P-VALUE**

The probability, assuming $H_0$ is true, that the test statistic would take a value as extreme or more extreme than that actually observed is called the **P-value** of the test. The smaller the $P$-value, the stronger the evidence against $H_0$ provided by the data.

The key to calculating the $P$-value is the sampling distribution of the test statistic. For the problems we consider in this chapter, we need only the standard Normal distribution for the test statistic $z$.

In Example 6.8, an SRS of $n = 200$ students resulted in $\hat{p} = 147/200 = 0.735$. This sample proportion corresponds to 1.812 standard deviations away from the hypothesized parameter value of 0.675. Because we are using a one-sided alternative and our expectation is for the proportion to be larger than 0.675, the $P$-value is the probability that we observed a $Z$ as extreme or more extreme than 1.812. More formally, this probability is

$$P(Z \geq 1.812)$$

where $Z$ has the standard Normal distribution $N(0, 1)$. Using Table A, this probability is equal to 0.035. This is slightly lower than what was reported in Example 6.8, which used the binomial distribution to compute the $P$-value.

If we had used the two-sided alternative in this example, we would need to consider the probability in both tails of the Normal distribution. In other words, a sample proportion of 0.615, which is just as far away from 0.675 as 0.735, provides just as much evidence against the null hypothesis. The $P$-value would then be

$$P(|Z| \geq 1.812) = 0.070$$

which is exactly twice as large as the one-sided $P$-value.

**USE YOUR KNOWLEDGE**

**6.13 The Normal curve and the $P$-value.** A test statistic for a two-sided significance test for a population proportion is $z = -1.63$. Sketch a standard Normal curve and mark this value of $z$ on it. Find the $P$-value and shade the appropriate areas under the curve to illustrate your calculations.

**6.14 More on the Normal curve and the $P$-value.** A test statistic for a two-sided significance test for a population proportion is $z = 2.42$. Sketch a standard Normal curve and mark this value of $z$ on it. Find the $P$-value and shade the appropriate areas under the curve to illustrate your calculations.
CHAPTER 6 Inference for Proportions

**Statistical significance**

We started our discussion of the reasoning of significance tests with the statement of null and alternative hypotheses. We then learned that a test statistic is the tool used to examine the compatibility of the observed data with the null hypothesis. Finally, we translated the test statistic into a $P$-value to quantify the evidence against $H_0$. One important final step is needed: to state our conclusion.

We can compare the $P$-value we calculated with a fixed value that we regard as decisive. This amounts to announcing in advance how much evidence against $H_0$ we will require to reject $H_0$. The decisive value of $P$ is called the *significance level*. It is commonly denoted by $\alpha$. If we choose $\alpha = 0.05$, we are requiring that the data give evidence against $H_0$ so strong that it would happen no more than 5% of the time (1 time in 20) when $H_0$ is true. If we choose $\alpha = 0.01$, we are insisting on stronger evidence against $H_0$, evidence so strong that it would appear only 1% of the time (1 time in 100) if $H_0$ is in fact true.

**STATISTICAL SIGNIFICANCE**

If the $P$-value is as small or smaller than $\alpha$, we say that the data are *statistically significant at level $\alpha$.*

In Example 6.8, the $P$-value is 0.04 (or 0.035 using the Normal approximation). If we choose $\alpha = 0.05$, we would state that there is enough evidence against $H_0$ to reject this hypothesis and conclude that the proportion at your university is higher than 0.675.

However, if we required stronger evidence and chose $\alpha = 0.01$, we would say that the data do not provide enough evidence to conclude that the proportion at your university is higher than 0.675. This does not mean we conclude that $H_0$ is true—that the proportion at your university is equal to 0.675—only that the level of evidence we required to reject $H_0$ was not met.

Our criminal court system follows a similar procedure in which a defendant is presumed innocent ($H_0$) until proven guilty. If the level of evidence presented is not strong enough for the jury to find the defendant guilty beyond a reasonable doubt, the defendant is acquitted. Acquittal does not imply innocence, only that the degree of evidence was not strong enough to prove guilt.

We will learn the details of many tests of significance in the following chapters. The proper test statistic is determined by the hypotheses and the data collection design. We use computer software or a calculator to find its numerical value and the $P$-value. The computer will not formulate your hypotheses for you, however. Nor will it decide if significance testing is appropriate or help you to interpret the $P$-value that it presents to you. The most difficult and important step is the last one: stating a conclusion.

**Significance test for a single proportion**

Recall that the sample proportion $\hat{p} = X/n$ is approximately Normal, with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. For confidence intervals, we substitute $\hat{p}$ for $p$ in the last expression to obtain the standard error.
When performing a significance test, however, the null hypothesis specifies a value for $p$, and we assume that this is the true value when calculating the $P$-value. Therefore, when we test $H_0: p = p_0$, we substitute $p_0$ into the expression for $\sigma_\hat{p}$ and then standardize $\hat{p}$. Here are the details.

**LARGE-SAMPLE SIGNIFICANCE TEST FOR A POPULATION PROPORTION**

Draw an SRS of size $n$ from a large population with an unknown proportion $p$ of successes. To test the hypothesis $H_0: p = p_0$, compute the $z$ statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

In terms of a standard Normal random variable $Z$, the approximate $P$-value for a test of $H_0$ against

- $H_a: p > p_0$ is $P(Z \geq z)$
- $H_a: p < p_0$ is $P(Z \leq z)$
- $H_a: p \neq p_0$ is $2P(Z \geq |z|)$

Use the large-sample $z$ significance test as long as the expected number of successes, $np_0$, and the expected number of failures, $n(1 - p_0)$, are both greater than 10.

This large-sample $z$ significance test also relies on the Normal approximation, so if our rule of thumb is not met, or the population is less than 10 times as large as the sample, other procedures should be used. Here is a large-sample example.

**EXAMPLE 6.9**

**Comparing two sunblock lotions.** Your company produces a sunblock lotion designed to protect the skin from exposure to both UVA and UVB radiation from the sun. You hire a testing firm to compare your product with the product sold by your major competitor. The testing firm exposes skin on the backs of a sample of 20 people to UVA and UVB rays and measures the protection provided by each product. For 13 of the subjects, your product provided better protection, while for the other 7 subjects, your competitor’s product provided better protection.
Do you have evidence to support a commercial claiming that your product provides superior UVA and UVB protection? To answer this, we first need to state the hypotheses. The parameter of interest is $p$, the proportion of people who would receive superior UVA and UVB protection from your product. If there is no difference between these two lotions, then we'd expect $p = 0.5$. (In other words, which product works better on someone is like flipping a fair coin.) This is our null hypothesis. As for $H_a$, we'll use the two-sided alternative even though your company hopes for $p > 0.5$.

So, to answer this claim, we have $n = 20$ subjects and $X = 13$ successes and want to test

$$H_0: p = 0.5$$
$$H_a: p \neq 0.5$$

The expected numbers of successes (your product provides better protection) and failures (your competitor’s product provides better protection) are $20 \times 0.5 = 10$ and $20 \times 0.5 = 10$. Both are at least 10, so we can use the $z$ test. The sample proportion is

$$\hat{p} = \frac{X}{n} = \frac{13}{20} = 0.65$$

The test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.65 - 0.5}{\sqrt{\frac{0.5(0.5)}{20}}} = 1.34$$

From Table A we find $P(Z \geq 1.34) = 0.9099$, so the probability in the upper tail is $1 - 0.9099 = 0.0901$. The $P$-value is the area in both tails, so $P = 2 \times 0.0901 = 0.1802$. Minitab and SAS outputs for this analysis appear in Figure 6.10. We conclude that the sunblock-testing data are compatible with the hypothesis of no difference between your product and your competitor’s ($\hat{p} = 0.65$, $z = 1.34$, $P = 0.18$). The data do not provide you with a basis to support your advertising claim.

Note that we used a two-sided hypothesis test when we compared the two sunblock lotions in Example 6.9. In settings like this, we must start with the view that either product could be better if we want to prove a claim of superiority. Thinking or hoping that your product is superior cannot be used to justify a one-sided test.

**USE YOUR KNOWLEDGE**

6.15 Draw a picture. Draw a picture of a standard Normal curve and shade the tail areas to illustrate the calculation of the $P$-value for Example 6.9.

6.16 What does the confidence interval tell us? Inspect the outputs in Figure 6.10 and report the confidence interval for the percent of people who would get better sun protection from your product than from your competitor’s. Be sure to convert from proportions to percents and to round appropriately. Interpret the confidence interval. In Chapter 8, we’ll
FIGURE 6.10 Minitab and SAS output for the significance test in Example 6.9.

discuss the relationship between confidence intervals and two-sided tests in more detail.

6.17 The effect of $X$. In Example 6.9, suppose that your product provided better UVA and UVB protection for 15 of the 20 subjects. Perform the significance test using these results and summarize the results.

6.18 The effect of $n$. In Example 6.9, consider what would have happened if you had paid for twice as many subjects to be tested. Assume that the results would be similar to those in your test—that is 65% of the subjects had better UVA and UVB protection with your product. Perform the significance test and summarize the results.

In Example 6.9, we treated an outcome as a success whenever your product provided better sun protection. Would we get the same results if we defined success as an outcome where your competitor’s product was superior? In this setting the null hypothesis is still $H_0: p = 0.5$. You will find that the $z$ test statistic is unchanged except for its sign and that the $P$-value remains the same.
USE YOUR KNOWLEDGE

6.19 Yes or no? In Example 6.9 we performed a significance test to compare your product with your competitor’s. Success was defined as the outcome where your product provided better protection. Now, take the viewpoint of your competitor where success is defined to be the outcome where your competitor’s product provides better protection. In other words, \( n \) remains the same (20) but \( X \) is now 7.

(a) Perform the two-sided significance test to verify that the \( P \)-value remains the same.

(b) Find the 95% confidence interval for this setting and compare it with the interval calculated where success is defined as the outcome in which your product provides better protection.

We do not often use significance tests for a single proportion, because it is uncommon to have a situation where there is a precise \( p_0 \) that we want to test. For physical experiments such as coin tossing or drawing cards from a well-shuffled deck, probability arguments lead to an ideal \( p_0 \). Even here, however, it can be argued, for example, that no real coin has a probability of heads exactly equal to 0.5. Data from past large samples can sometimes provide a \( p_0 \) for the null hypothesis of a significance test. In some types of epidemiology research, for example, “historical controls” from past studies serve as the benchmark for evaluating new treatments. Medical researchers argue about the validity of these approaches, because the past never quite resembles the present. In general, we prefer comparative studies whenever possible.

SECTION 6.1 SUMMARY

A number that describes a population is a parameter. A number that can be computed from the data is a statistic. The purpose of sampling or experimentation is usually inference: using sample statistics to make statements about unknown population parameters.

A statistic from a probability sample or randomized experiment has a sampling distribution that describes how the statistic varies in repeated data production. The sampling distribution answers the question “What would happen if we repeated the sample or experiment many times?” Formal statistical inference is based on the sampling distributions of statistics.

Inference about a population proportion \( p \) from an SRS of size \( n \) is based on the sample proportion \( \hat{p} = X/n \). When \( n \) is large, \( \hat{p} \) has approximately the Normal distribution with mean \( p \) and standard deviation \( \sqrt{p(1-p)/n} \).

The purpose of a confidence interval is to estimate an unknown parameter with an indication of how accurate the estimate is and of how confident we are that the result is correct.

Any confidence interval has two parts: an interval computed from the data and a confidence level. The interval often has the form

\[
\text{estimate} \pm \text{margin of error}
\]

The confidence level states the probability that the method will give a correct answer. That is, if you use 95% confidence intervals, in the long run 95% of your intervals will contain the true parameter value. When you apply the method once, you do not know whether your interval gave a correct value (this happens 95% of the time) or not (this happens 5% of the time).

For large samples, the margin of error for confidence level \( C \) of a proportion is

\[
m = z^* \text{SE}_{\hat{p}}
\]

where the critical value \( z^* \) is the value for the standard Normal density curve with area \( C \) between \(-z^*\) and \( z^*\), and the
The level of failures are both at least 15.

A test of significance is intended to assess the evidence provided by data against a null hypothesis $H_0$ in favor of an alternative hypothesis $H_a$. The hypotheses are stated in terms of population parameters. Usually $H_0$ is a statement that no effect or no difference is present, and $H_a$ says that there is an effect or difference, in a specific direction (one-sided alternative) or in either direction (two-sided alternative).

A student project used a confidence interval to describe the results in a final report. The confidence level was 110%.

6.20. What’s wrong? Explain what is wrong with each of the following:

(a) You can use a significance test to evaluate the hypothesis $H_0: \hat{p} = 0.6$ versus the two-sided alternative.

(b) The large-sample significance test for a population proportion is based on the binomial distribution.

(c) An approximate 95% confidence interval for an unknown proportion $p$ is $\hat{p}$ plus or minus its standard error.

6.21. What’s wrong? Explain what is wrong with each of the following:

(a) The margin of error for a confidence interval used for an opinion poll takes into account the fact that people who did not answer the poll questions may have had different responses from those who did answer the questions.

(b) If the $P$-value for a significance test is 0.35, we can conclude that the null hypothesis has a 35% chance of being true.

(c) A student project used a confidence interval to describe the results in a final report. The confidence level was 110%.

6.22. Draw some pictures. Consider the binomial setting with $n = 50$ and $p = 0.4$.

(a) The sample proportion $\hat{p}$ will have a distribution that is approximately Normal. Give the mean and the standard deviation of this Normal distribution.

(b) Draw a sketch of this Normal distribution. Mark the location of the mean.

(c) Find a value $p^*$ for which the probability is 95% that $\hat{p}$ will be between $\pm p^*$. Mark these two values on your sketch.

6.23. “Country food” and Inuits. “Country food” for Inuits includes seal, caribou, whale, ducks, fish, and berries and is an important part of the diet of the aboriginal people called Inuits who inhabit Inuit Nunaat, the northern region of what is now called Canada. A survey of Inuits in Inuit Nunaat reported that 3274 out of 5000 respondents said that at least half of the meat and fish that they eat is country food. Find the sample proportion and a 95% confidence interval for the population proportion of Inuits whose meat and fish consumption consists of at least half country food.
6.24. Most desirable mates. A poll of 5000 residents in Brazil, Canada, China, France, Malaysia, South Africa, and the United States asked about what profession they would prefer a marriage partner to have. The choice receiving the highest percent, 16% of the responses, was doctors, nurses, and other health care professionals.\(^{10}\)

(a) Find the sample proportion and a 95% confidence interval for the proportion of people who would prefer a doctor, nurse, or other health care professional as a marriage partner.

(b) Convert the estimate and the confidence interval to percents.

6.25. Guitar Hero and Rock Band. An electronic survey of size 7061 reported that 67% of players of Guitar Hero and Rock Band who do not currently play a musical instrument said that they are likely to begin playing a real musical instrument in the next two years. The reports describing the survey do not give the number of respondents who do not currently play a musical instrument.\(^{11}\) The reports describing the survey do not give the number of respondents who do not currently play a musical instrument.

(a) Explain why it is important to know the number of respondents who do not currently play a musical instrument.

(b) Assume that half of the respondents do not currently play a musical instrument. Find the count of players who said that they are likely to begin playing a real musical instrument in the next two years.

(c) Give a 99% confidence interval for the population proportion who would say that they are likely to begin playing a real musical instrument in the next two years.

(d) The survey collected data from two separate consumer panels. There were 3300 respondents from the LightSpeed consumer panel, and the others were from Guitar Center’s proprietary consumer panel. Comment on the sampling procedure used for this survey and how it would influence your interpretation of the findings.


(a) How would the result that you reported in part (c) change if only 25% of the respondents said that they did not currently play a musical instrument?

(b) Do the same calculations if the percent was 75%.

(c) The main conclusion of the survey that appeared in many news stories was that 67% of players of Guitar Hero and Rock Band who do not currently play a musical instrument said that they are likely to begin playing a real musical instrument in the next two years. What can you conclude about the effect of the three scenarios (part (b) in the previous exercise and parts (a) and (b) in this exercise) on the margin of error for the main result?

6.27. Are seniors prepared for class? The National Survey of Student Engagement found that 24% of seniors report that they often or very often went to class without completing readings or assignments.\(^{12}\) Assume that the sample size of seniors is 276,000.

(a) Find the margin of error for 99% confidence.

(b) Here are some items from the report that summarizes the survey. More than 537,000 students from 751 institutions in the United States and Canada participated. The average response rate was 33% and ranged from 15% to 87%. Institutions pay a participation fee of between $1800 and $7800 based on the size of their undergraduate enrollment. Discuss these as sources of error in this study. How do you think these errors would compare with the error that you calculated in part (a)?

6.28. Confidence level and interval width. Refer to Exercise 6.27. Would a 90% confidence interval be wider or narrower than the one that you found in that exercise? Verify your answer by computing the interval.

6.29. Can we use the \( z \) test? In each of the following cases state whether or not the Normal approximation to the binomial should be used for a significance test on the population proportion \( p \).

(a) \( n = 30 \) and \( H_0: p = 0.2 \).

(b) \( n = 30 \) and \( H_0: p = 0.6 \).

(c) \( n = 100 \) and \( H_0: p = 0.5 \).

(d) \( n = 200 \) and \( H_0: p = 0.01 \).

6.30. Instant versus fresh-brewed coffee. A matched pairs experiment compares the taste of instant versus fresh-brewed coffee. Each subject tastes two unmarked cups of coffee, one of each type, in random order and states which he or she prefers. Of the 40 subjects who
participate in the study, 12 prefer the instant coffee. Let \( p \) be the probability that a randomly chosen subject prefers fresh-brewed coffee to instant coffee. (In practical terms, \( p \) is the proportion of the population who prefer fresh-brewed coffee.)

(a) Test the claim that a majority of people prefer the taste of fresh-brewed coffee. Report the large-sample \( z \) statistic and its \( P \)-value.

(b) Draw a sketch of a standard Normal curve and mark the location of your \( z \) statistic. Shade the appropriate area that corresponds to the \( P \)-value.

(c) Is your result significant at the 5% level? What is your practical conclusion?

6.31. Long sermons. The National Congregations Study collected data in a one-hour interview with a key informant—that is, a minister, priest, rabbi, or other staff person or leader.\(^\text{13} \) One question asked concerned the length of the typical sermon. For this question 390 out of 1191 congregations reported that the typical sermon lasted more than 30 minutes.

(a) Use the large-sample inference procedures to construct a 95% confidence interval for the true proportion of congregations in which the typical sermon lasts more than 30 minutes.

(b) The respondents to this question were not asked to use a stopwatch to record the lengths of a random sample of sermons at their congregations. They responded based on their impressions of the sermons. Do you think that ministers, priests, rabbis, or other staff persons or leaders might perceive sermon lengths differently from the people listening to the sermons? Discuss how your ideas would influence your interpretation of the results of this study.

6.32. Do you enjoy driving your car? The Pew Research Center polled \( n = 1048 \) U.S. drivers and found that 69% enjoyed driving their automobiles.\(^\text{14} \)

(a) Construct a 95% confidence interval for the proportion of U.S. drivers who enjoy driving their automobiles.

(b) In 1991, a Gallup Poll reported this percent to be 79%. Using the data from this poll, test the claim that the percent of drivers who enjoy driving their cars has declined since 1991. Report the large-sample \( z \) statistic and its \( P \)-value.

6.33. Getting angry at other drivers. Refer to Exercise 6.32. The same Pew Poll found that 38% of the respondents “shouted, cursed or made gestures to other drivers” in the last year.

(a) Construct a 95% confidence interval for the true proportion of U.S. drivers who did these actions in the last year.

(b) Does the fact that the respondent is self-reporting these actions affect the way that you interpret the results? Write a short paragraph explaining your answer.

6.34. Cheating during a test. A national survey of high school students conducted by the Josephson Institute of Ethics was sent to 43,000 high school students, and 40,774 were returned. One question asked students if they had cheated during a test in the last school year.\(^\text{15} \) Of those who returned the survey, 14,028 responded that they had cheated at least two times in the last year.

(a) What is the sample proportion of respondents who cheated at least twice?

(b) Compute the 95% confidence interval for the true proportion of students who have cheated on at least two tests in the last year.

(c) Compute the nonresponse rate for this study. Does this influence how you interpret these results? Write a short discussion of this issue.

6.35. Pet ownership among older adults. In a study of the relationship between pet ownership and physical activity in older adults,\(^\text{16} \) 594 subjects reported that they owned a pet, while 1939 reported that they did not. Give a 95% confidence interval for the proportion of older adults in this population who are pet owners.

6.36. Annual income of older adults. In the study described in the previous exercise, 1434 subjects out of a total of 2533 reported that their annual income was $25,000 or more.

(a) Give a 95% confidence interval for the true proportion of subjects in this population with incomes of at least $25,000.

(b) Do you think that some respondents might not give truthful answers to a question about their income? Discuss the possible effects on your estimate and confidence interval.

(c) In the previous exercise, the question analyzed concerned pet ownership. Compare this question with the income question with respect to the possibility that the respondents were not truthful.

6.37. Dogs sniffing out cancer. Can dogs detect lung cancer by sniffing exhaled breath samples? In one study, researchers performed 125 trials.\(^\text{17} \) In each trial, a sniffer dog smelled five breath samples, consisting of four control
samples and one cancer sample. A correct response involved the dog lying down next to the cancer sample. Collectively, the dogs correctly identified the cancer sample in 110 of these trials. Construct a 95% confidence interval for the true proportion of times these dogs will correctly identify a lung cancer sample.

6.38. Bicycle accidents and alcohol. In the United States approximately 900 people die in bicycle accidents each year. One study examined the records of 1711 bicyclists aged 15 or older who were fatally injured in bicycle accidents between 1987 and 1991 and were tested for alcohol. Of these, 542 tested positive for alcohol (blood alcohol concentration of 0.01% or higher).18

(a) Summarize the data with appropriate descriptive statistics.

(b) To do statistical inference for these data, we think in terms of a model where \( p \) is a parameter that represents the probability that a tested bicycle rider is positive for alcohol. Find a 99% confidence interval for \( p \).

(c) Can you conclude from your analysis of this study that alcohol causes fatal bicycle accidents? Explain.

(d) In this study 386 bicyclists had blood alcohol levels above 0.10%, a level defining legally drunk in many states at the time. Give a 99% confidence interval for the proportion who were legally drunk according to this criterion.

6.2 Comparing Two Proportions

Because comparative studies are so common, we often want to compare the proportions of two groups (such as men and women) that have some characteristic. In the previous section, we learned how to estimate a single proportion. Our problem now concerns the comparison of two proportions.

We call the two groups being compared Population 1 and Population 2 and the two population proportions of “successes” \( p_1 \) and \( p_2 \). The data consist of two independent SRSs, of size \( n_1 \) from Population 1 and size \( n_2 \) from Population 2. The proportion of successes in each sample estimates the corresponding population proportion. Here is the notation we will use in this section:

<table>
<thead>
<tr>
<th>Population</th>
<th>Population proportion</th>
<th>Sample size</th>
<th>Count of successes</th>
<th>Sample proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( p_1 )</td>
<td>( n_1 )</td>
<td>( X_1 )</td>
<td>( \hat{p}_1 = X_1/n_1 )</td>
</tr>
<tr>
<td>2</td>
<td>( p_2 )</td>
<td>( n_2 )</td>
<td>( X_2 )</td>
<td>( \hat{p}_2 = X_2/n_2 )</td>
</tr>
</tbody>
</table>

To compare the two populations, we use the difference between the two sample proportions:

\[
D = \hat{p}_1 - \hat{p}_2
\]

Normal approximation for the difference between two proportions

Inference procedures for comparing proportions are \( z \) procedures based on the Normal approximation and on standardizing the difference \( D \). The first step is to obtain the mean and standard deviation of \( D \). By the addition rule for means, the mean of \( D \) is the difference between the means:

\[
\mu_D = \mu_{\hat{p}_1} - \mu_{\hat{p}_2} = p_1 - p_2
\]

That is, the difference \( D = \hat{p}_1 - \hat{p}_2 \) between the sample proportions is an unbiased estimator of the population difference \( p_1 - p_2 \). Similarly, the addition
rule for variances tells us that the variance of $D$ is the sum of the variances:

$$
\sigma_D^2 = \sigma_1^2 + \sigma_2^2
$$

$$
= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}
$$

Therefore, when $n_1$ and $n_2$ are large, $D$ is approximately Normal with mean $\mu_D = p_1 - p_2$ and standard deviation

$$
\sigma_D = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}
$$

---

**USE YOUR KNOWLEDGE**

**6.39 Rules for means and variances.** Suppose that $p_1 = 0.4$, $n_1 = 25$, $p_2 = 0.5$, $n_2 = 30$. Find the mean and the standard deviation of the sampling distribution of $p_1 - p_2$.

**6.40 Effect of the sample sizes.** Suppose that $p_1 = 0.4$, $n_1 = 100$, $p_2 = 0.5$, $n_2 = 120$.

(a) Find the mean and the standard deviation of the sampling distribution of $p_1 - p_2$.

(b) The sample sizes here are four times as large as those in the previous exercise, while the population proportions are the same. Compare the results for this exercise with those that you found in the previous exercise. What is the effect of multiplying the sample sizes by 4?

**6.41 Rules for means and variances.** It is quite easy to verify the formulas for the mean and standard deviation of the difference $D$.

(a) What are the means and standard deviations of the two sample proportions $\hat{p}_1$ and $\hat{p}_2$?

(b) Use the addition rule for means of random variables: what is the mean of $D = \hat{p}_1 - \hat{p}_2$?

(c) The two samples are independent. Use the addition rule for variances of random variables: what is the variance of $D$?

---

**Large-sample confidence interval for the difference between two proportions**

To obtain a confidence interval for $p_1 - p_2$, we once again replace the unknown parameters in the standard deviation by estimates to obtain an estimated standard deviation, or standard error. Here is the confidence interval we want.
LARGE-SAMPLE CONFIDENCE INTERVAL FOR COMPARING TWO PROPORTIONS

Choose an SRS of size \( n_1 \) from a large population having proportion \( p_1 \) of successes and an independent SRS of size \( n_2 \) from another population having proportion \( p_2 \) of successes. The estimate of the difference between the population proportions is

\[
D = \hat{p}_1 - \hat{p}_2
\]

The standard error of \( D \) is

\[
SE_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}
\]

and the margin of error for confidence level \( C \) is

\[
m = z^* SE_D
\]

where the critical value \( z^* \) is the value for the standard Normal density curve with area \( C \) between \(-z^*\) and \( z^*\). An approximate level \( C \) confidence interval for \( p_1 - p_2 \) is

\[
D \pm m
\]

Use this method for 90%, 95%, or 99% confidence when the number of successes and the number of failures in each sample are at least 10.

EXAMPLE 6.10

**Gender and the proportion of frequent binge drinkers.** Many studies have documented binge drinking as a major problem among college students.\(^{19}\) Here are some data that let us compare men and women:

<table>
<thead>
<tr>
<th>Population</th>
<th>( n )</th>
<th>( X )</th>
<th>( \hat{p} = X/n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (men)</td>
<td>5,348</td>
<td>1,392</td>
<td>0.260</td>
</tr>
<tr>
<td>2 (women)</td>
<td>8,471</td>
<td>1,748</td>
<td>0.206</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>13,819</td>
<td>3,140</td>
<td>0.227</td>
</tr>
</tbody>
</table>

In this table the \( \hat{p} \) column gives the sample proportions of frequent binge drinkers.

Let’s find a 95% confidence interval for the difference between the proportions of men and of women who are frequent binge drinkers. Output from Minitab and CrunchIt! is given in Figure 6.11. To perform the computations using our formulas, we first find the difference in the proportions:

\[
D = \hat{p}_1 - \hat{p}_2
\]

\[
= 0.260 - 0.206
\]

\[
= 0.054
\]
Then we calculate the standard error of $D$:

$$SE_D = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

$$= \sqrt{\frac{(0.260)(0.740)}{5348} + \frac{(0.206)(0.794)}{8471}}$$

$$= 0.00744$$

For 95% confidence, we have $z^* = 1.96$, so the margin of error is

$$m = z^*SE_D$$

$$= (1.96)(0.00744)$$

$$= 0.015$$

The 95% confidence interval is

$$D \pm m = 0.054 \pm 0.015$$

$$= (0.039, 0.069)$$

With 95% confidence we can say that the difference in the proportions is between 0.039 and 0.069. Alternatively, we can report that the difference in the percent of men who are frequent binge drinkers and the percent of women who are frequent binge drinkers is 5.4%, with a 95% margin of error of 1.5%.
In this example men and women were not sampled separately. The sample sizes are in fact random and reflect the gender distributions of the colleges that were randomly chosen. Two-sample significance tests and confidence intervals are still approximately correct in this situation. The authors of the report note that women are overrepresented partly because 6 of the 140 colleges in the study were women’s colleges.

In the example above we chose men to be the first population. Had we chosen women to be the first population, the estimate of the difference would be negative (−0.054). Because it is easier to discuss positive numbers, we generally choose the population with the higher proportion to be the first population.

6.42 Gender and commercial preference. A study was designed to compare two energy drink commercials. Each participant was shown the commercials in random order and asked to select the better one. Commercial A was selected by 44 out of 100 women and 79 out of 140 men. Give an estimate of the difference in gender proportions that favored Commercial A. Also construct a large-sample 95% confidence interval for this difference.

6.43 Gender and commercial preference, revisited. Refer to Exercise 6.42. Construct a 95% confidence interval for the difference in gender proportions that favor Commercial B. Explain how you could have obtained these results from the calculations you did in Exercise 6.42.

Significance test for a difference in proportions

Although we prefer to compare two proportions by giving a confidence interval for the difference between the two population proportions, it is sometimes useful to test the null hypothesis that the two population proportions are the same.

We standardize \( D = \hat{p}_1 - \hat{p}_2 \) by subtracting its mean \( p_1 - p_2 \) and then dividing by its standard deviation

\[
\sigma_D = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}
\]

If \( n_1 \) and \( n_2 \) are large, the standardized difference is approximately \( N(0, 1) \). For the large-sample confidence interval, we used sample estimates in place of the unknown population values in the expression for \( \sigma_D \). Although this approach would lead to a valid significance test, we instead adopt the more common practice of replacing the unknown \( \sigma_D \) with an estimate that takes into account our null hypothesis \( H_0: p_1 = p_2 \). If these two proportions are equal, then we can view all the data as coming from a single population. Let \( p \) denote the common value of \( p_1 \) and \( p_2 \). Then the standard deviation of \( D = \hat{p}_1 - \hat{p}_2 \) is

\[
\sigma_D = \sqrt{\frac{p(1 - p)}{n_1} + \frac{p(1 - p)}{n_2}} = \sqrt{p(1 - p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}
\]
We estimate the common value of $p$ by the overall proportion of successes in the two samples:

$$\hat{p} = \frac{\text{number of successes in both samples}}{\text{number of observations in both samples}} = \frac{X_1 + X_2}{n_1 + n_2}$$

This estimate of $p$ is called the **pooled estimate** because it combines, or pools, the information from both samples.

To estimate $\sigma_D$ under the null hypothesis, we substitute $\hat{p}$ for $p$ in the expression for $\sigma_D$. The result is a standard error for $D$ that assumes $H_0: p_1 = p_2$:

$$SE_{dp} = \sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

The subscript on $SE_{dp}$ reminds us that we pooled data from the two samples to construct the estimate.

**SIGNIFICANCE TEST FOR COMPARING TWO PROPORTIONS**

To test the hypothesis

$$H_0: p_1 = p_2$$

compute the $z$ **statistic**

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE_{dp}}$$

where the **pooled standard error** is

$$SE_{dp} = \sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

and where

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

In terms of a standard Normal random variable $Z$, the $P$-value for a test of $H_0$ against

- $H_0: p_1 > p_2$ is $P(Z \geq z)$
- $H_0: p_1 < p_2$ is $P(Z \leq z)$
- $H_0: p_1 \neq p_2$ is $2P(Z \geq |z|)$

As a general rule, we will use this test when the number of successes and the number of failures in each of the samples are at least 5.
CHAPTER 6 Inference for Proportions

This \( z \) test has the general form given on page xxx and is based on the Normal approximation to the binomial distribution.

**EXAMPLE 6.11**

**Gender and the proportion of frequent binge drinkers: the \( z \) test.** Are men and women college students equally likely to be frequent binge drinkers? We examine the survey data in Example 6.10 (page xxx) to answer this question. Here is the data summary:

<table>
<thead>
<tr>
<th>Population</th>
<th>( n )</th>
<th>( X )</th>
<th>( \hat{p} = X/n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (men)</td>
<td>5,348</td>
<td>1,392</td>
<td>0.260</td>
</tr>
<tr>
<td>2 (women)</td>
<td>8,471</td>
<td>1,748</td>
<td>0.206</td>
</tr>
<tr>
<td>Total</td>
<td>13,819</td>
<td>3,140</td>
<td>0.227</td>
</tr>
</tbody>
</table>

The sample proportions are certainly quite different, but we will perform a significance test to see if the difference is large enough to lead us to believe that the population proportions are not equal. Formally, we test the hypotheses

\[
H_0: p_1 = p_2 \\
H_a: p_1 \neq p_2
\]

The pooled estimate of the common value of \( p \) is

\[
\hat{p} = \frac{1392 + 1748}{5348 + 8471} = \frac{3140}{13,819} = 0.227
\]

Note that this is the estimate on the bottom line of the data summary.

The test statistic is calculated as follows:

\[
SE_{\hat{p}} = \sqrt{(0.227)(0.773)\left(\frac{1}{5348} + \frac{1}{8471}\right)} = 0.007316
\]

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{SE_{\hat{p}}} = \frac{0.260 - 0.206}{0.007316} = 7.37
\]

The \( P \)-value is \( 2P(Z \geq 7.37) \). The largest value of \( z \) in Table A is 3.49, so from this table we can conclude \( P < 2 \times 0.0002 = 0.0004 \). Most software reports this result as 0 or a very small number. Output from Minitab and CrunchIt! is given in Figure 6.12. Minitab reports the \( P \)-value as 0.000. This means that the calculated value is less than 0.0005; this is certainly a very small number. CrunchIt! gives \( < 0.0001 \).

The exact value is not particularly important. It is clear that we should reject the null hypothesis because the chance of getting these sample results is so small if the null hypothesis is true. For most situations, 0.001 (1 chance in 1000) is sufficiently small. We report: among college students in the study, 26.0% of the men and 20.6% of the women were frequent binge drinkers. The difference is statistically significant (\( z = 7.37, P < 0.001 \)).
We could have argued that we expect the proportion to be higher for men than for women in this example. This would justify using the one-sided alternative $H_a: p_1 > p_2$. The $P$-value would be half of the value obtained for the two-sided test. Because the $z$ statistic is so large, this distinction is of no practical importance.

**USE YOUR KNOWLEDGE**

**6.44 Gender and commercial preference: the $z$ test.** Refer to Exercise 6.42 (page xxx). Test that the proportions of women and men who liked Commercial A are the same versus the two-sided alternative at the 5% level.

**6.45 Changing the alternative hypothesis.** Refer to the previous exercise. Does your conclusion change if you test whether the proportion of men who favor Commercial A is larger than the proportion of females? Explain.
SECTION 6.2 SUMMARY

The large-sample estimate of the difference in two population proportions is
\[ D = \hat{p}_1 - \hat{p}_2 \]
where \( \hat{p}_1 \) and \( \hat{p}_2 \) are the sample proportions
\[ \hat{p}_1 = \frac{X_1}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{X_2}{n_2} \]
The standard error of the difference \( D \) is
\[ SE_D = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \]
The margin of error for confidence level \( C \) is
\[ m = z^* SE_D \]
where \( z^* \) is the value for the standard Normal density curve with area \( C \) between \( -z^* \) and \( z^* \). The large-sample level \( C \) confidence interval is
\[ D \pm m \]

SECTION 6.2 EXERCISES

For Exercises 6.39 to 6.41, see pages xxx-xxx; for Exercises 6.42 and 6.43, see page xxx; and for Exercises 6.44 and 6.45, see page xxx.

6.46. Draw a picture. Suppose that there are two binomial populations. For the first, the true proportion of successes is 0.4; for the second, it is 0.5. Consider taking independent samples from these populations, 50 from the first and 60 from the second.

(a) Find the mean and the standard deviation of the distribution of \( \hat{p}_1 - \hat{p}_2 \).

(b) This distribution is approximately Normal. Sketch this Normal distribution and mark the location of the mean.

(c) Find a value \( d \) for which the probability is 0.95 that the difference in sample proportions is within \( \pm d \). Mark these values on your sketch.

6.47. What’s wrong? For each of the following, explain what is wrong and why.

(a) A \( z \) statistic is used to test the null hypothesis that \( \hat{p}_1 = \hat{p}_2 \).

(b) If two sample proportions are equal, then the sample counts are equal.

(c) A 95% confidence interval for the difference in two proportions includes errors due to nonresponse.

6.48. Podcast downloading. The Podcast Alley Web site recently reported that they have 91,701 podcasts available for downloading, with 6,070,164 episodes. The Pew Research Center performed two surveys about podcast downloading. The first was conducted between February and April 2006 and surveyed 2822 Internet users. It found that 198 of these said that they had downloaded a podcast to listen to it or view it later at least once. In a more recent survey, conducted in May 2008, there were 1553 Internet users. Of this total, 295 said that they had downloaded a podcast to listen to it or view it later.

(a) Refer to the table that appears at the beginning of this section (page xxx). Fill in the numerical values of all quantities that are known.

(b) Find the estimate of the difference between the proportion of Internet users who had downloaded podcasts as of February to April 2006 and the proportion as of May 2008.

(c) Is the large-sample confidence interval for the difference in two proportions appropriate to use in this setting? Explain your answer.
(d) Find the 95% confidence interval for the difference.
(e) Convert your estimated difference and confidence interval to percents.
(f) One of the surveys was conducted between February and April, whereas the other was conducted in May. Do you think that this difference should have any effect on the interpretation of the results? Be sure to explain your answer.

6.49. Significance test for podcast downloading. Refer to the previous exercise. Test the null hypothesis that the two proportions are equal versus the two-sided alternative. Report the test statistic with the \( P \)-value and interpret your results.

6.50. Are more Internet users downloading podcasts? Refer to the previous two exercises. The ratio of the proportion in the 2008 sample to the proportion in the 2006 sample is about 2.7.
(a) Can you conclude that 2.7 times as many people are downloading podcasts? Explain why or why not.
(b) Can you conclude from the data available that there has been an increase from 2006 to 2008 in the number of people who download podcasts? If your answer is no, explain what additional data you would need or what additional assumptions you would need to be able to draw this conclusion.

6.51. Adult gamers versus teen gamers. A Pew Internet Project Data Memo presented data comparing adult gamers with teen gamers with respect to the devices on which they play. The data are from two surveys. The adult survey had 1063 gamers, while the teen survey had 1064 gamers. The memo reports that 54% of adult gamers played on game consoles (Xbox, PlayStation, Wii, etc.), while 89% of teen gamers played on game consoles.\(^{22}\)
(a) Refer to the table that appears at the beginning of this section (page xxx). Fill in the numerical values of all quantities that are known.
(b) Find the estimate of the difference between the proportion of teen gamers who played on game consoles and the proportion of adults who played on these devices.
(c) Is the large-sample confidence interval for the difference in two proportions appropriate to use in this setting? Explain your answer.
(d) Find the 95% confidence interval for the difference.
(e) Convert your estimated difference and confidence interval to percents.
(f) The adult survey was conducted between October and December 2008, whereas the teen survey was conducted between November 2007 and February 2008. Do you think that this difference should have any effect on the interpretation of the results? Be sure to explain your answer.

6.52. Significance test for gaming on consoles. Refer to the previous exercise. Test the null hypothesis that the two proportions are equal versus the two-sided alternative. Report the test statistic with the \( P \)-value and summarize your conclusion.

6.53. Gamers on computers. The report described in Exercise 6.51 also presented data from the same surveys for gaming on computers (desktops or laptops). These devices were used by 73% of adult gamers and by 76% of teen gamers. Answer the questions given in Exercise 6.51 for gaming on computers.

6.54. Significance test for gaming on computers. Refer to the previous two exercises. The ratio of the proportion in the 2008 sample to the proportion in the 2006 sample is about 2.7.
(a) Can you conclude that 2.7 times as many people are gaming on computers? Explain why or why not.
(b) Can you conclude from the data available that there has been an increase from 2006 to 2008 in the number of people who download podcasts? If your answer is no, explain what additional data you would need or what additional assumptions you would need to be able to draw this conclusion.

6.55. Can we compare gaming on consoles with gaming on computers? Refer to the previous four exercises. Do you think that you can use the large-sample confidence intervals for a difference in proportions to compare teens’ use of computers with teens’ use of consoles? Write a short paragraph giving the reason for your answer. \( (\text{Hint: Look carefully in the box on page xxx giving the assumptions needed for this procedure.}) \)

6.56. \( \hat{p}_1 - \hat{p}_2 \) and the Normal distribution. Refer to Exercise 6.46. Assume that all the conditions for that exercise remain the same, with the exception that \( n_2 = 1000 \).
(a) Find the mean and standard deviation of \( \hat{p}_1 - \hat{p}_2 \).
(b) Find the mean and standard deviation of \( \hat{p}_1 - 0.5 \).
(c) Because \( n_2 \) is very large, we expect \( \hat{p}_2 \) to be very close to 0.5. How close?
(d) Summarize what you have found in parts (a), (b), and (c) of this exercise. Interpret your results in terms of inference for comparing two proportions when the size of one of the samples is much larger than the size of the other.

6.57. Peer-to-peer music downloading. The NPD Group reported that the percent of Internet users who download music via peer-to-peer (P2P) services was 9% in late 2010, compared with 16% in late 2007.\(^{23}\) The filing of lawsuits by the recording industry may be a reason why this percent has decreased. Assume that the sample sizes...
are both 5549, the sample size reported for the 2010 survey. Using a significance test, evaluate whether or not there has been a change in the percent of Internet users who download music via P2P services. Provide all details for the test and summarize your conclusion. Also report a 95% confidence interval for the difference in proportions (2007 versus 2010) and explain what information is provided in the interval that is not given in the significance test results.

6.58. More on downloading music via P2P. Refer to the previous exercise. Because we are not exactly sure about the size of the 2007 sample, redo the calculations for the significance test and the confidence interval under the following assumptions, and summarize the effects of the sample sizes on the results.

(a) The first sample size is 500.
(b) The first sample size is 1500.
(c) The first sample size is 3000.

6.59. Gender bias in textbooks. To what extent do syntax textbooks, which analyze the structure of sentences, illustrate gender bias? A study of this question sampled sentences from 10 texts. One part of the study examined the use of the words “girl,” “boy,” “man,” and “woman.” We will call the first two words juvenile and the last two adult. Is the proportion of female references that are juvenile (girl) equal to the proportion of male references that are juvenile (boy)? Here are data from one of the texts:

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>X (juvenile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>60</td>
<td>48</td>
</tr>
<tr>
<td>Male</td>
<td>132</td>
<td>52</td>
</tr>
</tbody>
</table>

(a) Find the proportion of juvenile references for females and its standard error. Do the same for the males.
(b) Give a 90% confidence interval for the difference and briefly summarize what the data show.
(c) Use a test of significance to examine whether the two proportions are equal.

6.60. Cheating during a test: 2002 versus 2010. In Exercise 6.34 (page xxx), you examined the proportion of high school students who cheated on tests at least twice during the past year. Also available are results from other years. A reported 14,028 out of 40,774 students said they cheated at least twice in 2010. A reported 5794 out of 12,121 students said they cheated at least twice in 2002. Give an estimate of the difference between these two proportions with a 90% confidence interval.

6.61. Bicycle accidents, alcohol, and gender. In Exercise 6.38 (page xxx) we examined the percent of fatally injured bicyclists tested for alcohol who tested positive. Here we examine the same data with respect to gender.

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>X (tested positive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>191</td>
<td>27</td>
</tr>
<tr>
<td>Male</td>
<td>1520</td>
<td>515</td>
</tr>
</tbody>
</table>

(a) Summarize the data by giving the estimates of the two population proportions and a 95% confidence interval for their difference. Briefly summarize what the data show.
(b) Use a test of significance to examine whether the two proportions are equal.

6.62. Video game genres. U.S. computer and video game software sales were $25.1 billion in 2010. A survey of 1102 teens collected data about their video game use. According to the survey, the most popular game genres are:

<table>
<thead>
<tr>
<th>Genre</th>
<th>Examples</th>
<th>Percent who play</th>
</tr>
</thead>
<tbody>
<tr>
<td>Racing</td>
<td>NASCAR, Mario Kart, Burnout</td>
<td>74</td>
</tr>
<tr>
<td>Puzzle</td>
<td>Bejeweled, Tetris, Solitaire</td>
<td>72</td>
</tr>
<tr>
<td>Sports</td>
<td>Madden, FIFA, Tony Hawk</td>
<td>68</td>
</tr>
<tr>
<td>Action</td>
<td>Grand Theft Auto, Devil</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>May Cry, Ratchet and Clank</td>
<td></td>
</tr>
<tr>
<td>Adventure</td>
<td>Legend of Zelda, Tomb Raider</td>
<td>66</td>
</tr>
<tr>
<td>Rhythm</td>
<td>Guitar Hero, Dance Dance</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Revolution, Lumines</td>
<td></td>
</tr>
</tbody>
</table>
Give a 95% confidence interval for the proportion who play games in each of these six genres.

6.63. **Too many errors.** Refer to the previous exercise. The chance that each of the six intervals that you calculated includes the true proportion for that genre is approximately 95%. In other words, the chance that you make an error and your interval misses the true value is approximately 5%.

(a) Explain why the chance that at least one of your intervals does not contain the true value of the parameter is greater than 5%.

(b) One way to deal with this problem is to adjust the confidence level for each interval so that the overall probability of at least one miss is 5%. One simple way to do this is to use a Bonferroni procedure. Here is the basic idea: You have an error budget of 5% and you choose to spend it equally on six intervals. Each interval has a budget of $0.05/6 = 0.0083$. So each confidence interval should have a 0.83% chance of missing the true value. In other words, the confidence level for each interval should be $1 - 0.0083 = 0.9917$.

(c) Use Table A to find the value of $z$ for a large-sample confidence interval for a single proportion corresponding to 99.17% confidence.

(d) Calculate the six confidence intervals using the Bonferroni procedure.

6.64. **Wireless only.** Are customers giving up their landlines and relying on wireless for all of their phone needs? Surveys have collected data to answer this question. In June 2006, 10.5% of households were wireless only. Assume that this survey is based on sampling 15,000 households.

(a) Convert the percent to a proportion. Then use the proportion and the sample size to find the count of households who were wireless only.

(b) Find a 95% confidence interval for the proportion of households that were wireless only in June 2006.

6.65. **Change in wireless only.** Refer to the previous exercise. The percent increased to 31.6% in June 2011. Assume the same sample size for this sample.

(a) Find the proportion and the count for this sample.

(b) Compute the 95% confidence interval for the proportion.

(c) Convert the estimate and confidence interval in terms of proportions to an estimate and confidence interval in terms of percents.

(d) Find the estimate of the difference between the proportions of households that are wireless only in June 2011 and the households that are wireless only in June 2006.

(e) Give the margin of error for 95% confidence for the difference in proportions.

6.66. **Student employment during the school year.** A study of 1430 undergraduate students reported that 994 work 10 or more hours a week during the school year. Give a 95% confidence interval for the proportion of all undergraduate students who work 10 or more hours a week during the school year.

6.67. **Examine the effect of the sample size.** Refer to the previous exercise. Assume a variety of different scenarios where the sample size changes but the proportion in the sample who work 10 or more hours a week during the school year remains the same. Write a short report summarizing your results and conclusions. Be sure to include numerical and graphical summaries of what you have found.

6.68. **Using a handheld phone while driving.** Refer to Exercise 6.32 (page xxx). This same poll found that 58% of the respondents talked on a handheld phone while driving in the last year. Construct a 90% confidence interval for the proportion of U.S. drivers who talked on a handheld phone while driving in the last year.

6.69. **Gender and using a handheld phone while driving.** Refer to the previous exercise. In this same report, this percent was broken down into 59% for men and 56% for women. Assuming that among the 1048 respondents, there were an equal number of men and women, construct a 95% confidence interval for the difference in these proportions.

6.70. **Even more on downloading music from the Internet.** The following quotation is from a survey of Internet users. The sample size for the survey was 1371. Since 18% of those surveyed said they download music, the sample size for this subsample is 247.

Among current music down loaders, 38% say they are downloading less because of the RIAA suits . . . About a third of current music down loaders say they use peer-to-peer networks . . . 24% of them say they swap files using email and instant messaging; 20% download files from music-related Web sites like those run by music magazines or musician homepages. And while online music services like iTunes are far from trumping the popularity of file-sharing networks, 17% of current music down loaders say they are using these paid services. Overall, 7% of Internet users say they have
CHAPTER 6 Inference for Proportions

bought music at these new services at one time or another, including 3% who currently use paid services.\textsuperscript{28}

(a) For each percent quoted, give the margin of error. You should express these in percents, as given in the quotation.

(b) Rewrite the paragraph more concisely and include the margins of error.

(c) Pick either side A or side B below and give arguments in favor of the view that you select.

(A) The margins of error should be included because they are necessary for the reader to properly interpret the results.

(B) The margins of error interfere with the flow of the important ideas. It would be better to just report one margin of error and say that all the others are no greater than this number.

If you choose View B, be sure to give the value of the margin of error that you report.

6.71. Parental pressure to succeed in school. A Pew Research Center Poll used telephone interviews to ask American adults if parents are pushing their kids too hard to succeed in school. Of those responding, 64% said parents are placing too little pressure on their children.\textsuperscript{29}

Assuming that this is an SRS of 1200 U.S. residents over the age of 18, give the 95% margin of error for this estimate.

6.72. Improving the time to repair golf clubs. The Ping Company makes custom-built golf clubs and competes in the $4 billion golf equipment industry. To improve its business processes, Ping decided to seek ISO 9001 certification.\textsuperscript{30} As part of this process, a study of the time it took to repair golf clubs that were sent to the company by mail determined that 16% of orders were sent back to the customers in 5 days or less. Ping examined the processing of repair orders and made changes. Following the changes, 90% of orders were completed within 5 days. Assume that each of the estimated percents is based on a random sample of 200 orders.

(a) How many orders were completed in 5 days or less before the changes? Give a 95% confidence interval for the proportion of orders completed in this time.

(b) Do the same for orders after the changes.

(c) Give a 95% confidence interval for the improvement. Express this both for a difference in proportions and for a difference in percents.

6.73. Gallup Poll study. Go to the Gallup Poll Web site at gallup.com/ and find two Gallup Daily Polls that interest you. Summarize the results of the polls giving margins of error and comparisons of interest. (For this exercise, you may assume that the data come from SRSs.)

6.74. Brand loyalty and the Chicago Cubs. According to literature on brand loyalty, consumers who are loyal to a brand are likely to consistently select the same product. This type of consistency could come from a positive childhood association. To examine brand loyalty among fans of the Chicago Cubs, 371 Cubs fans among patrons of a restaurant located in Lakeview were surveyed prior to a game at Wrigley Field, the Cubs’ home field.\textsuperscript{31} The respondents were classified as “die-hard fans” or “less loyal fans.” Of the 134 die-hard fans, 90.3% reported that they had watched or listened to Cubs games when they were children. Among the 237 less loyal fans, 67.9% said that they had watched or listened as children.

(a) Find the number of die-hard Cubs fans who watched or listened to games when they were children. Do the same for the less loyal fans.

(b) Use a significance test to compare the die-hard fans with the less loyal fans with respect to their childhood experiences relative to the team.

(c) Express the results with a 95% confidence interval for the difference in proportions.

6.75. Brand loyalty in action. The study mentioned in the previous exercise found that two-thirds of the die-hard fans attended Cubs games at least once a month, but only 20% of the less loyal fans attended this often. Analyze these data using a significance test and a confidence interval. Write a short summary of your findings.

6.76. More on gender bias in textbooks. Refer to the study of gender bias and textbooks described in Exercise 6.59 (page xxx). Here are the counts of “girl,” “woman,” “boy,” and “man” for all the syntax texts studied. The one we analyzed in Exercise 6.59 was number 6.
For each text perform the significance test to compare the proportions of juvenile references for females and males. Summarize the results of the significance tests for the 10 texts studied. The researchers who conducted the study note that the authors of the last 3 texts are women, while the other 7 texts were written by men. Do you see any pattern that suggests that the gender of the author is associated with the results?

6.77. Even more on gender bias in textbooks. Refer to the previous exercise. Let’s now combine the categories “girl” with “woman” and “boy” with “man.” For each text calculate the proportion of male references and test the hypothesis that male and female references are equally likely (that is, the proportion of male references is equal to 0.5). Summarize the results of your 10 tests. Is there a pattern that suggests a relation with the gender of the author?

6.78. Parental pressure and gender. The Pew Research Center Poll in Exercise 6.71 (page xxx) also reported that 65% of the men and 62% of the women thought parents are placing too little pressure on their children to succeed in school. Assuming that the respondents were 52% women, compare the proportions with a significance test and give a 95% confidence interval for the difference. Write a summary of your results.

6.79. Sample size and the P-value. In this exercise we examine the effect of the sample size on the significance test for comparing two proportions. In each case suppose that \( \hat{p}_1 = 0.5 \) and \( \hat{p}_2 = 0.4 \), and take \( n \) to be the common value of \( n_1 \) and \( n_2 \). Use the \( z \) statistic to test \( H_0: p_1 = p_2 \) versus the alternative \( H_a: p_1 \neq p_2 \). Compute the statistic and the associated \( P \)-value for the following values of \( n \): 40, 50, 80, 100, 400, 500, and 1000. Summarize the results in a table. Explain what you observe about the effect of the sample size on statistical significance when the sample proportions \( \hat{p}_1 \) and \( \hat{p}_2 \) are unchanged.

6.80. A corporate liability trial. A major court case on the health effects of drinking contaminated water took place in the town of Woburn, Massachusetts. A town well in Woburn was contaminated by industrial chemicals. During the period that residents drank water from this well, there were 16 birth defects among 414 births. In years when the contaminated well was shut off and water was supplied from other wells, there were 3 birth defects among 228 births. The plaintiffs suing the firm responsible for the contamination claimed that these data show that the rate of birth defects was higher when the contaminated well was in use. How statistically significant is the evidence? What assumptions does your analysis require? Do these assumptions seem reasonable in this case?

6.81. Attitudes toward student loan debt. The National Student Loan Survey asked the student loan borrowers in their sample about attitudes toward debt. Here are some of the questions they asked, with the percent who responded in a particular way:

(a) “To what extent do you feel burdened by your student loan payments?” 55.5% said they felt burdened.
(b) “If you could begin again, taking into account your current experience, what would you borrow?” 54.4% said they would borrow less.
(c) “Since leaving school, my education loans have not caused me more financial hardship than I had anticipated at the time I took out the loans.” 34.3% disagreed.
(d) “Making loan payments is unpleasant but I know that the benefits of education loans are worth it.” 58.9% agreed.
(e) “I am satisfied that the education I invested in with my student loan(s) was worth the investment for career opportunities.” 58.9% agreed.
(f) “I am satisfied that the education I invested in with my student loan(s) was worth the investment for personal growth.” 71.5% agreed.

Assume that the sample size is 1280 for all of these questions. Compute a 95% confidence interval for each of the questions, and write a short report about what student loan borrowers think about their debt.