Hans-Peter Merten/Getty Images

CHAPTER 8

Inference for Proportions

Introduction

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. When we record categorical variables, our data consist of *counts* or of *percents* obtained from counts.

In these settings, our goal is to say something about the corresponding *population proportions*. Just as in the case of inference about population means, we may be concerned with a single population or with comparing two populations. Inference about one or two proportions is very similar to inference about means, which we discussed in Chapter 7. In particular, inference for both means and proportions is based on sampling distributions that are approximately Normal.

We begin in Section 8.1 with inference about a single population proportion. Section 8.2 concerns methods for comparing two proportions.

8.1 Inference for a Single Proportion

When you complete this section, you will be able to:

- Identify the sample proportion, the sample size, and the count for a single sample. Use this information to estimate the population proportion.
- Describe the relationship between the population proportion and the sample proportion.
- Identify the standard error for a sample proportion and the margin of error for confidence level *C*.
- Apply the guidelines for when to use the large-sample confidence interval for a population proportion.
- Find and interpret the large-sample confidence interval for a single proportion.
- Apply the guidelines for when to use the large-sample significance test for a population proportion.
- Use the large-sample significance test to test a null hypothesis about a population proportion.
- Find the sample size needed for a desired margin of error.
- Find the sample size needed for a significance test.

LOOK BACK
simple random sample, p. 191
sampling distribution of a count, p. 314

We want to estimate the proportion p of some characteristic in a large population. For example, we may want to know the proportion of likely voters who approve of the president's conduct in office. We select a simple random sample (SRS) of size n from the population and record the count X of "successes" (such as Yes answers to a question about the president). A "success" response represents the characteristic of interest in this example.

In statistical terms, we are concerned with inference about the probability p of a success in the binomial setting. The sample proportion of successes $\hat{p} = X/n$ estimates the unknown population proportion p. If the population is much larger than the sample (at least 20 times as large), the count X has approximately the binomial distribution $B(n, p)$.[1]

EXAMPLE 8.1

ROBOT

Robotics and jobs. A Pew survey asked a panel of experts whether or not they thought that networked, automated, artificial intelligence (AI), and robotic devices will have displaced more jobs than they have created (net jobs) by 2025.[2]

The sample size is the number of experts who responded to the Pew survey question, $n = 1896$. The report on the survey tells us that 48% of the respondents said they "believe net jobs will decrease by 2025 due to networked, automated, artificial intelligence (AI), and robotic devices." Thus, the sample proportion is $\hat{p} = 0.48$. We can calculate the count X from the information given; it is the sample size times the proportion responding Yes, $X = n\hat{p} = 1896(0.48) = 910$.

USE YOUR KNOWLEDGE

8.1 Smartphones and purchases. A Google research study asked 5013 smartphone users about how they used their phones. In response to a question about purchases, 2657 reported that they purchased an item after using their smartphone to search for information about the item.[3]

(a) What is the sample size n for this survey?

(b) In this setting, describe the population proportion p in a short sentence.

(c) What is the value of the count X? Describe the count in a short sentence.

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

8.2 Coca-Cola and demographics. A Pew survey interviewed 162 CEOs from U.S. companies. The report of the survey quotes Muhtar Kent, Coca-Cola Company chairman and CEO, on the importance of demographics in developing customer strategies. Kent notes that the population of the United States is aging and that there is a need to provide products that appeal to this segment of the market. The survey found that 52% of the CEOs in the sample are planning to change their customer growth and retention strategies.

(a) How many CEOs participated in the survey? What is the sample size n for the survey?

(b) What is the count X of those who said that they are planning to change their customer growth and retention strategies?

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

LOOK BACK
anecdotal data, p. 164

(d) The quotes from Muhtar Kent in the report could be viewed as anecdotal data. Do you think that these quotes are useful to explain and interpret the results of the survey? Write a short paragraph discussing your answer.

If the sample size n is very small, we must base tests and confidence intervals for p on the binomial distributions. These are awkward to work with because of the discreteness of the binomial distributions.[4] But we know that when the these counts are large, both the count X and the sample proportion $\hat{p}$ are approximately Normal. We will consider only inference procedures based on the Normal approximation. These procedures are similar to those for inference about the mean of a Normal distribution.

LOOK BACK
Normal approximations for counts and proportions, p. 322

Large-sample confidence interval for a single proportion

The unknown population proportion p is estimated by the sample proportion $\hat{p} = X/n$. If the sample size n is sufficiently large, the sampling distribution of $\hat{p}$ is approximately Normal, with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. This means that approximately 95% of the time $\hat{p}$ will be within $2\sqrt{p(1-p)/n}$ of the unknown population proportion p.

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 formula by the sample proportion $\hat{p}$. As we did in Chapter 7, we use the term *standard error* for the standard deviation of a statistic that is estimated from data. Here is a summary of the procedure.

LOOK BACK
standard error, p. 408

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

$$\mathrm{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

and the **margin of error** for confidence level C is

$$m = z^*\mathrm{SE}_{\hat{p}}$$

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% ($z^* = 1.645$), 95% ($z^* = 1.96$), or 99% ($z^* = 2.576$) confidence when the number of successes and the number of failures are both at least 10.

Table D includes a line at the bottom with values of z^* for selected values of C. Use Table A for other values of C.

EXAMPLE 8.2

Inference for robotics and jobs. The sample survey in Example 8.1 found that 910 of a sample of 1896 experts reported that they think net jobs will decrease by 2025 because of robots and related technology developments. Thus, the sample size is $n = 1896$ and the count is $X = 910$. The sample proportion is

$$\hat{p} = \frac{X}{n} = \frac{910}{1896} = 0.47996$$

The standard error is

$$\mathrm{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.47996(1-0.47996)}{1896}} = 0.011474$$

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

$$m = 1.96\mathrm{SE}_{\hat{p}} = (1.96)(0.011474) = 0.022489$$

The confidence interval is

$$\hat{p} \pm m = 0.480 \pm 0.022$$

We are 95% confident that between 45.8% and 50.2% of CEOs would report that they think net jobs will decrease by 2025 because of robots and related technology developments.

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, "48.0% with a margin of error of 2.2%." *You should always focus on what is important. Reporting extra digits that are not needed can divert attention from the main point of your summary.* There is no additional information to be gained by reporting $\hat{p} = 0.47996$ with a margin of error of 0.022489. Do you think it would be better to report 48% with a 2% margin of error?

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. Likewise, if the response rate is low, your estimate and standard error may be biased.

Although the calculations for statistical inference for a single proportion are relatively straightforward and can be done with a calculator or in a spreadsheet, we prefer to use software.

EXAMPLE 8.3

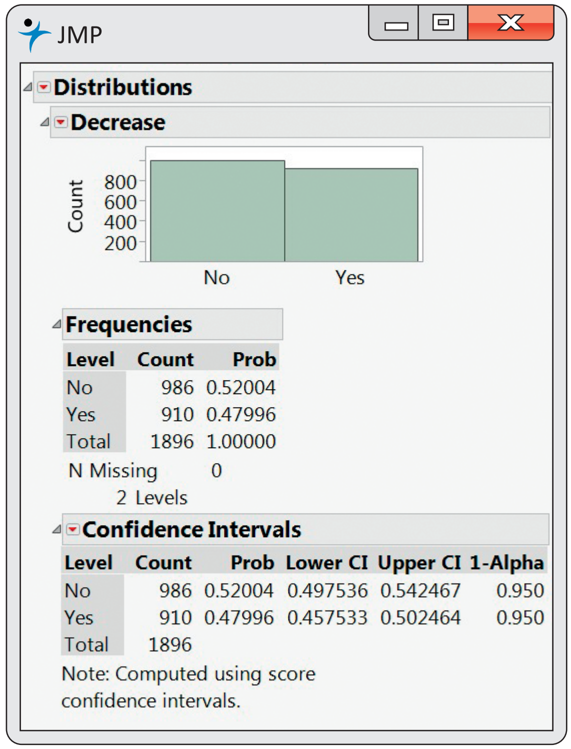
Robotics and jobs confidence interval using software. Figure 8.1 shows a spreadsheet for the robotics and jobs example that could be used as input for statistical software. Note that there are 1896 experts who expressed an opinion in this example. The sheet specifies a value for each of these 1896 cases: there are 910 cases with the value Yes and 986 cases with the value No. An alternative spreadsheet would not summarize the responses but rather would list all 1896 cases with the response for each case.

Excel

	A	B
1	Decrease	Count
2	Yes	910
3	No	986

FIGURE 8.1 The robotics and jobs data in an Excel spreadsheet for the confidence interval, Example 8.3.

Figure 8.2 gives output from JMP and Minitab for these data. There are differences in the displays, but it is easy to find the 95% confidence interval. For JMP, the confidence interval is on the line with "Level" equal to Yes under the headings "Lower CI" and "Upper CI." Minitab gives the output



Frequencies

Level	Count	Prob
No	986	0.52004
Yes	910	0.47996
Total	1896	1.00000

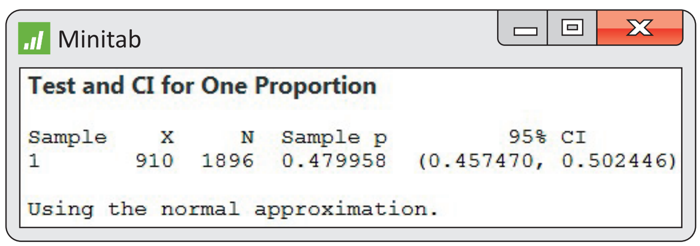
N Missing 0
2 Levels

Confidence Intervals

Level	Count	Prob	Lower CI	Upper CI	1-Alpha
No	986	0.52004	0.497536	0.542467	0.950
Yes	910	0.47996	0.457533	0.502464	0.950
Total	1896				

Note: Computed using score confidence intervals.

(a)



Test and CI for One Proportion

```
Sample     X     N  Sample p          95% CI
1        910  1896  0.479958  (0.457470, 0.502446)

Using the normal approximation.
```

(b)

FIGURE 8.2 (a) JMP and (b) Minitab output for the robotics and jobs survey, Example 8.3.

in the form of an interval under the heading "95% CI." Notice that the confidence intervals are similar but not identical. Minitab notes that the Normal approximation is used. This is the large-sample method that we described. JMP notes that an alternative method, using score functions, is used.

CAUTION

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.*

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 10. For smaller sample sizes, we recommend exact methods that use the binomial distribution. These, as well as other alternative procedures, such as the score function, are available as the default or as options in many statistical software packages. We do not cover them here. There is also an intermediate case between large samples and very small samples where a slight modification of the large-sample approach works quite well. This method is called the "plus four" procedure and is described next.

USE YOUR KNOWLEDGE

8.3 Robotics and jobs. Refer to Example 8.1 (page 484).

(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.

(d) State your conclusion and interpret the meaning of the confidence interval in part (c).

8.4 Coca-Cola and demographics. Refer to Exercise 8.2 (page 485).

(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.

(d) State your conclusion and interpret the meaning of the confidence interval in part (c).

BEYOND THE BASICS

The Plus Four Confidence Interval for a Single Proportion

Computer studies reveal that confidence intervals based on the large-sample approach can be quite inaccurate when the number of successes and the number of failures are not at least 10. When this occurs, a simple adjustment to the confidence interval works very well in practice. The adjustment is based on assuming that the sample contains four additional observations, two of which are successes and two of which are failures. The estimator of the population proportion based on this *plus four* rule is

$$\tilde{p} = \frac{X + 2}{n + 4}$$

plus four estimate

This estimate was first suggested by Edwin Bidwell Wilson in 1927, and it is sometimes called the Agresti-Coull interval.[5] We call it the **plus four estimate**. The confidence interval is based on the z statistic obtained by standardizing the plus four estimate $\tilde{p}$. Because $\tilde{p}$ is the sample proportion for our modified sample of size $n + 4$, it isn't surprising that the distribution of $\tilde{p}$ is close to the Normal distribution with mean p and standard deviation $\sqrt{p(1 - p)/(n + 4)}$. To get a confidence interval, we estimate p by $\tilde{p}$ in this standard deviation to get the standard error of $\tilde{p}$. Here is an example.

EXAMPLE 8.4


Asia Images Group/Getty Images

Percent of equol producers. Research has shown that there are many health benefits associated with a diet that contains soy foods. Substances in soy called isoflavones are known to be responsible for these benefits. When soy foods are consumed, some subjects produce a chemical called equol, and it is thought that production of equol is a key factor in the health benefits of a soy diet. Unfortunately, not all people are equol producers; there appear to be two distinct subpopulations: equol producers and equol nonproducers.

A nutrition researcher planning some bone health experiments would like to include some equol producers and some nonproducers among her subjects. A preliminary sample of 12 female subjects were measured, and four were found to be equol producers. We would like to estimate the proportion of equol producers in the population from which this researcher will draw her subjects.

The plus four estimate of the proportion of equol producers is

$$\tilde{p} = \frac{4+2}{12+4} = \frac{6}{16} = 0.375$$

For a 95% confidence interval, we use Table D to find $z^* = 1.96$. We first compute the standard error

$$\begin{aligned} \text{SE}_{\tilde{p}} &= \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} \\ &= \sqrt{\frac{(0.375)(1-0.375)}{16}} \\ &= 0.12103 \end{aligned}$$

and then the margin of error

$$\begin{aligned} m &= z^*\text{SE}_{\tilde{p}} \\ &= (1.96)(0.12103) \\ &= 0.237 \end{aligned}$$

So the confidence interval is

$$\begin{aligned} \tilde{p} \pm m &= 0.375 \pm 0.237 \\ &= (0.138, 0.612) \end{aligned}$$

We estimate with 95% confidence that between 14% and 61% of women from this population are equol producers. Note that the interval is very wide because the sample size is very small. Compare this result with the large-sample confidence interval.

If the true proportion of equol users is near 14%, the lower limit of this interval, there may not be a sufficient number of equol producers in the study if subjects are tested only after they are enrolled in the experiment. It may be necessary to determine whether or not a potential subject is an equol producer. The study could then be designed to have the same number of equol producers and nonproducers.

Significance test for a single proportion

LOOK BACK
Normal approximation for proportions, p. 322

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.

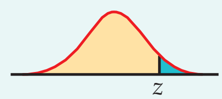
LARGE-SAMPLE SIGNIFICANCE TEST FOR A POPULATION PROPORTION

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**

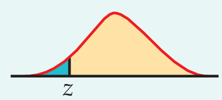
$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

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

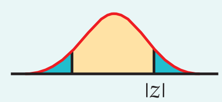
H_a: $p > p_0$ is $P(Z \geq z)$



H_a: $p < p_0$ is $P(Z \leq z)$



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



We recommend 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 at least 10.

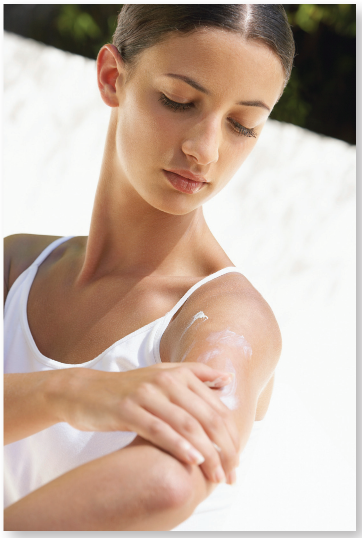
LOOK BACK
sign test for matched pairs, p. 473

If the numbers of successes and failures are not both at least 10, or if the population is less than 20 times as large as the sample, other procedures should be used. One such approach is to use the binomial distribution as we did with the sign test. Here is a large-sample example.

EXAMPLE 8.5

SUNBL

Comparing two sunblock lotions. Your company produces a sunblock lotion designed to protect the skin from both UVA and UVB exposure to the sun. You hire a company to compare your product with the product sold by your major competitor. The testing company 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


Fancy/Alamy

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? For the data we have $n = 20$ subjects and $X = 13$ successes. The parameter p is the proportion of people who would receive superior UVA and UVB protection from your product. To answer the claim question, we test

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

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

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

From Table A, we find $P(Z < 1.34) = 0.9099$, so the probability in the upper tail is $1 - 0.9099 = 0.0901$. The P-value is the area in both tails, $P = 2 \times 0.0901 = 0.1802$.

We conclude that the sunblock testing data are compatible with the hypothesis of no difference between your product and your competitor's product ($\hat{p} = 0.65$, $z = 1.34$, $P = 0.18$). The data do not support your proposed advertising claim.

Note that we have used the two-sided alternative for this example. 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.

Although these calculations are not particularly difficult to do using a calculator, we prefer to use software. Here are some details.

EXAMPLE 8.6

SUNBL

Sunblock significance tests using software. JMP and Minitab outputs for the analysis in Example 8.5 appear in Figure 8.3. First, JMP uses a slightly different way of reporting the results. Two ways of performing the significance test are labeled in the column "Test." The one that corresponds to the procedure that we have described is on the second line, labeled "Pearson." The P-value under the heading "Prob>Chisq" is 0.1797, which is very close to the 0.1802 that we calculated using Table A. Minitab reports the value of the test statistic z, and the P-value is rounded to 0.180.

JMP

Distributions

Product

Frequencies

Level	Count	Prob
Theirs	7	0.35000
Yours	13	0.65000
Total	20	1.00000

N Missing 0
2 Levels

Test Probabilities

Level	Estim Prob	Hypoth Prob
Theirs	0.35000	0.50000
Yours	0.65000	0.50000

Test	ChiSquare	DF	Prob>Chisq
Likelihood Ratio	1.8280	1	0.1764
Pearson	1.8000	1	0.1797

(a)

Minitab

Test and CI for One Proportion

```
Test of p = 0.5 vs p ≠ 0.5


Sample    X   N  Sample p        95% CI            Z-Value  P-Value
1        13  20  0.650000  (0.440963, 0.859037)     1.34     0.180

Using the normal approximation.
```

(b)

FIGURE 8.3 (a) JMP and (b) Minitab output for comparing sunblock lotions, Example 8.5.

USE YOUR KNOWLEDGE

8.5 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 8.5.

8.6 What does the confidence interval tell us? Inspect the outputs in Figure 8.3. 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 and compare this way of analyzing data with the significance test.

8.7 The effect of *X*. In Example 8.5 (page 491), suppose that your product provided better UVA and UVB protection for 15 of the 20 subjects. Perform the significance test and summarize the results.

8.8 The effect of *n*. In Example 8.5 (page 491), 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 Example 8.5, 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 8.5, 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

8.9 Redefining success. In Example 8.5 (page 491), 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, but X is now 7.

(a) Perform the two-sided significance test and report the results. How do these compare with what we found in Example 8.5?

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

CAUTION

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. The matched pairs study of Example 8.5 is an example of a comparative study that involved a single proportion.

Choosing a sample size for a confidence interval

LOOK BACK
choosing sample size, p. 351

In Chapter 6, we showed how to choose the sample size n to obtain a confidence interval with specified margin of error m for a Normal mean. We also discussed the effect of the sample size on the power of a significance test for a Normal mean. Because we are using a Normal approximation for inference about a population proportion, sample size selection proceeds in much the same way.

Recall that the margin of error for the large-sample confidence interval for a population proportion is

$$m = z^* \mathrm{SE}_{\hat{p}} = z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Choosing a confidence level C fixes the critical value z^*. The margin of error also depends on the value of $\hat{p}$ and the sample size n. Because we don't know the value of $\hat{p}$ until we gather the data, we must guess a value to use in the calculations. We will call the guessed value p^*. There are two common ways to get p^*:

1. Use the sample estimate from a pilot study or from similar studies done earlier.
2. Use $p^* = 0.5$. Because the margin of error is largest when $\hat{p} = 0.5$, this choice gives a sample size that is somewhat larger than we really need for the confidence level we choose. It is a safe choice no matter what the data later show.

Once we have chosen p^* and the margin of error m that we want, we can find the n we need to achieve this margin of error. Here is the result.

SAMPLE SIZE FOR DESIRED MARGIN OF ERROR

The level C confidence interval for a proportion p will have a margin of error approximately equal to a specified value m when the sample size satisfies

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1 - p^*)$$

Here, z^* is the critical value for confidence level C, and p^* is a guessed value for the proportion of successes in the future sample.

The margin of error will be less than or equal to m if p^* is chosen to be 0.5. Substituting $p^* = 0.5$ into the formula above gives

$$n = \frac{1}{4}\left(\frac{z^*}{m}\right)^2$$

The value of n obtained by this method is not particularly sensitive to the choice of p^* when p^* is fairly close to 0.5. However, if the value of p is likely to be smaller than about 0.3 or larger than about 0.7, use of $p^* = 0.5$ may result in a sample size that is much larger than needed.

EXAMPLE 8.7

Planning a survey of students. A large university is interested in assessing student satisfaction with the overall campus environment. The plan is to distribute a questionnaire to an SRS of students, but before proceeding, the university wants to determine how many students to sample. The questionnaire asks about a student's degree of satisfaction with various student services, each measured on a five-point scale. The university is interested in the proportion p of students who are satisfied (that is, who choose either "satisfied" or "very satisfied," the two highest levels on the five-point scale).

The university wants to estimate p with 95% confidence and a margin of error less than or equal to 3%, or 0.03. For planning purposes, it is willing to use $p^* = 0.5$. To find the sample size required,

$$n = \frac{1}{4}\left(\frac{z^*}{m}\right)^2 = \frac{1}{4}\left(\frac{1.96}{0.03}\right)^2 = 1067.1$$

Round up to get $n = 1068$. (Always round up. Rounding down would give a margin of error slightly greater than 0.03.)

Similarly, for a 2.5% margin of error, we have (after rounding up)

$$n = \frac{1}{4}\left(\frac{1.96}{0.025}\right)^2 = 1537$$

and for a 2% margin of error,

$$n = \frac{1}{4}\left(\frac{1.96}{0.02}\right)^2 = 2401$$

News reports frequently describe the results of surveys with sample sizes between 1000 and 1500 and a margin of error of about 3%. These surveys generally use sampling procedures more complicated than simple random sampling, so the calculation of confidence intervals is more involved than what we have studied in this section. The calculations in Example 8.7 show in principle how such surveys are planned.

In practice, many factors influence the choice of a sample size. The following example illustrates one set of factors.

EXAMPLE 8.8

iStockphoto

Assessing interest in Pilates classes. The Division of Recreational Sports (Rec Sports) at a major university is responsible for offering comprehensive recreational programs, services, and facilities to the students. Rec Sports is continually examining its programs to determine how well it is meeting the needs of the students. Rec Sports is considering adding some new programs and would like to know how much interest there is in a new exercise program based on the Pilates method.[6] It will take a survey of undergraduate students. In the past, Rec Sports emailed short surveys to all undergraduate students. The response rate obtained in this way was about 5%. This time, it will send emails to a simple random sample of the students and will follow up with additional emails and eventually a phone call to get a higher response rate. Because of limited staff and the work involved with the follow-up, it would like to use a sample size of about 200 responses. It assumes that the new procedures will improve the response rate to 90%, so it will contact 225 students in the hope that these will provide at least 200 valid responses. One of the questions it will ask is, "Have you ever heard about the Pilates method of exercise?"

The primary purpose of the survey is to estimate various sample proportions for undergraduate students. Will the proposed sample size of $n = 200$ be adequate to provide Rec Sports with the needed information? To address this question, we calculate the margins of error of 95% confidence intervals for various values of $\hat{p}$.

EXAMPLE 8.9

Margins of error. In the Rec Sports survey, the margin of error of a 95% confidence interval for any value of $\hat{p}$ and $n = 200$ is

$$\begin{aligned} m &= z^*\mathrm{SE}_{\hat{p}} \\ &= 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{200}} \\ &= 0.139\sqrt{\hat{p}(1-\hat{p})} \end{aligned}$$

The results for various values of $\hat{p}$ are

$\hat{p}$	m	$\hat{p}$	m
0.05	0.030	0.60	0.068
0.10	0.042	0.70	0.064
0.20	0.056	0.80	0.056
0.30	0.064	0.90	0.042
0.40	0.068	0.95	0.030
0.50	0.070		

Rec Sports judged these margins of error to be acceptable, and it contacted 225 students, hoping to achieve a sample size of 200 for its survey.

The table in Example 8.9 illustrates two points. First, the margins of error for $\hat{p} = 0.05$ and $\hat{p} = 0.95$ are the same. The margins of error will always be the same for $\hat{p}$ and $1 - \hat{p}$. This is a direct consequence of the form of the confidence interval. Second, the margin of error varies between only 0.064 and 0.070 as $\hat{p}$ varies from 0.3 to 0.7, and the margin of error is greatest when $\hat{p} = 0.5$, as we claimed earlier (page xxx). It is true in general that the margin of error will vary relatively little for values of $\hat{p}$ between 0.3 and 0.7. Therefore, when planning a study, it is not necessary to have a very precise guess for p. If $p^* = 0.5$ is used and the observed $\hat{p}$ is between 0.3 and 0.7, the actual interval will be a little shorter than needed, but the difference will be small.

Again it is important to emphasize that these calculations consider only the effects of sampling variability that are quantified in the margin of error. Other sources of error, such as nonresponse and possible misinterpretation of questions, are not included in the table of margins of error for Example 8.9. Rec Sports is trying to minimize these kinds of errors. It performed a pilot study using a small group of current users of its facilities to check the wording of the questions, and for the final survey it devised a careful plan to follow up with the students who did not respond to the initial email.

USE YOUR KNOWLEDGE

8.10 Confidence level and sample size. Refer to Example 8.7 (page 494). Suppose that the university was interested in a 95% confidence interval with margin of error 0.04. Would the required sample size be smaller or larger than 1068 students? Verify this by performing the calculation.

8.11 Make a plot. Use the values for $\hat{p}$ and m given in Example 8.9 to draw a plot of the sample proportion versus the margin of error. Summarize the major features of your plot.

Choosing a sample size for a significance test

LOOK BACK
power, p. 391

In Chapter 6, we also introduced the idea of power for a significance test. These ideas apply to the significance test for a proportion that we studied in this section. There are some more complicated details, but the basic ideas are the same. Fortunately, software can take care of the details, and we can concentrate on the input and output.

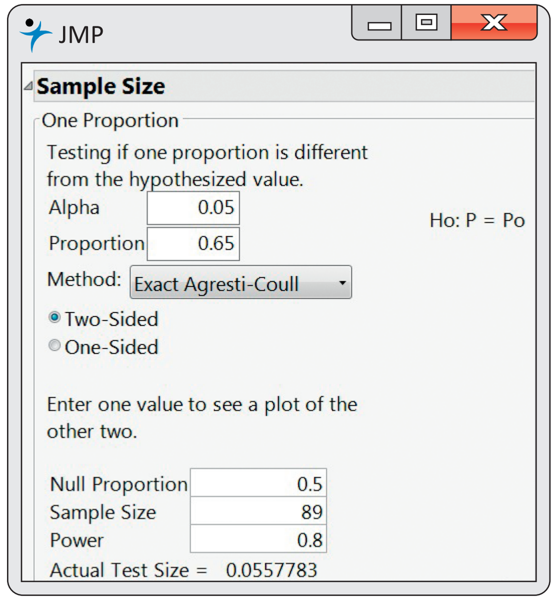
To find the required sample size, we need to specify

- The value of p_0 in the null hypothesis H_0: $p = p_0$.
- The alternative hypothesis, two-sided (H_a: $p \neq p_0$), one-sided (H_a: $p > p_0$ or H_a: $p < p_0$).
- A value of p for the alternative hypothesis.
- The Type I error (α, the probability of rejecting the null hypothesis when it is true); usually we choose 5% ($\alpha = 0.05$) for the Type I error.
- Power (probability of rejecting the null hypothesis when it is false); usually we choose 80% (0.80) for power.

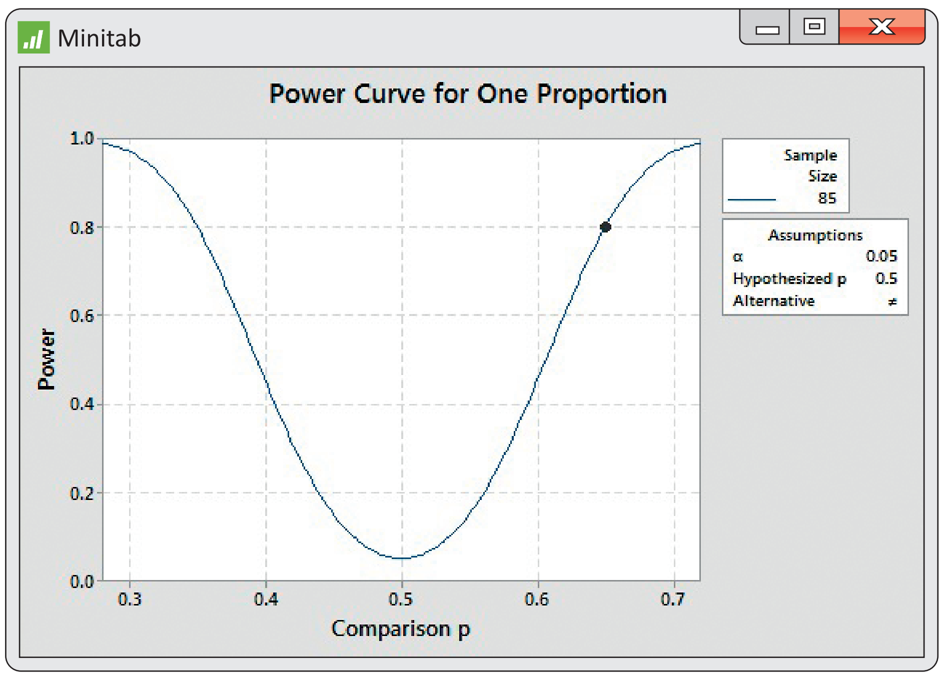
EXAMPLE 8.10

Sample size for comparing two sunblock lotions. In Example 8.5 (page 491), we performed the significance test for comparing two sunblock lotions in a setting where each subject used the two lotions and the product that provided better protection was recorded. Although your product performed better 13 times in 20 trials, the value of $\hat{p} = 13/20 = 0.65$ was not sufficiently far from the null hypothesized value of $p_0 = 0.5$ for us to reject the H_0, ($p = 0.18$). Let's suppose that the true percent of the time that your lotion would perform better is $p_0 = 0.65$, and we plan to test the null hypothesis H_0: $p = 0.5$ versus the two-sided alternative H_a: $p \neq 0.5$ using a Type I error probability of 0.05.

What sample size n should we choose if we want to have an 80% chance of rejecting H_0? Outputs from JMP and Minitab are given in Figure 8.4. JMP indicates that $n = 89$ should be used, while Minitab suggests $n = 85$. The difference is due to the different methods that can be used for these calculations.



(a)



(b)

FIGURE 8.4 (a) JMP and (b) Minitab output for sample size needed to compare sunblock lotions, Example 8.10.

Note that Minitab provides a graph as a function of the value of the proportion for the alternative hypothesis. Similar plots can be produced by JMP. In some situations, you might want to specify the sample size n and have software compute the power. This option is available in JMP, Minitab, and other software.

USE YOUR KNOWLEDGE

8.12 Compute the sample size for a different alternative. Refer to Example 8.10 (page 498). Use software to find the sample size needed for a two-sided test of the null hypothesis that $p = 0.5$ versus the two-sided alternative with $\alpha = 0.05$ and 80% power if the alternative is $p = 0.7$.

8.13 Compute the power for a given sample size. Consider the setting in Example 8.10 (page 498). You have a budget that will allow you to test 100 subjects. Use software to find the power of the test for this value of n.

SECTION 8.1 SUMMARY

- 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 $\sqrt{p(1-p)/n}$.

- For large samples, the **margin of error for confidence level *C*** 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 **standard error of $\hat{p}$** is

$$\text{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- The **level *C* large-sample confidence interval** is

$$\hat{p} \pm m$$

We recommend using this interval for 90%, 95%, and 99% confidence whenever the number of successes and the number of failures are both at least 10. When sample sizes are smaller, alternative procedures such as the **plus four estimate of the population proportion** are recommended.

- The **sample size** required to obtain a confidence interval of approximate margin of error m for a proportion is found from

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$$

where p^* is a guessed value for the proportion and z^* is the standard Normal critical value for the desired level of confidence. To ensure that the margin of error of the interval is less than or equal to m no matter what $\hat{p}$ may be, use

$$n = \frac{1}{4}\left(\frac{z^*}{m}\right)^2$$

- Tests of H_0: $p = p_0$ are based on the **z statistic**

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

with P-values calculated from the $N(0, 1)$ distribution. Use this procedure when the expected number of successes, np_0, and the expected number of failures, $n(1 - p_0)$, are both greater than 10.

- Software can be used to determine the sample sizes for significance tests.

SECTION 8.1 EXERCISES

For Exercises 8.1 and 8.2, see page 485; for Exercises 8.3 and 8.4, see page 489; for Exercises 8.5 through 8.8, see pages 493–494; for Exercise 8.9, see page 494; for Exercises 8.10 and 8.11, see page 498; and for Exercises 8.12 and 8.13, see page 500.

8.14 How did you use your cell phone? A Pew Internet poll asked cell phone owners about how they used their cell phones. One question asked whether or not during the past 30 days they had used their phone while in a store to call a friend or family member for advice about a purchase they were considering. The poll surveyed 1003 adults living in the United States by telephone. Of these, 462 responded that they had used their cell phone while in a store within the last 30 days to call a friend or family member for advice about a purchase they were considering.[7]

(a) Identify the sample size and the count.

(b) Calculate the sample proportion.

(c) Explain the relationship between the population proportion and the sample proportion.

8.15 Do you eat breakfast? A random sample of 300 students from your college is asked if they regularly eat breakfast. One hundred and nine students responded that they did eat breakfast regularly.

(a) Identify the sample size and the count.

(b) Calculate the sample proportion.

(c) Explain the relationship between the population proportion and the sample proportion.

8.16 Would you recommend the service to a friend? An automobile dealership asks all its customers who used its service department in a given two-week period if they would recommend the service to a friend. A total of 250 customers used the service during the two-week period, and 210 said that they would recommend the service to a friend.

(a) Identify the sample size and the count.

(b) Calculate the sample proportion.

(c) Explain the relationship between the population proportion and the sample proportion.

8.17 How did you use your cell phone? Refer to Exercise 8.14.

(a) Report the sample proportion, the standard error of the sample proportion, and the margin of error for 95% confidence.

(b) Are the guidelines for when to use the large-sample confidence interval for a population proportion satisfied in this setting? Explain your answer.

(c) Find the 95% large-sample confidence interval for the population proportion.

(d) Write a short statement explaining the meaning of your confidence interval.

8.18 Do you eat breakfast? Refer to Exercise 8.15.

(a) Report the sample proportion, the standard error of the sample proportion, and the margin of error for 95% confidence.

(b) Are the guidelines for when to use the large-sample confidence interval for a population proportion satisfied in this setting? Explain your answer.

(c) Find the 95% large-sample confidence interval for the population proportion.

(d) Write a short statement explaining the meaning of your confidence interval.

8.19 Would you recommend the service to a friend? Refer to Exercise 8.16.

(a) Report the sample proportion, the standard error of the sample proportion, and the margin of error for 95% confidence.

(b) Are the guidelines for when to use the large-sample confidence interval for a population proportion satisfied in this setting? Explain your answer.

(c) Find the 95% large-sample confidence interval for the population proportion.

(d) Write a short statement explaining the meaning of you confidence interval.

8.20 Whole grain versus regular grain? A study of young children was designed to increase their intake of whole-grain, rather than regular-grain, snacks. At the end of the study, the 86 children who participated in the study were presented with a choice between a regular-grain snack and a whole-grain alternative. The whole-grain alternative was chosen by 48 children. You want to examine the possibility that the children are equally likely to choose each type of snack.

(a) Formulate the null and alternative hypotheses for this setting.

(b) Are the guidelines for using the large-sample significance test satisfied for testing this null hypothesis? Explain your answer.

(c) Perform the significance test and summarize your results in a short paragraph.

8.21 Find the sample size. You are planning a survey similar to the one about cell phone use described in Exercise 8.14. You will report your results with a large-sample confidence interval. How large a sample do you need to be sure that the margin of error will not be greater than 0.05? Show your work.

8.22 What's wrong? Explain what is wrong with each of the following:

(a) An approximate 90% confidence interval for an unknown proportion p is $\hat{p}$ plus or minus its standard error.

(b) You can use a significance test to evaluate the hypothesis H_0: $\hat{p} = 0.3$ versus the one-sided alternative.

(c) The large-sample significance test for a population proportion is based on a t statistic.

8.23 What's wrong? Explain what is wrong with each of the following:

(a) A student project used a confidence interval to describe the results in a final report. The confidence level was 115%.

(b) 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.

(c) If the P-value for a significance test is 0.50, we can conclude that the null hypothesis has a 50% chance of being true.

8.24 Draw some pictures. Consider the binomial setting with $n = 800$ and $p = 0.3$.

(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.

8.25 Country food and Inuits. Country food includes seals, caribou, whales, ducks, fish, and berries and is an important part of the diet of the aboriginal people called Inuits who inhabit Inuit Nunangat, the northern region of what is now called Canada. A survey of Inuits in Inuit Nunangat reported that 3274 out of 5000 respondents said that at least half of the meat and fish that they eat is country food.[8] 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.

8.26 Soft drink consumption in New Zealand. A survey commissioned by the Southern Cross Healthcare Group reported that 16% of New Zealanders consume five or more servings of soft drinks per week. The data were obtained by an online survey of 2006 randomly selected New Zealanders over 15 years of age.[9]

(a) What number of survey respondents reported that they consume five or more servings of soft drinks per week? You will need to round your answer. Why?

(b) Find a 95% confidence interval for the proportion of New Zealanders who report that they consume five or more servings of soft drinks per week.

(c) Convert the estimate and your confidence interval to percents.

(d) Discuss reasons why the estimate might be biased.

8.27 Violent video games. A survey of 1050 parents who have a child under the age of 18 living at home asked about their opinions regarding violent video games. A report describing the results of the survey stated that 89% of parents say that violence in today's video games is a problem.[10]

(a) What number of survey respondents reported that they thought that violence in today's video games is a problem? You will need to round your answer. Why?

(b) Find a 95% confidence interval for the proportion of parents who think that violence in today's video games is a problem.

(c) Convert the estimate and your confidence interval to percents.

(d) Discuss reasons why the estimate might be biased.

8.28 Bullying. Refer to the previous exercise. The survey also reported that 93% of the parents surveyed said that bullying contributes to violence in the United States. Answer the questions in the previous exercise for this item on the survey.

8.29 $\hat{p}$ and the Normal distribution. Consider the binomial setting with $n = 40$. You are testing the null hypothesis that $p = 0.4$ versus the two-sided alternative with a 5% chance of rejecting the null hypothesis when it is true.

(a) Find the values of the sample proportion $\hat{p}$ that will lead to rejection of the null hypothesis.

(b) Repeat part (a) assuming a sample size of $n = 80$.

(c) Make a sketch illustrating what you have found in parts (a) and (b). What does your sketch show about the effect of the sample size in this setting?

8.30 Students doing community service. In a sample of 159,949 first-year college students, the National Survey of Student Engagement reported that 39% participated in community service or volunteer work.[11]

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

(b) Here are some facts from the report that summarizes the survey. The students were from 617 four-year colleges and universities. The response rate was 36%. Institutions paid a participation fee of between \$1800 and \$7800 based on the size of their undergraduate enrollment. Discuss these facts as possible sources of error in this study. How do you think these errors would compare with the margin of error that you calculated in part (a)?

8.31 Plans to study abroad. The survey described in the previous exercise also asked about items related to academics. In response to one of these questions, 42% of first-year students reported that they plan to study abroad.

(a) Based on the information available, how many students plan to study abroad?

(b) Give a 99% confidence interval for the population proportion of first-year college students who plan to study abroad.

8.32 Student credit cards. In a survey of 1430 undergraduate students, 1087 reported that they had one or more credit cards.[12] Give a 95% confidence interval for the proportion of all college students who have at least one credit card.

8.33 How many credit cards? The summary of the survey described in the previous exercise reported that 43% of undergraduates had four or more credit cards. Give a 95% confidence interval for the proportion of all college students who have four or more credit cards.

8.34 How would the confidence interval change? Refer to Exercise 8.33.

(a) Would a 80% confidence interval be wider or narrower than the one that you found in Exercise 8.33? Verify your answer by computing the interval.

(b) Would a 98% confidence interval be wider or narrower than the one that you found in that exercise? Verify your results by computing the interval.

8.35 Do students report Internet sources? The National Survey of Student Engagement found that 87% of students report that their peers at least "sometimes" copy information from the Internet in their papers without reporting the source.[13] Assume that the sample size is 430,000.

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

(b) Here are some items from the report that summarizes the survey. More than 430,000 students from 730 four-year colleges and universities participated. The average response rate was 43% and ranged from 15% to 89%. Institutions pay a participation fee of between \$3000 and \$7500 based on the size of their undergraduate enrollment. Discuss these facts as possible sources of error in this study. How do you think these errors would compare with the error that you calculated in part (a)?

8.36 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. Explain your answers.

(a) $n = 30$ and H_0: $p = 0.3$.

(b) $n = 60$ and H_0: $p = 0.2$.

(c) $n = 100$ and H_0: $p = 0.12$.

(d) $n = 150$ and H_0: $p = 0.04$.

8.37 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.[14] One question 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 estimate the true proportion for this question with a 95% confidence interval.

(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.

8.38 Instant versus fresh-brewed coffee. A matched pairs experiment compares the taste of instant with 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 60 subjects who participate in the study, 21 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?

8.39 Tossing a coin 10,000 times! The South African mathematician John Kerrich, while a prisoner of war during World War II, tossed a coin 10,000 times and obtained 5067 heads.

(a) Is this significant evidence at the 5% level that the probability that Kerrich's coin comes up heads is not 0.5? Use a sketch of the standard Normal distribution to illustrate the P-value.

(b) Use a 95% confidence interval to find the range of probabilities of heads that would not be rejected at the 5% level.

8.40 Is there interest in a new product? One of your employees has suggested that your company develop a new product. You decide to take a random sample of your customers and ask whether or not there is interest in the new product. The response is on a 1 to 5 scale with 1 indicating "definitely would not purchase"; 2, "probably would not purchase"; 3, "not sure"; 4, "probably would purchase"; and 5, "definitely would purchase." For an initial analysis, you will record the responses 1, 2, and 3 as No and 4 and 5 as Yes. What sample size would you use if you wanted the 95% margin of error to be 0.25 or less?

8.41 More information is needed. Refer to the previous exercise. Suppose that after reviewing the results of the previous survey, you proceeded with preliminary development of the product. Now you are at the stage where you need to decide whether or not to make a major investment to produce and market it. You will use another random sample of your customers, but now you want the margin of error to be smaller. What sample size would you use if you wanted the 95% margin of error to be 0.015 or less?

8.42 Sample size needed for an evaluation. You are planning an evaluation of a semester-long alcohol awareness campaign at your college. Previous evaluations indicate that about 20% of the students surveyed will respond Yes to the question "Did the campaign alter your behavior toward alcohol consumption?" How large a sample of students should you take if you want the margin of error for 95% confidence to be about 0.07?

8.43 Sample size needed for an evaluation, continued. The evaluation in the previous exercise will also have questions that have not been asked before, so you do not have previous information about the possible value of p. Repeat the preceding calculation for the following values of p^*: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. Summarize the results in a table and graphically. What sample size will you use?

8.44 Are the customers dissatisfied? An automobile manufacturer would like to know what proportion of its customers are dissatisfied with the service received from their local dealer. The customer relations department will survey a random sample of customers and compute a 95% confidence interval for the proportion who are dissatisfied. From past studies, it believes that this proportion will be about 0.25. Find the sample size needed if the margin of error of the confidence interval is to be no more than 0.035.

8.45 Sample size for coffee. Refer to Exercise 8.38 where we analyzed data from a matched pairs study that compared preferences for instant versus fresh-brewed coffee. Suppose that you want to design a similar study. The null hypothesis is that instant and fresh-brewed are equally likely to be preferred and the alternative is two-sided. You will use $\alpha = 0.05$. What is the sample size needed to detect a preference of 60% for fresh-brewed with 0.80 probability?

8.46 Sample size for tossing a coin. Refer to Exercise 8.39 where we analyzed the 10,000 coin tosses made by John Kerrich. Suppose that you want to design a study that would test the hypothesis that a coin is fair versus the alternative that the probability of a head is 0.51. Using a two-sided test with $\alpha = 0.05$. what sample size would be needed to have 0.80 power to detect this alternative?

8.2 Comparing Two Proportions

When you complete this section, you will be able to:

- Identify the counts and sample sizes for a comparison between two proportions, compute the sample proportions, and find their difference.
- Apply the guidelines for when to use the large-sample confidence interval for a difference between two proportions.
- Apply the large-sample method to find the confidence interval for a difference between two proportions and interpret the confidence interval.
- Apply the guidelines for when to use the large-sample significance test for a difference between two proportions.
- Apply the large-sample method to perform a significance test for comparing two proportions and interpret the results of the significance test.
- Find the sample size needed for a desired margin of error for the difference in proportions.
- Find the sample size needed for a significance test for comparing two proportions.
- Calculate and interpret the relative risk.

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:

Population	Population proportion	Sample size	Count of successes	Sample proportion
1	p_1	n_1	X_1	$\hat{p}_1 = X_1/n_1$
2	p_2	n_2	X_2	$\hat{p}_2 = X_2/n_2$

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

$$D = \hat{p}_1 - \hat{p}_2$$

When both sample sizes are sufficiently large, the sampling distribution of the difference D is approximately Normal.

Inference procedures for comparing proportions are z procedures based on the Normal approximation and on standardizing the difference D. The first

LOOK BACK

addition rule for means, p. 254

addition rule for variances, p. 258

step is to obtain the mean and standard deviation of D. By the addition rule for means, the mean of D is the difference of 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:

$$\begin{aligned}\sigma_D^2 &= \sigma_{\hat{p}_1}^2 + \sigma_{\hat{p}_2}^2 \\ &= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\end{aligned}$$

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

8.47 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 $p_1 - p_2$.

8.48 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?

8.49 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 a difference in proportions

To obtain a confidence interval for $p_1 - p_2$, we once again replace the unknown parameters in the standard deviation with 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 in the population proportions is

$$D = \hat{p}_1 - \hat{p}_2$$

The **standard error of *D*** is

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

and the **margin of error** for confidence level C is

$$m = z^*\mathrm{SE}_D$$

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 both 10 or more.

EXAMPLE 8.11

INSTAG

Who uses Instagram? A recent study compared the proportions of young women and men who use Instagram.[15] A total of 1069 young women and men were surveyed. These are the cases for the study. The response variable is User with values Yes and No. The explanatory variable is Sex with values "Men" and "Women." Here are the data:

Sex	n	X	$\hat{p} = X/n$
Women	537	328	0.6108
Men	532	234	0.4398
Total	1069	562	0.5257

In this table, the $\hat{p}$ column gives the sample proportions of women and men who use Instagram. The proportion for the total sample is given in the last entry in this column.

Let's find a 95% confidence interval for the difference between the proportions of women and of men who use Instagram. We first find the difference in the proportions:

$$\begin{aligned} D &= \hat{p}_1 - \hat{p}_2 \\ &= 0.6108 - 0.4398 \\ &= 0.1710 \end{aligned}$$

Then we calculate the standard error of D:

$$\begin{aligned}\mathrm{SE}_D &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\\ &= \sqrt{\frac{(0.6108)(1-0.6108)}{537} + \frac{(0.4398)(1-0.4398)}{532}}\\ &= 0.0301\end{aligned}$$

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

$$\begin{aligned}m &= z^*\mathrm{SE}_D\\ &= (1.96)(0.0301)\\ &= 0.0590\end{aligned}$$

The 95% confidence interval is

$$\begin{aligned}D \pm m &= 0.1710 \pm 0.0590\\ &= (0.112, 0.230)\end{aligned}$$

With 95% confidence, we can say that the difference in the proportions is between 0.112 and 0.230. Alternatively, we can report that the difference between the percent of women who are Instagram users and the percent of men who are Instagram users is 17.1%, with a 95% margin of error of 5.9%.

In this example, men and women were not sampled separately. The sample sizes are, in fact, random and reflect the gender distributions of the subjects who responded to the survey. Two-sample significance tests and confidence intervals are still approximately correct in this situation.

In the preceding example, we chose women to be the first population. Had we chosen men to be the first population, the estimate of the difference would be negative (-0.1710). Because it is easier to discuss positive numbers, we generally choose the first population to be the one with the higher proportion.

EXAMPLE 8.12

INSTAG

Instagram confidence interval from software. Figure 8.5 shows a spreadsheet that can be used as input to software. Output from JMP and Minitab is given in Figure 8.6. Compare these outputs with the calculations that we performed in Example 8.11.

Excel

	A	B	C
1	Sex	User	Count
2	1Women	Yes	328
3	1Women	No	209
4	2Men	Yes	234
5	2Men	No	298

FIGURE 8.5 Spreadsheet that can be used as input to software that computes the confidence interval for the Instagram data, Example 8.11.

JMP

Contingency Analysis of User By Sex

Freq: Count

Two Sample Test for Proportions

Description	Proportion Difference	Lower 95%	Upper 95%
P(Yes|1Women)-P(Yes|2Men)	0.170951	0.111429	0.2292

(a)

Minitab

```
Sample     X    N  Sample p
1        328  537  0.610801
2        234  532  0.439850


Difference = p (1) - p (2)
Estimate for difference:  0.170951
95% CI for difference:  (0.111963, 0.229940)
```

(b)

FIGURE 8.6 (a) JMP and (b) Minitab output for the Instagram confidence interval, Example 8.11.

USE YOUR KNOWLEDGE

8.50 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 54 out of 115 women and 83 out of 145 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.

8.51 Gender and commercial preference, revisited. Refer to Exercise 8.50. Construct a 95% confidence interval for the difference in proportions that favor Commercial B. Explain how you could have obtained these results from the calculations you did in Exercise 8.50.

BEYOND THE BASICS

The Plus Four Confidence Interval for a Difference in Proportions

Just as in the case of estimating a single proportion, a small modification of the sample proportions can greatly improve the accuracy of confidence intervals.[16] As before, we add two successes and two failures to the actual data, but now we divide them equally between the two samples. That is, we *add one success and one failure to each sample*. This method can be used for 90%, 95%, or 99% confidence when both sample sizes are at least five. Here is an example.

EXAMPLE 8.13

Gender and sexual maturity. In studies that look for a difference between genders, a major concern is whether or not apparent differences are due to other variables that are associated with gender. Because boys mature more slowly than girls, a study of adolescents that compares boys and girls of the same age may confuse a gender effect with an effect of sexual maturity. The "Tanner score" is a commonly used measure of sexual maturity.[17] Subjects are asked to determine their score by placing a mark next to a rough drawing of an individual at their level of sexual maturity. There are five different drawings, so the score is an integer between 1 and 5.

A pilot study included 12 girls and 12 boys from a population that will be used for a large experiment. Four of the boys and three of the girls had Tanner scores of 4 or 5, a high level of sexual maturity. Let's find a 95% confidence interval for the difference between the proportions of boys and girls who have high (4 or 5) Tanner scores in this population. The numbers of successes and failures in both groups are not all at least 10, so the large-sample approach is not recommended. On the other hand, the sample sizes are both at least 5, so the plus four method is appropriate.

The plus four estimate of the population proportion for boys is

$$\tilde{p}_1 = \frac{X_1 + 1}{n_1 + 2} = \frac{4 + 1}{12 + 2} = 0.3571$$

For girls, the estimate is

$$\tilde{p}_2 = \frac{X_2 + 1}{n_2 + 2} = \frac{3 + 1}{12 + 2} = 0.2857$$

Therefore, the estimate of the difference is

$$\tilde{D} = \tilde{p}_1 - \tilde{p}_2 = 0.3571 - 0.2857 = 0.071$$

The standard error of $\tilde{D}$ is

$$\begin{aligned} SE_{\tilde{D}} &= \sqrt{\frac{\tilde{p}_1(1 - \tilde{p}_1)}{n_1 + 2} + \frac{\tilde{p}_2(1 - \tilde{p}_2)}{n_2 + 2}} \\ &= \sqrt{\frac{(0.3571)(1 - 0.3571)}{12 + 2} + \frac{(0.2857)(1 - 0.2857)}{12 + 2}} \\ &= 0.1760 \end{aligned}$$

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

$$m = z^* SE_{\tilde{D}} = (1.96)(0.1760) = 0.345$$

The confidence interval is

$$\begin{aligned} \tilde{D} \pm m &= 0.071 \pm 0.345 \\ &= (-0.274, 0.416) \end{aligned}$$

With 95% confidence, we can say that the difference in the proportions is between −0.274 and 0.416. Alternatively, we can report that the difference in the proportions of boys and girls with high Tanner scores in this population is 7.1% with a 95% margin of error of 34.5%.

The very large margin of error in this example indicates that either boys or girls could be more sexually mature in this population and that the difference could be quite large. *Although the interval includes the possibility that there is no difference, corresponding to* $p_1 = p_2$ *or* $p_1 - p_2 = 0$, *we should not conclude that there is* ***no*** *difference in the proportions*. With small sample sizes such as these, the data do not provide us with a lot of information for our inference. This fact is expressed quantitatively through the very large margin of error.

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 σ_D. Although this approach would lead to a valid significance test, we instead adopt the more common practice of replacing the unknown σ_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

$$\begin{aligned}\sigma_D &= \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}} \\ &= \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\end{aligned}$$

We estimate the common value of p by the overall proportion of successes in the two samples:

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

pooled estimate of p

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

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

$$\mathrm{SE}_{Dp} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

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**

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\mathrm{SE}_{Dp}}$$

where the **pooled standard error** is

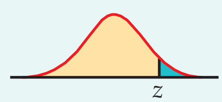
$$\mathrm{SE}_{Dp} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

and where the **pooled estimate** of the common value of p_1 and p_2 is

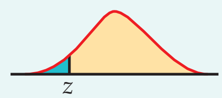
$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

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

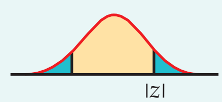
$H_a: p_1 > p_2$ is $P(Z \geq z)$



$H_a: p_1 < p_2$ is $P(Z \leq z)$



$H_a: p_1 \neq p_2$ is $2P(Z \geq |z|)$



This z test is based on the Normal approximation to the binomial distribution. As a general rule, we will use it when the number of successes and the number of failures in each of the samples are at least 5.

EXAMPLE 8.14

INSTAG

Sex and Instagram use: The z test. Are young women and men equally likely to say they use Instagram? We examine the data in Example 8.11 (page 597) to answer this question. Here is the data summary:

Sex	n	X	$\hat{p} = X/n$
Women	537	328	0.6108
Men	532	234	0.4398
Total	1069	562	0.5257

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

$$\hat{p} = \frac{328 + 234}{537 + 532} = \frac{562}{1069} = 0.5257$$

Note that this is the estimate on the bottom line of the preceding data summary. The test statistic is calculated as follows:

$$\mathrm{SE}_{Dp} = \sqrt{(0.5257)(1 - 0.5257)\left(\frac{1}{537} + \frac{1}{532}\right)} = 0.03055$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\mathrm{SE}_{Dp}} = \frac{0.6108 - 0.4398}{0.03055}$$
$$= 5.60$$

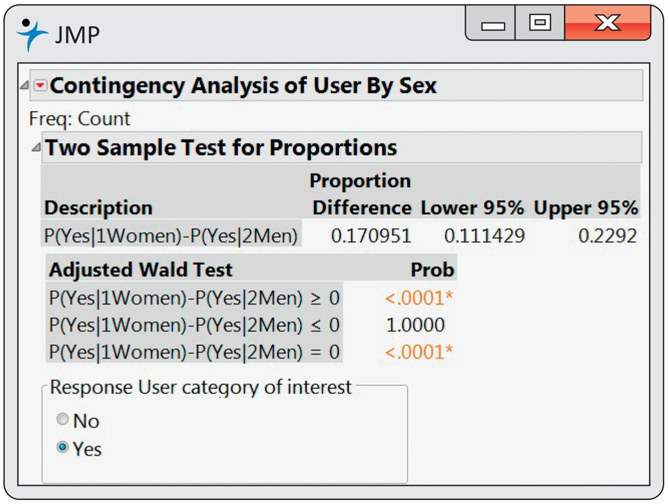
The P-value is $2P(Z \geq 5.60)$. Note that the largest value for z in Table A is 3.49. Therefore, from Table A, we can conclude that $P < 2(1 - 0.9998) = 0.0004$, although we know that the true P value is smaller.

Here is our summary: 61% of the women and 44% of the men are Instagram users; the difference is statistically significant ($z = 5.60$, $P < 0.0004$).

EXAMPLE 8.15

Output for the Instagram significance test. We prefer to use software to obtain the significance test results for comparing the Instagram use of young women and men. Output from JMP and Minitab is given in Figure 8.7. JMP reports the significance tests for the two-sided alternative and for the two one-sided alternatives. We are interested in the two-sided alternative.

FIGURE 8.7 (a) JMP and (b) Minitab output for the Instagram significance test, Example 8.15.



(a)

FIGURE 8.7 *Continued*

```
Minitab
Test and CI for Two Proportions

Sample     X    N   Sample p
1        328  537   0.610801
2        234  532   0.439850


Difference = p (1) - p (2)
Estimate for difference:   0.170951
95% CI for difference:   (0.111963, 0.229940)
Test for difference = 0 (vs ≠ 0):   Z = 5.60

Fisher's exact test: P-Value = 0.000
```

(b)

Therefore, we report the P-value as < 0.0001. Minitab reports the test statistic, $z = 5.60$, and gives the P-value as 0.000 (this means $P < 0.0005$) for the Fisher exact test. This test is an alternative to the large-sample significance test that we have discussed. It is preferred by many, particularly for small sample sizes.

Do you think that we could have argued that the proportion would be higher for women than for men before looking at the data in this example? This would allow us to use 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. Do you think that this approach is justified?

USE YOUR KNOWLEDGE

8.52 Gender and commercial preference: the z test. Refer to Exercise 8.50 (page 509). Test whether the proportions of women and men who liked Commercial A are the same versus the two-sided alternative at the 5% level.

8.53 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.

Choosing a sample size for two sample proportions

In Section 8.1, we studied methods for determining the sample size using two settings. First, we used the margin of error for a confidence interval for a single proportion as the criterion for choosing n (page xxx). Second, we used the power of the significance test for a single proportion as the determining factor (page xxx). We follow the same approach here for comparing two proportions.

Use the margin of error Recall that the large-sample estimate of the difference in proportions is

$$D = \hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

the standard error of the difference 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 z^* is the value for the standard Normal density curve with area C between $-z^*$ and z^*.

For a single proportion, we guessed a value for the true proportion and computed the margins of error for various choices of n. Here, we use the same idea but we need to guess values for the two proportions. We can display the results in a table, as in Example 8.9 (page 497), or in a graph, as in Exercise 8.43 (page 504).

SAMPLE SIZE FOR DESIRED MARGIN OF ERROR

The level C confidence interval for a difference in two proportions will have a margin of error approximately equal to a specified value m when the sample size for each of the two proportions is

$$n = \left(\frac{z^*}{m}\right)^2 (p_1^*(1-p_1^*) + p_2^*(1-p_2^*))$$

Here, z^* is the critical value for confidence C, and p_1^* and p_2^* are guessed values for p_1 and p_2, the proportions of successes in the future sample.

The margin of error will be less than or equal to m if p_1^* and p_2^* are chosen to be 0.5. The common sample size required is then given by

$$n = \left(\frac{1}{2}\right)\left(\frac{z^*}{m}\right)^2$$

Note that to use the confidence interval, which is based on the Normal approximation, we still require that the number of successes and the number of failures in each of the samples are at least 10.

EXAMPLE 8.16

Confidence interval–based sample sizes for preferences of women and men. Consider the setting in Exercise 8.50, where we compared the preferences of women and men for two commercials. Suppose we want to do a study in which we perform a similar comparison using a 95% confidence interval that will have a margin of error of 0.1 or less. What should we choose for our sample size? Using $m = 0.1$ and z^* in our formula, we have

$$n = \left(\frac{1}{2}\right)\left(\frac{z^*}{m}\right)^2 = \left(\frac{1}{2}\right)\left(\frac{1.96}{0.1}\right)^2 = 192.08$$

We would include 192 women and 192 men in our study.

Note that we have rounded the calculated value, 192.08, down because it is very close to 192. The normal procedure would be to round the calculated value up to the next larger integer.

USE YOUR KNOWLEDGE

8.54 What would the margin of error be? Consider the setting in Example 8.50.

(a) Compute the margins of error for $n_1 = 24$ and $n_2 = 24$ for each of the following scenarios: $p_1 = 0.6$, $p_2 = 0.5$; $p_1 = 0.7$, $p_2 = 0.5$; and $p_1 = 0.8$, $p_2 = 0.5$.

(b) If you think that any of these scenarios is likely to fit your study, should you reconsider your choice of $n_1 = 24$ and $n_2 = 24$? Explain your answer.

Use the power of the significance test When we studied using power to compute the sample size needed for a significance test for a single proportion, we used software. We will do the same for the significance test for comparing two proportions.

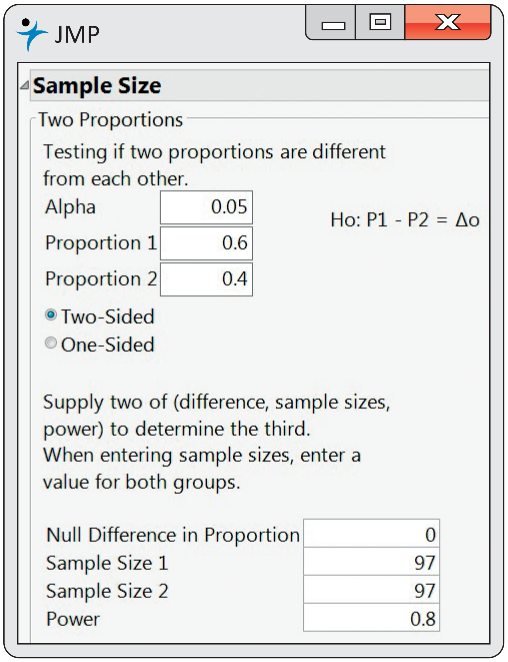
Some software allows us to consider significance tests that are a little more general than the version we studied in this section. Specifically, we used the null hypothesis H_0: $p_1 = p_2$, which we can rewrite as H_0: $p_1 - p_2 = 0$. The generalization allows us to use values different from zero in the alternative way of writing H_0. Therefore, we write H_0: $p_1 - p_2 = \Delta_0$ for the null hypothesis, and we will need to specify $\Delta_0 = 0$ for the significance test that we studied.

Here is a summary of the inputs needed for software to perform the calculations:

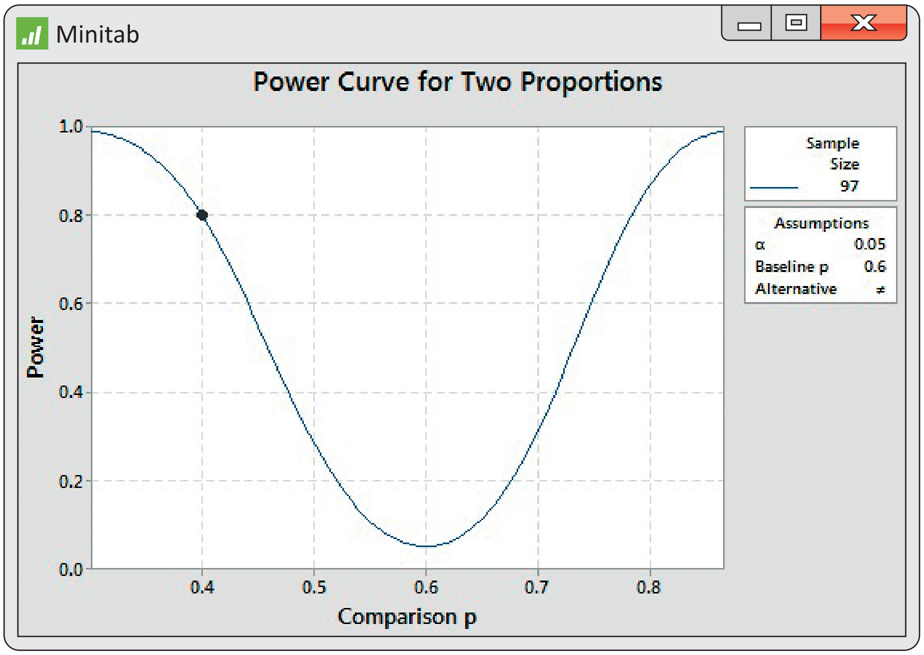
- The value of Δ_0 in the null hypothesis H_0: $p_1 - p_2 = \Delta_0$.
- The alternative hypothesis, two-sided (H_a: $p_1 \neq p_2$) or one-sided (H_a: $p_1 > p_2$ or H_a: $p_1 < p_2$).
- Values for p_1 and p_2 in the alternative hypothesis.
- The Type I error (α, the probability of rejecting the null hypothesis when it is true); usually we choose 5% ($\alpha = 0.05$) for the Type I error.
- Power (probability of rejecting the null hypothesis when it is false); usually we choose 80% (0.80) for power.

EXAMPLE 8.17

Sample sizes for preferences of women and men. Refer to Example 8.16 where we used the margin of error to find the sample sizes for comparing the preferences of women and men for two commercials. Let's find the sample sizes required for a significance test that the two proportions who prefer Commercial A are equal ($\Delta_0 = 0$) using a two-sided alternative with $p_1 = 0.6$ and $p_2 = 0.4$, $\alpha = 0.05$, and 80% (0.80) power. Outputs from JMP and Minitab are given in Figure 8.8. We need $n_1 = 97$ women and $n_2 = 97$ men for our study.



(a)



(b)

FIGURE 8.8 (a) JMP and (b) Minitab output for finding the sample size, Example 8.17.

Note that the Minitab output [Figure 8.8(b)] gives the power curve for different alternatives. All of these have $p_1 = 0.6$, which Minitab calls the “Comparison p,” while p_2 varies from 0.3 to 0.9. We see that the power is essentially 100% (1) at these extremes. It is 0.05, the type I error, at $p_2 = 0.6$, which corresponds to the null hypothesis.

USE YOUR KNOWLEDGE

8.55 Find the sample sizes. Consider the setting in Example 8.17. Change p_1 to 0.85 and p_2 to 0.90. Find the required sample sizes.

BEYOND THE BASICS

Relative Risk

We compared Instagram use for women and men by reporting the difference in the proportions with a confidence interval. Another way to compare two proportions is to take the ratio. This approach can be used in any setting and it is particularly common in medical settings.

risk

relative risk

We think of each proportion as a **risk** that something (usually bad) will happen. We then compare these two risks with the ratio of the two proportions, which is called the **relative risk** (RR). Note that a relative risk of 1 means that the two proportions, $\hat{p}_1$ and $\hat{p}_2$, are equal. The procedure for calculating confidence intervals for relative risk is based on the same kind of principles that we have studied, but the details are somewhat more complicated. Fortunately, we can leave the details to software and concentrate on interpretation and communication of the results.

EXAMPLE 8.18

Aspirin and blood clots: Relative risk. A study of patients who had blood clots (venous thromboembolism) and had completed the standard treatment were randomly assigned to receive a low-dose aspirin or a placebo treatment. The 822 patients in the study were randomized to the treatments, 411 to each. Patients were monitored for several years for the occurrence of several related medical conditions. Counts of patients who experienced one or more of these conditions were reported for each year after the study began.[18] The following table gives the data for a composite of events, termed "major vascular events." Here, X is the number of patients who had a major event.

Population	n	X	$\hat{p} = X/n$
1 (aspirin)	411	45	0.1095
2 (placebo)	411	73	0.1776
Total	822	118	0.1436

The relative risk is

$$\text{RR} = \frac{\hat{p}_1}{\hat{p}_2} = \frac{45/411}{73/411} = 0.6164$$

Software gives the 95% confidence interval as 0.4364 to 0.8707. Taking aspirin has reduced the occurrence of major events to 62% of what it is for patients taking the placebo. The 95% confidence interval is 44% to 87%.

Note that the confidence interval is not symmetric about the estimate. Relative risk is one of many situations where this occurs.

SECTION 8.2 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

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

- The **margin of error for confidence level *C*** is

$$m = z^*\text{SE}_D$$

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

$$D \pm m$$

We recommend using this interval for 90%, 95%, or 99% confidence when the number of successes and the number of failures in both samples are all at least 10. When sample sizes are smaller, alternative procedures such as the **plus four estimate of the difference in two population proportions** are recommended.

- Significance tests of H_0: $p_1 = p_2$ use the **z statistic**

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\text{SE}_{Dp}}$$

with P-values from the $N(0, 1)$ distribution. In this statistic,

$$\text{SE}_{Dp} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

and $\hat{p}$ is the **pooled estimate** of the common value of p_1 and p_2:

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

Use this test when the number of successes and the number of failures in each of the samples are at least 5.

- **Relative risk** is the ratio of two sample proportions:

$$\text{RR} = \frac{\hat{p}_1}{\hat{p}_2}$$

Confidence intervals for relative risk are often used to summarize the comparison of two proportions.

SECTION 8.2 EXERCISES

For Exercises 8.47, 8.48, and 8.49, see page 506; for Exercises 8.50 and 8.51, see page 509; for Exercises 8.52 and 8.53, see page 514; for Exercise 8.54, see page 516; and for Exercise 8.55, see page 518.

8.56 Identify the key elements. For each of the following scenarios, identify the populations, the counts, and the sample sizes; compute the two proportions and find their difference.

(a) A study of tipping behaviors examined the relationship between the color of the shirt worn by the server and whether or not the customer left a tip.[19] There were 418 male customers in the study; 40 of the 69 who were served by a server wearing a red shirt left a tip. Of the 349 who were served by a server wearing a different colored shirt, 130 left a tip.

(b) A sample of 40 runners will be used to compare two new routines for stretching. The runners will be randomly assigned to one of the routines which they will follow for two weeks. Satisfaction with the routines will be measured using a questionnaire at the end of the two-week period. For the first routine, nine runners said that they were satisfied or very satisfied. For the second routine, six runners said that they were satisfied or very satisfied.

8.57 Apply the confidence interval guidelines. Refer to the previous exercise. For each of the scenarios, determine whether or not the guidelines for using the large-sample method for a 95% confidence interval are satisfied. Explain your answers.

8.58 Find the 95% confidence interval. Refer to Exercise 8.56. For each scenario, find the large-sample 95% confidence interval for the difference in proportions and use the scenario to explain the meaning of the confidence interval.

8.59 Apply the significance test guidelines. Refer to Exercise 8.56. For each of the scenarios, determine whether or not the guidelines for using the large-sample significance test are satisfied. Explain your answers.

8.60 Perform the significance test. Refer to Exercise 8.56. For each scenario, perform the large-sample significance test and use the scenario to explain the meaning of the significance test.

8.61 Find the relative risk. Refer to Exercise 8.56. For each scenario, find the relative risk. Be sure to give a justification for your choice of proportions to use in the numerator and the denominator of the ratio. Use the scenarios to explain the meaning of the relative risk.

8.62 Teeth and military service. In 1898, the United States and Spain fought a war over the U.S. intervention in the Cuban War of Independence. At that time, the U.S. military was concerned about the nutrition of its recruits. Many did not have a sufficient number of teeth to chew the food provided to soldiers. As a result, it was likely that they would be undernourished and unable to fulfill their duties as soldiers. The requirements at that time specified that a recruit must have "at least four sound double teeth, one above and one below on each side of the mouth, and so opposed" so that they could chew food. Of the 58,952 recruits who were under the age of 20, 68 were rejected for this reason. For the 43,786 recruits who were 40 or over, 3801 were rejected.[20]

(a) Find the proportion of rejects for each age group.

(b) Find a 99% confidence interval for the difference in the proportions.

(c) Use a significance test to compare the proportions. Write a short paragraph describing your results and conclusions.

(d) Are the guidelines for the use of the large-sample approach satisfied for your work in parts (b) and (c)? Explain your answers.

8.63 Physical education requirements. In the 1920s, about 97% of U.S. colleges and universities required a physical education course for graduation. Today, about 40% require such a course. A recent study of physical education requirements included 354 institutions: 225 private and 129 public. Among the private institutions, 60 required a physical education course, while among the public institutions, 101 required a course.[21]

(a) What are the explanatory and response variables for this exercise? Justify your answers.

(b) What are the populations?

(c) What are the statistics?

(d) Use a 95% confidence interval to compare the private and the public institutions with regard to the physical education requirement.

(e) Use a significance test to compare the private and the public institutions with regard to the physical education requirement.

(f) For parts (d) and (e), verify that the guidelines for using the large-sample methods are satisfied.

(g) Summarize your analysis of these data in a short paragraph.

8.64 Exergaming in Canada. Exergames are active video games such as rhythmic dancing games, virtual bicycles, balance board simulators, and virtual sports simulators that require a screen and a console. A study of exergaming practiced by students from grades 10 and 11 in Montreal, Canada, examined many factors related to participation in exergaming.[22] Of the 358 students who reported that they stressed about their health, 29.9% said that they were exergamers. Of the 851 students who reported that they did not stress about their health, 20.8% said that they were exergamers.

(a) Define the two populations to be compared for this exercise.

(b) What are the counts, the sample sizes, and the proportions?

(c) Are the guidelines for the use of the large-sample confidence interval satisfied?

(d) Are the guidelines for the use of the large-sample significance test satisfied?

8.65 Confidence interval for exergaming in Canada. Refer to the previous exercise. Find the 95% confidence interval for the difference in proportions. Write a short statement interpreting this result.

8.66 Significance test for exergaming in Canada. Refer to Exercise 8.64. Use a significance test to compare the proportions. Write a short statement interpreting this result.

8.67 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.[23]

(a) Refer to the table that appears on 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 between 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.

8.68 Significance test for gaming on computers. Refer to the previous exercise. Test the null hypothesis that the two proportions are equal. Report the test statistic with the *P*-value and summarize your conclusion.

8.69 Gamers on computers. The report described in Exercise 8.67 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 8.67 for gaming on computers.

8.70 Significance test for gaming on consoles. Refer to the previous exercise. Test the null hypothesis that the two proportions are equal. Report the test statistic with the *P*-value and summarize your conclusion.

8.71 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. (*Hint:* Look carefully in the box giving the assumptions needed for this procedure.)

8.72 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.

8.73 Find the power. Consider testing the null hypothesis that two proportions are equal versus the two-sided alternative with $\alpha = 0.05$, 80% power, and equal sample sizes in the two groups.

(a) For each of the following situations, find the required sample size: (i) $p_1 = 0.1$ and $p_2 = 0.2$ (ii) $p_1 = 0.2$ and $p_2 = 0.3$, (iii) $p_1 = 0.3$ and $p_2 = 0.4$, (iv) $p_1 = 0.4$ and $p_2 = 0.5$, (v) $p_1 = 0.5$ and $p_2 = 0.6$, (vi) $p_1 = 0.6$ and $p_2 = 0.7$, (vii) $p_1 = 0.7$ and $p_2 = 0.8$, and (viii) $p_1 = 0.8$ and $p_2 = 0.9$.

(b) Write a short summary describing your results.

CHAPTER 8 EXERCISES

8.74 The future of gamification. Gamification is an interactive design that includes rewards such as points, payments, and gifts. A Pew survey of 1021 technology stakeholders and critics was conducted to predict the future of gamification. A report on the survey said that 42% of those surveyed thought that there would be no major increases in gamification by 2020. On the other hand, 53% said that they believed that there would be significant advances in the adoption and use of gamification by 2020.[24] Analyze these data using the methods that you learned in this chapter and write a short report summarizing your work.

8.75 Where do you get your news? A report produced by the Pew Research Center's Project for Excellence in Journalism summarized the results of a survey on how people get their news. Of the 2342 people in the survey who own a desktop or laptop, 1639 reported that they get their news from the desktop or laptop.[25]

(a) Identify the sample size and the count.

(b) Find the sample proportion and its standard error.

(c) Find and interpret the 95% confidence interval for the population proportion.

(d) Are the guidelines for use of the large-sample confidence interval satisfied? Explain your answer.

8.76 Is the calcium intake adequate? Young children need calcium in their diet to support the growth of their bones. The Institute of Medicine provides guidelines for how much calcium should be consumed by people of different ages.[26] One study examined whether or not a sample of children consumed an adequate amount of calcium based on these guidelines. Because there are different guidelines for children aged 5 to 10 years and those aged 11 to 13 years, the children were classified into these two age groups. Each student's calcium intake was classified as meeting or not meeting the guideline. There were 2029 children in the study. Here are the data:[27]

	Age (years)	
Met requirement	**5 to 10**	**11 to 13**
No	194	557
Yes	861	417

Identify the populations, the counts, and the sample sizes for comparing the extent to which the two age groups of children met the calcium intake requirement.

8.77 Use a confidence interval for the comparison. Refer to the previous exercise. Use a 95% confidence interval for the comparison and explain what the confidence interval tells us. Be sure to include a justification for the use of the large-sample procedure for this comparison.

8.78 Use a significance test for the comparison. Refer to Exercise 8.76. Use a significance test to make the comparison. Interpret the result of your test. Be sure to include a justification for the use of the large-sample procedure for this comparison.

8.79 Confidence interval or significance test? Refer to Exercises 8.76, 8.77, and 8.78. Do you prefer to use the confidence interval or the significance test for this comparison? Give reasons for your answer.

8.80 Changing majors during college. In a random sample of 975 students from a large public university, it was found that 463 of the students changed majors during their college years.

(a) Give a 95% confidence interval for the proportion of students at this university who change majors.

(b) Express your results from part (a) in terms of the *percent* of students who change majors.

(c) University officials concerned with counseling students are interested in the number of students who change majors rather than the proportion. The university has 37,500 undergraduate students. Convert the confidence interval you found in part (a) to a confidence interval for the *number* of students who change majors during their college years.

8.81 Facebook users. A Pew survey of 1802 Internet users found that 67% use Facebook.[28]

(a) How many of those surveyed used Facebook?

(b) Give a 95% confidence interval for the proportion of Internet users who use Facebook.

(c) Convert the confidence interval that you found in part (b) to a confidence interval for the percent of Internet users who use Facebook.

8.82 Twitter users. Refer to the previous exercise. The same survey reported that 16% of Internet users use Twitter. Answer the questions in the previous exercise for Twitter use.

8.83 Facebook versus Twitter. Refer to Exercises 8.81 and 8.82. Can you use the data provided in these two exercises to compare the proportion of Facebook users with the proportion of Twitter users? If your answer is Yes, do the comparison. If your answer is No, explain why you cannot make the comparison.

8.84 Video game genres. U.S. computer and video game software sales were $13.26 billion in 2012.[29]

A survey of 1102 teens collected data about video game use by teens. According to the survey, the following are the most popular game genres:[30]

Genre	Examples	Percent who play
Racing	NASCAR, Mario Kart, Burnout	74
Puzzle	Bejeweled, Tetris, Solitaire	72
Sports	Madden, FIFA, Tony Hawk	68
Action	Grand Theft Auto, Devil May Cry, Ratchet and Clank	67
Adventure	Legend of Zelda, Tomb Raider	66
Rhythm	Guitar Hero, Dance Dance Revolution, Lumines	61

Give a 95% confidence interval for the proportion who play games in each of these six genres.

8.85 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 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 $0.05/6 = 0.008$. So, each confidence interval should have a 0.8% chance of missing the true value. In other words, the confidence level for each interval should be $1 - 0.008 = 0.992$. Use Table A to find the value of z^* for a large-sample confidence interval for a single proportion corresponding to 99.2% confidence.

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

8.86 Changes in credit card usage by undergraduates. In Exercise 8.32 (page 503), we looked at data from a survey of 1430 undergraduate students and their credit card use. In the sample, 43% said that they had four or more credit cards. A similar study performed four years earlier by the same organization reported that 32% of the sample said that they had four or more credit cards.[31] Assume that the sample sizes for the two studies are the same. Find a 95% confidence interval for the change in the percent of undergraduates who report having four or more credit cards.

8.87 Do the significance test for the change. Refer to the previous exercise. Perform the significance test for comparing the two proportions. Report your test statistic, the *P*-value, and summarize your conclusion.

8.88 We did not know the sample size. Refer to the previous two exercises. We did not report the sample size for the earlier study, but it is reasonable to assume that it is close to the sample size for the later study.

(a) Suppose that the sample size for the earlier study was only 800. Redo the confidence interval and significance test calculations for this scenario.

(b) Suppose that the sample size for the earlier study was 2500. Redo the confidence interval and significance test calculations for this scenario.

(c) Compare your results for parts (a) and (b) of this exercise with the results that you found in the previous two exercises. Write a short paragraph about the effects of assuming a value for the sample size on your conclusions.

8.89 Student employment during the school year. A study of 1530 undergraduate students reported that 1006 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.

8.90 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.

8.91 Gender and soft drink consumption. Refer to Exercise 8.26 (page 502). This survey found that 16% of the 2006 New Zealanders surveyed reported that they consumed five or more servings of soft drinks per week. The corresponding percents for men and women were 17% and 15%, respectively. Assuming that the numbers of men and women in the survey are approximately equal, do the data suggest that the proportions vary by gender? Explain your methods, assumptions, results, and conclusions.

8.92 Examine the effect of the sample size. Refer to the previous exercise. Assume the following values for the total sample size: 1000, 4000, 10,000. Also assume that the sample proportions do not change. For each of these scenarios, redo the calculations that you performed in the previous exercise. Write a short paragraph summarizing the effect of the sample size on the results.

8.93 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 $\hat{p}_1 = 0.65$ and $\hat{p}_2 = 0.45$, 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: 60, 70, 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.

8.94 Sample size and the margin of error. In Section 8.1, we studied the effect of the sample size on the margin of error of the confidence interval for a single proportion. In this exercise, we perform some calculations to observe this effect for the two-sample problem. Suppose that $\hat{p}_1 = 0.7$ and $\hat{p}_2 = 0.5$ and n represents the common value of n_1 and n_2. Compute the 95% margins of error for the difference between the two proportions for n = 60, 70, 80, 100, 400, 500, and 1000. Present the results in a table and with a graph. Write a short summary of your findings.

8.95 Calculating sample sizes for the two-sample problem. For a single proportion, the margin of error of a confidence interval is largest for any given sample size n and confidence level C when $\hat{p} = 0.5$. This led us to use $p^* = 0.5$ for planning purposes. The same kind of result is true for the two-sample problem. The margin of error of the confidence interval for the difference between two proportions is largest when $\hat{p}_1 = \hat{p}_2 = 0.5$. You are planning a survey and will calculate a 95% confidence interval for the difference between two proportions when the data are collected. You would like the margin of error of the interval to be less than or equal to 0.055. You will use the same sample size n for both populations.

(a) How large a value of n is needed?

(b) Give a general formula for n in terms of the desired margin of error m and the critical value z^*.

8.96 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 three 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.[32] How statistically significant is the evidence? What assumptions does your analysis require? Do these assumptions seem reasonable in this case?

8.97 Statistics and the law. *Castaneda v. Partida* is an important court case in which statistical methods were used as part of a legal argument.[33] When reviewing this case, the Supreme Court used the phrase "two or three standard deviations" as a criterion for statistical significance. This Supreme Court review has served as the basis for many subsequent applications of statistical methods in legal settings. (The two or three standard deviations referred to by the Court are values of the z statistic and correspond to P-values of approximately 0.05 and 0.0026.) In *Castaneda,* the plaintiffs alleged that the method for selecting juries in a county in Texas was biased against Mexican Americans. For the period of time at issue, there were 181,535 persons eligible for jury duty, of whom 143,611 were Mexican Americans. Of the 870 people selected for jury duty, 339 were Mexican Americans.

(a) What proportion of eligible jurors were Mexican Americans? Let this value be p_0.

(b) Let p be the probability that a randomly selected juror is a Mexican American. The null hypothesis to be tested is H_0: $p = p_0$. Find the value of $\hat{p}$ for this problem, compute the z statistic, and find the P-value. What do you conclude? (A finding of statistical significance in this circumstance does not constitute proof of discrimination. It can be used, however, to establish a prima facie case. The burden of proof then shifts to the defense.)

(c) We can reformulate this exercise as a two-sample problem. Here we wish to compare the proportion of Mexican Americans among those selected as jurors with the proportion of Mexican Americans among those not selected as jurors. Let p_1 be the probability that a randomly selected juror is a Mexican American and let p_2 be the probability that a randomly selected nonjuror is a Mexican American. Find the z statistic and its P-value. How do your answers compare with your results in part (b)?