Inference about a Population Proportion

Our discussion of statistical inference to this point has concerned making inferences about population means. Now we turn to questions about the proportion of some outcome in a population. Here are some examples that call for inference about population proportions.

EXAMPLE 18.1 Risksy behavior in the age of AIDS

How common is behavior that puts people at risk of AIDS? The National AIDS Behavioral Surveys interviewed a random sample of 2673 adult heterosexuals. Of these, 170 had more than one sexual partner in the past year. That’s 6.36% of the sample. Based on these data, what can we say about the percent of all adult heterosexuals who have multiple partners? We want to estimate a single population proportion. This chapter concerns inference about one proportion.

EXAMPLE 18.2 Does preschool make a difference?

Do preschool programs for poor children make a difference in later life? A study looked at 62 children who were enrolled in a Michigan preschool in the late 1960s and at a control group of 61 similar children who were not enrolled. At 27 years of age, 61% of the preschool group and 80% of the control group had required the help
of a social service agency (mainly welfare) in the previous ten years. Is this significant evidence that preschool for poor children reduces later use of social services? We want to compare two population proportions. This is the topic of Chapter 19.

To do inference about a population mean \( \mu \), we use the mean \( \bar{x} \) of a random sample from the population. The reasoning of inference starts with the sampling distribution of \( \bar{x} \). Now we follow the same pattern, replacing means by proportions.

**The sample proportion \( \hat{p} \)**

We are interested in the unknown proportion \( p \) of a population that has some outcome. For convenience, call the outcome we are looking for a “success.” In Example 18.1, the population is adult heterosexuals, and the parameter \( p \) is the proportion who have had more than one sexual partner in the past year. To estimate \( p \), the National AIDS Behavioral Surveys used random dialing of telephone numbers to contact a sample of 2673 people. Of these, 170 said they had multiple sexual partners. The statistic that estimates the parameter \( p \) is the sample proportion

\[
\hat{p} = \frac{\text{count of successes in the sample}}{\text{count of observations in the sample}} = \frac{170}{2673} = 0.0636
\]

Read the sample proportion \( \hat{p} \) as “p-hat.”

**APPLY YOUR KNOWLEDGE**

In each of the following settings:

(a) Describe the population and explain in words what the parameter \( p \) is.
(b) Give the numerical value of the statistic \( \hat{p} \) that estimates \( p \).

18.1 Do college students pray? A study of religious practices among college students interviewed a sample of 127 students; 107 of the students said that they prayed at least once in a while.

18.2 Information online. A random sample of 1318 Internet users was asked where they will go for information the next time they need information about health or medicine; 606 said that they would use the Internet.

**The sampling distribution of \( \hat{p} \)**

How good is the statistic \( \hat{p} \) as an estimate of the parameter \( p \)? To find out, we ask, “What would happen if we took many samples?” The sampling distribution of \( \hat{p} \) answers this question. Here are the facts.
The sampling distribution of \( \hat{p} \)

**SAMPLING DISTRIBUTION OF A SAMPLE PROPORTION**

Choose an SRS of size \( n \) from a large population that contains population proportion \( p \) of “successes.” Let \( \hat{p} \) be the sample proportion of successes,

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

Then:

- As the sample size increases, the sampling distribution of \( \hat{p} \) becomes approximately Normal.
- The mean of the sampling distribution is \( p \).
- The standard deviation of the sampling distribution is

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

Figure 18.1 summarizes these facts in a form that helps you recall the big idea of a sampling distribution. The behavior of sample proportions \( \hat{p} \) is similar to the behavior of sample means \( \bar{x} \). When the sample size \( n \) is large, the sampling distribution is approximately Normal. The larger the sample, the more nearly Normal the distribution is. The mean of the sampling distribution is the true value of the population proportion \( p \). That is, \( \hat{p} \) is an unbiased estimator of \( p \).

The standard deviation of \( \hat{p} \) gets smaller as the sample size \( n \) gets larger, so that estimation is likely to be more accurate when the sample is larger. But the standard deviation gets smaller only at the rate \( \sqrt{n} \). We need four times as many observations to cut the standard deviation in half.

You should not use the Normal approximation to the distribution of \( \hat{p} \) when the sample size \( n \) is small. What is more, the formula given for the standard
deviation of \( \hat{p} \) is not accurate unless the population is much larger than the sample—say, at least 10 times larger. We will give guidelines to help you decide when methods for inference based on this sampling distribution are trustworthy.

**EXAMPLE 18.3** Asking about risky behavior

Suppose that in fact 6% of all adult heterosexuals had more than one sexual partner in the past year (and would admit it when asked). The National AIDS Behavioral Surveys interviewed a random sample of 2673 people from this population. What is the probability that at least 5% of such a sample admit to having more than one partner?

If the sample size is \( n = 2673 \) and the population proportion is \( p = 0.06 \), the sample proportion \( \hat{p} \) has mean 0.06 and standard deviation

\[
\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.06(0.94)}{2673}} = \sqrt{0.00000211} = 0.00459
\]

We want the probability that \( \hat{p} \) is 0.05 or greater. Standardize \( \hat{p} \) by subtracting the mean 0.06 and dividing by the standard deviation 0.00459. This produces a new statistic that has approximately the standard Normal distribution. As usual, we call this statistic \( z \):

\[
z = \frac{\hat{p} - 0.06}{0.00459}
\]

Figure 18.2 shows the probability we want as an area under the standard Normal curve.

\[
P(\hat{p} \geq 0.05) = P \left( z \geq \frac{0.05 - 0.06}{0.00459} \right)
\]

\[
= P(Z \geq -2.18) = 1 - 0.0146 = 0.9854
\]

If we repeat the National AIDS Behavioral Surveys many times, more than 98% of all the samples will contain at least 5% of respondents who admit to more than one sexual partner.

The Normal approximation for the sampling distribution of \( \hat{p} \) works poorly when \( p \) is close to 0. You can see that if \( p = 0 \), any sample must contain only failures. That is, \( \hat{p} = 0 \) always and there is no Normal distribution in sight. In just the same way, the approximation works poorly when \( p \) is close to 1. In practice, we need larger \( n \) for values of \( p \) near 0 or 1.

**APPLY YOUR KNOWLEDGE**

18.3 Student drinking. The College Alcohol Study interviewed an SRS of 14,941 college students about their drinking habits. Suppose that
half of all college students “drink to get drunk” at least once in a while. That is, \( p = 0.5 \).

(a) What are the mean and standard deviation of the proportion \( \hat{p} \) of the sample who drink to get drunk?

(b) Use the Normal approximation to find the probability that \( \hat{p} \) is between 0.49 and 0.51.

18.4 Harley motorcycles. Harley-Davidson motorcycles make up 14% of all the motorcycles registered in the United States. You plan to interview an SRS of 500 motorcycle owners.

(a) What is the approximate distribution of the proportion of your sample who own Harleys?

(b) Is your sample likely to contain 20% or more who own Harleys? Is it likely to contain at least 15% Harley owners? Do Normal probability calculations to answer these questions.

18.5 Student drinking, continued. Suppose that half of all college students drink to get drunk at least once in a while. Exercise 18.3 asks the probability that the sample proportion \( \hat{p} \) estimates \( p = 0.5 \) within ±1 percentage point. Find this probability for SRSs of sizes 1000, 4000, and 16,000. What general fact do your results illustrate?

**Conditions for inference**

Inference about a population proportion \( p \) is based on the sampling distribution of the sample proportion \( \hat{p} \). More specifically, we use the \( z \) statistic that results from standardizing \( \hat{p} \):

\[
z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}
\]
The statistic $z$ has approximately the standard Normal distribution $N(0, 1)$ if the sample is not too small and the sample is not a large part of the entire population.

To test the hypothesis $H_0: p = p_0$, use the test statistic

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

This statistic has approximately the standard Normal distribution when $H_0$ is true.

To obtain a level $C$ confidence interval for $p$, we would like to use

$$\hat{p} \pm z^* \sqrt{\frac{p(1 - p)}{n}}$$

with the critical value $z^*$ chosen to cover the central area $C$ under the standard Normal curve. Figure 18.3 shows why. Because we don’t know the value of $p$, we replace the standard deviation by the standard error of $\hat{p}$

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

to get a confidence interval of the familiar form

$$\text{estimate} \pm z^*SE_{\text{estimate}}$$

Figure 18.3 With probability $C$, $\hat{p}$ lies within $\pm z^* \sqrt{\hat{p}(1 - \hat{p})/n}$ of the unknown population proportion $p$. That is, $p$ lies within $\pm z^* \sqrt{p(1 - p)/n}$ of $\hat{p}$ in those samples.
When we estimate a mean \( \mu \), there is a separate parameter \( \sigma \) that describes the spread of the distribution. Estimating \( \sigma \) gave us the one-sample \( t \) statistic.

When we estimate a proportion \( p \), there is just one parameter, and the standard deviation of \( \hat{p} \) depends on \( p \). We don’t get a \( t \) statistic—we just make the Normal approximation less accurate when we replace \( p \) by \( \hat{p} \). Here is a summary of the conditions we need for inference.

### CONDITIONS FOR INFERENCE ABOUT A PROPORTION

- The data are an SRS from the population of interest. This is, as usual, the most important condition.
- The population is at least 10 times as large as the sample. This condition ensures that the standard deviation of \( \hat{p} \) is close to \( \sqrt{\hat{p}(1-\hat{p})/n} \).
- The sample size \( n \) is large enough to ensure that the distribution of \( z \) is close to standard Normal. We will see that different inference procedures require different answers to the question “how large is large enough?”

### APPLY YOUR KNOWLEDGE

18.6 **No inference.** Tonya wants to estimate what proportion of the students in her dormitory like the dorm food. She interviews an SRS of 50 of the 175 students living in the dormitory. She finds that 14 think the dorm food is good. Tonya can’t use the methods of this chapter to get a confidence interval. Why not?

18.7 **No inference.** A television news program conducts a call-in poll about a proposed city ban on handgun ownership. Of the 2372 calls, 1921 oppose the ban. We can’t use these data as the basis for inference about the proportion of all citizens who oppose the ordinance. Why not?

### Large-sample confidence intervals for a proportion

Here is the basic confidence interval for a proportion. Unfortunately, this interval can be trusted only for quite large samples. Our rule of thumb for “how large” takes into account the fact that \( n \) must be larger if the sample proportion \( \hat{p} \) suggests that \( p \) may be close to 0 or 1. Because \( n\hat{p} \) is just the count of successes in the sample and \( n(1-\hat{p}) \) is the count of failures, we can use these counts to get a simple guideline for “\( n \) is large and \( \hat{p} \) is not too close to 0 or 1.”
CHAPTER 18 • Inference about a Population Proportion

**LARGE-SAMPLE CONFIDENCE INTERVAL FOR A POPULATION PROPORTION**

Draw an SRS of size \( n \) from a population with unknown proportion \( p \) of successes. An approximate level \( C \) confidence interval for \( p \) is

\[
\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}
\]

where \( z^* \) is the critical value for the standard Normal density curve with area \( C \) between \(-z^*\) and \( z^*\).

Use this interval only when the counts of successes and failures in the sample are both at least 15.\(^9\)

**EXAMPLE 18.4** Estimating risky behavior

The National AIDS Behavioral Surveys found that 170 of a sample of 2673 adult heterosexuals had multiple partners. That is, \( \hat{p} = 0.0636 \). We will act as if the sample were an SRS.

A 99% confidence interval for the proportion \( p \) of all adult heterosexuals with multiple partners uses the standard Normal critical value \( z^* = 2.576 \). (Look in the bottom row of Table C for standard Normal critical values.) The confidence interval is

\[
\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 0.0636 \pm 2.576 \sqrt{\frac{0.0636 \times 0.9364}{2673}} = 0.0636 \pm 0.0122 = 0.0514 \text{ to } 0.0758
\]

We are 99% confident that the percent of adult heterosexuals who had more than one sexual partner in the past year lies between about 5% and 7.6%.

**EXAMPLE 18.5** Are the conditions met?

We used the National AIDS Behavioral Surveys data to give a confidence interval for the proportion of adult heterosexuals who have had multiple sexual partners. Does the sample meet the requirements for inference?

- The sampling design was a complex stratified sample, and the survey used inference procedures for that design. The overall effect is close to an SRS, however.
- The number of adult heterosexuals (the population) is much larger than 10 times the sample size, \( n = 2673 \).
- The numbers of successes (170) and failures (2503) in the sample are both much larger than 15.
The second and third requirements are easily met. The first requirement, that the sample be an SRS, is only approximately met. As usual, the practical problems of a large sample survey pose a greater threat to the AIDS survey’s conclusions. Only people in households with telephones could be reached. This is acceptable for surveys of the general population, because about 94% of American households have telephones. However, some groups at high risk for AIDS, like intravenous drug users, often don’t live in settled households and are underrepresented in the sample. About 30% of the people reached refused to cooperate. A nonresponse rate of 30% is not unusual in large sample surveys, but it may cause some bias if those who refuse differ systematically from those who cooperate. The survey used statistical methods that adjust for unequal response rates in different groups. Finally, some respondents may not have told the truth when asked about their sexual behavior. The survey team tried hard to make respondents feel comfortable. For example, Hispanic women were interviewed only by Hispanic women, and Spanish speakers were interviewed by Spanish speakers with the same regional accent (Cuban, Mexican, or Puerto Rican). Nonetheless, the survey report says that some bias is probably present:

It is more likely that the present figures are underestimates; some respondents may underreport their numbers of sexual partners and intravenous drug use because of embarrassment and fear of reprisal, or they may forget or not know details of their own or of their partner’s HIV risk and their antibody testing history.³

Reading the report of a large study like the National AIDS Behavioral Surveys reminds us that statistics in practice involves much more than recipes for inference.

**APPLY YOUR KNOWLEDGE**

18.8 **No confidence interval.** In the National AIDS Behavioral Surveys sample of 2673 adult heterosexuals, 0.2% (that’s 0.002 as a decimal fraction) had both received a blood transfusion and had a sexual partner from a group at high risk of AIDS. Explain why we can’t use the large-sample confidence interval to estimate the proportion \( p \) in the population who share these two risk factors.

18.9 **How common is SAT coaching?** A random sample of students who took the SAT college entrance examination twice found that 427 of the respondents had paid for coaching courses and that the remaining 2733 had not.⁶ Give a 99% confidence interval for the proportion of coaching among students who retake the SAT.

18.10 **The millennium begins with optimism.** In January of the year 2000, a Gallup Poll asked a random sample of 1633 adults, “In general, are
you satisfied or dissatisfied with the way things are going in the United States at this time?" It found that 1127 said that they were satisfied. Write a short report of this finding, as if you were writing for a newspaper. Be sure to include a margin of error.

18.11 Teens and their TV sets. The New York Times and CBS News conducted a nationwide survey of 1048 randomly selected 13- to 17-year-olds. Of these teenagers, 692 had a television in their room. (a) Check that we can use the large-sample confidence interval. (b) Give a 95% confidence interval for the proportion of all teens who have a TV set in their room. (c) The news article says, "In theory, in 19 cases out of 20, the survey results will differ by no more than three percentage points in either direction from what would have been obtained by seeking out all American teenagers." Explain how your results agree with this statement.

Accurate confidence intervals for a proportion*

The confidence interval $\hat{p} \pm z^* \sqrt{\hat{p}(1 - \hat{p})/n}$ for a sample proportion $\hat{p}$ is easy to calculate. It is also easy to understand because it rests directly on the approximately Normal distribution of $\hat{p}$. Unfortunately, this interval is often quite inaccurate unless the sample is very large. More specifically, the actual confidence level is usually less than the confidence level you asked for in choosing the critical value $z^*$. That’s bad. What is worse, accuracy does not consistently get better as the sample size $n$ increases. There are "lucky" and "unlucky" combinations of $n$ and the true population proportion $p$. Here is a quote from a recent computational study (the “standard interval” is our large-sample interval):

For instance, when $n$ is 100, the actual coverage probability of the nominal 95% standard interval is 0.952 if $p$ is 0.106, but only 0.911 if $p$ is 0.107. The behavior of the coverage probability can be even more erratic as a function of $n$. If the true $p$ is 0.5, the actual coverage of the nominal 95% interval is 0.953 at the rather small sample size $n = 14$, but falls to 0.919 at the much larger sample size $n = 40$.

What should we do? Fortunately, there is a simple modification that is almost magically effective in improving the accuracy of the confidence interval. We call it the “plus four” method, because all you need to do is add four imaginary observations, two successes and two failures. If you have $X$ successes in your $n$ observations, act as if you had $X+2$ successes in $n+4$ observations. Here is the formula.*

*Although this material is optional, it is essential unless you calculate confidence intervals for a proportion only for very large samples.
ACCURATE CONFIDENCE INTERVALS FOR A PROPORTION

Choose an SRS of size $n$ from a large population that contains population proportion $p$ of “successes.” The plus four estimate of $p$ is

$$\hat{p} = \frac{\text{count of successes in the sample} + 2}{n + 4}$$

An approximate level $C$ confidence interval for $p$ is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n + 4}}$$

where $z^*$ is the critical value for the standard Normal density curve with area $C$ between $-z^*$ and $z^*$.

Use this interval when $C$ is at least 90% and the sample size $n$ is at least 10.

EXAMPLE 18.6 Shaq’s free throws

Shaquille O’Neal of the Los Angeles Lakers, the dominant center in professional basketball, has one weakness: he is a poor free-throw shooter. In his career prior to the 2000 season, Shaq made just 53.3% of his free throws. Before that season, he worked with a coach on his technique. In the first two games of the following season, he made 26 out of 39 free throws.

We can consider these as an SRS of size 39 from the population of free throws shot under game conditions with the new technique. We want a 95% confidence interval for the proportion $p$ of free throws that Shaq will make.

First, let’s calculate the large-sample interval. The sample proportion is

$$\hat{p} = \frac{26}{39} = 0.667$$

and the confidence interval based on 26 successes in 39 observations is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 0.667 \pm 1.96 \sqrt{\frac{0.667(0.333)}{39}}$$

$$= 0.667 \pm 0.148$$

We can’t trust this result. The plus four estimate of $p$ is

$$\tilde{p} = \frac{26 + 2}{39 + 4} = \frac{28}{43} = 0.651$$

The plus four confidence interval is the same as the large-sample interval based on 28 successes in 43 observations. Here it is:

$$\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}} = 0.651 \pm 1.960 \sqrt{\frac{0.651(0.349)}{43}}$$

$$= 0.651 \pm 0.142$$
We estimate with 95% confidence that Shaq will make between 50.9% and 79.3% of his free throws with the new technique. The second interval is accurate because we used the plus four method. It is too wide to be very helpful because the sample is small. We should wait for more data.

How much more accurate is the plus four interval? Computer studies have asked how large \( n \) must be to guarantee that the actual probability that a 95% confidence interval covers the true parameter value is at least 0.94 for all larger samples. If \( p = 0.1 \), for example, the answer is \( n = 646 \) for the large-sample interval and \( n = 11 \) for the plus four interval. The consensus of computational and theoretical studies is that plus four is very much better than the large-sample interval for many combinations of \( n \) and \( p \). We recommend that you always use the plus four interval.

**APPLY YOUR KNOWLEDGE**

**18.12 Drug-detecting rats?** Dogs are big and expensive. Rats are small and cheap. Might rats be trained to replace dogs in sniffing out illegal drugs? A first study of this idea trained rats to rear up on their hind legs when they smelled simulated cocaine. To see how well rats performed after training, they were let loose on a surface with many cups sunk in it, one of which contained simulated cocaine. Four out of six trained rats succeeded in 80 out of 80 trials. If a rat succeeds in every one of 80 trials, how should we estimate that rat’s long-term success rate \( p \)?

(a) Explain why the estimate \( \hat{p} = \frac{80}{80} = 1 \) is almost certainly too high.

(b) Find the plus four estimate \( \tilde{p} \). It is more reasonable. This example shows how \( \tilde{p} \) improves on \( \hat{p} \) when a sample has almost all successes or almost all failures. That’s why the guidelines for using the plus four interval can ignore the “near 0 or 1” issue.

**18.13 High-risk behavior.** In the National AIDS Behavioral Surveys sample of 2673 adult heterosexuals, 5 respondents had both received a blood transfusion and had a sexual partner from a group at high risk of AIDS. (a) You should not use the large-sample confidence interval for the proportion \( p \) in the population who share these two risk factors. Why not?

(b) The plus four method adds four observations, two successes and two failures. What are the sample size and the count of successes after you do this? What is the plus four estimate \( \tilde{p} \) of \( p \)?

(c) Give the plus four 95% confidence interval for \( p \).

**18.14 Fear of crime among older black women.** The elderly fear crime more than younger people, even though they are less likely to be victims of crime. One of the few studies that looked at older blacks recruited a random sample of 56 black women over the age of 65 from
Choosing the sample size

Atlantic City, New Jersey. Of these women, 27 said that they “felt vulnerable” to crime.12

(a) Give the two estimates $\hat{p}$ and $\tilde{p}$ of the proportion $p$ of all elderly black women in Atlantic City who feel vulnerable to crime. There is little difference between them. This is generally true when $\hat{p}$ is not close to either 0 or 1.

(b) Give both the large-sample 95% confidence interval and the plus four 95% confidence interval for $p$. The plus four interval is a bit narrower. This is generally true when $\hat{p}$ is not close to either 0 or 1.

18.15 Do college students pray? Social scientists asked 127 undergraduate students “from courses in psychology and communications” about prayer and found that 107 prayed at least a few times a year.13

(a) Give the plus four 99% confidence interval for the proportion $p$ of all students who pray.

(b) To use any inference procedure, we must be willing to regard these 127 students, as far as their religious behavior goes, as an SRS from the population of all undergraduate students. Do you think it is reasonable to do this? Why or why not?

Choosing the sample size

In planning a study, we may want to choose a sample size that will allow us to estimate the parameter within a given margin of error. We saw earlier (page 331) how to do this for a population mean. The method is similar for estimating a population proportion.

The margin of error in the large-sample confidence interval for $p$ is

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

Here $z^*$ is the standard Normal critical value for the level of confidence we want. Because the margin of error involves the sample proportion of successes $\hat{p}$, we need to guess this value when choosing $n$. Call our guess $p^*$. Here are two ways to get $p^*$:

1. Use a guess $\hat{p}$ based on a pilot study or on past experience with similar studies. You should do several calculations that cover the range of $\hat{p}$-values you might get.

2. Use $p^* = 0.5$ as the guess. The margin of error $m$ is largest when $\hat{p} = 0.5$, so this guess is conservative in the sense that if we get any other $\hat{p}$ when we do our study, we will get a margin of error smaller than planned.

Once you have a guess $p^*$, the recipe for the margin of error can be solved to give the sample size $n$ needed. Here is the result for the large-sample confidence interval. For simplicity, use this result even if you plan to use the plus four interval.
CHAPTER 18 • Inference about a Population Proportion

SAMPLE SIZE FOR DESIRED MARGIN OF ERROR

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

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

where \( p^* \) is a guessed value for the sample proportion. The margin of error will be less than or equal to \( m \) if you take the guess \( p^* \) to be 0.5.

Which method for finding the guess \( p^* \) should you use? The \( n \) you get doesn't change much when you change \( p^* \) as long as \( p^* \) is not too far from 0.5. So use the conservative guess \( p^* = 0.5 \). If the true \( \hat{p} \) is close to 0 or 1, using \( p^* = 0.5 \) as your guess will give a sample much larger than you need. So try to use a better guess from a pilot study when you suspect that \( \hat{p} \) will be less than 0.3 or greater than 0.7.

EXAMPLE 18.7 Planning a poll

Gloria Chavez and Ronald Flynn are the candidates for mayor in a large city. You are planning a sample survey to determine what percent of the voters plan to vote for Chavez. This is a population proportion \( p \). You will contact an SRS of registered voters in the city. You want to estimate \( p \) with 95% confidence and a margin of error no greater than 3%, or 0.03. How large a sample do you need?

The winner’s share in all but the most lopsided elections is between 30% and 70% of the vote. So use the guess \( p^* = 0.5 \). The sample size you need is

\[
n = \left( \frac{1.96}{m} \right)^2 (0.5)(1 - 0.5) = 1067.1
\]

You should round the result up to \( n = 1068 \). (Rounding down would give a margin of error slightly greater than 0.03.) If you want a 2.5% margin of error, we have (after rounding up)

\[
n = \left( \frac{1.96}{0.025} \right)^2 (0.5)(1 - 0.5) = 1537
\]

For a 2% margin of error the sample size you need is

\[
n = \left( \frac{1.96}{0.02} \right)^2 (0.5)(1 - 0.5) = 2401
\]

As usual, smaller margins of error call for larger samples.

APPLY YOUR KNOWLEDGE

18.16 Canadians and doctor-assisted suicide. A Gallup Poll asked a sample of Canadian adults if they thought the law should allow doctors to end the life of a patient who is in great pain and near death if the patient
makes a request in writing. The poll included 270 people in Quebec, 221 of whom agreed that doctor-assisted suicide should be allowed.14

(a) What is the margin of error of the large-sample 95% confidence interval for the proportion of all Quebec adults who would allow doctor-assisted suicide?
(b) How large a sample is needed to get the common ±3 percentage point margin of error?

18.17 Can you taste PTC? PTC is a substance that has a strong bitter taste for some people and is tasteless for others. The ability to taste PTC is inherited. About 75% of Italians can taste PTC, for example. You want to estimate the proportion of Americans with at least one Italian grandparent who can taste PTC. Starting with the 75% estimate for Italians, how large a sample must you test in order to estimate the proportion of PTC tasters within ±0.04 with 90% confidence?

Significance tests for a proportion

Tests of hypotheses about a population proportion \( p \) are based on the sampling distribution of the sample proportion \( \hat{p} \), taking the null hypothesis to be true. Because \( H_0 \) fixes a value of \( p \), the inaccuracy that plagues the large-sample confidence interval does not affect tests. Here is the rule for tests.

**SIGNIFICANCE TESTS FOR A PROPORTION**

To test the hypothesis \( H_0: p = p_0 \), compute the \( z \) statistic

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

In terms of a variable \( Z \) having the standard Normal distribution, the approximate \( P \)-value for a test of \( H_0 \) against

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

Use this test when the sample size \( n \) is so large that both \( np_0 \) and \( n(1 - p_0) \) are 10 or more.
EXAMPLE 18.8 Is this coin fair?

A coin that is balanced should come up heads half the time in the long run. The population for coin tossing contains the results of tossing the coin forever. The parameter \( p \) is the probability of a head, which is the proportion of all tosses that give a head. The tosses we actually make are an SRS from this population.

The French naturalist Count Buffon (1707–1788) tossed a coin 4040 times. He got 2048 heads. The sample proportion of heads is

\[
\hat{p} = \frac{2048}{4040} = 0.5069
\]

That’s a bit more than one-half. Is this evidence that Buffon’s coin was not balanced?

Step 1. Hypotheses. The null hypothesis says that the coin is balanced \((p = 0.5)\). The alternative hypothesis is two-sided, because we did not suspect before seeing the data that the coin favored either heads or tails. We therefore test the hypotheses

\[
H_0: p = 0.5 \\
H_a: p \neq 0.5
\]

The null hypothesis gives \( p \) the value \( p_0 = 0.5 \).

Step 2. Test statistic. The \( z \) test statistic is

\[
z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.5069 - 0.5}{\sqrt{\frac{0.5(0.5)}{4040}}} = 0.88
\]

Step 3. \( P \)-value. Because the test is two-sided, the \( P \)-value is the area under the standard Normal curve more than 0.88 away from 0 in either direction. Figure 18.4 shows this area. In Table A we read that the area below \(-0.88\) is 0.1894. The \( P \)-value is twice this area:

\[
P = 2(0.1894) = 0.3788
\]

You can approximate the \( P \)-value by comparing \( z = 0.88 \) with the critical values in the last row of Table C. It lies between the values for tail areas 0.15 and 0.20, so the two-sided \( P \)-value lies between 0.30 and 0.40. This is enough to reach a conclusion.

Conclusion. A proportion of heads as far from one-half as Buffon’s would happen more than 30% of the time when a balanced coin is tossed 4040 times. Buffon’s result doesn’t show that his coin is unbalanced.

In Example 18.8, we failed to find good evidence against \( H_0: p = 0.5 \). We cannot conclude that \( H_0 \) is true, that is, that the coin is perfectly balanced. No doubt \( p \) is not exactly 0.5. The test of significance shows only that the results of Buffon’s 4040 tosses do not distinguish this coin from one that is perfectly...
Significance tests for a proportion

Figure 18.4 The P-value for the two-sided test of Example 18.8.

balanced. To see what values of $p$ are consistent with the sample results, use a confidence interval.

**Example 18.9** Estimating the chance of a head

With 2048 successes in 4040 trials, the large-sample and plus four intervals will be almost identical. The 95% large-sample confidence interval for the probability $p$ that Buffon’s coin gives a head is

$$
\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 0.5069 \pm 1.960 \sqrt{\frac{(0.5069)(0.4931)}{4040}}
$$

$$
= 0.5069 \pm 0.0154
$$

$$
= 0.4915 \text{ to } 0.5223
$$

We are 95% confident that the probability of a head is between 0.4915 and 0.5223.

The confidence interval is more informative than the test in Example 18.8. We would not be surprised if the true probability of a head for Buffon’s coin were something like 0.51.

**Apply Your Knowledge**

18.18 Spinning pennies. Spinning a coin, unlike tossing it, may not give heads and tails equal probabilities. I spun a penny 200 times and got 83 heads. How significant is this evidence against equal probabilities?

State hypotheses, give the test statistic, use Table C to approximate its $P$-value, and state your conclusion.

18.19 Teens and their TV sets. A random sample of 1048 13- to 17-year-olds found that 692 had a television set in their room. Is this
good evidence that more than half of all teens have a TV in their room? State hypotheses, give the test statistic, use Table A to find its $P$-value, and state your conclusion.

18.20 No test. Explain why we can’t use the $z$ test for a proportion in these situations:
(a) You toss a coin 10 times in order to test the hypothesis $H_0: p = 0.5$ that the coin is balanced.
(b) A college president says, “99% of the alumni support my firing of Coach Boggs.” You contact an SRS of 200 of the college’s 15,000 living alumni to test the hypothesis $H_0: p = 0.99$.

**Chapter 18 SUMMARY**

Tests and confidence intervals for a population proportion $p$ when the data are an SRS of size $n$ are based on the sample proportion $\hat{p}$.

When $n$ is large, $\hat{p}$ has approximately the Normal distribution with mean $p$ and standard deviation $\sqrt{p(1-p)/n}$.

The level $C$ large-sample confidence interval for $p$ is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $z^*$ is the critical value for the standard Normal curve with area $C$ between $z^*$ and $-z^*$.

The true confidence level of the large-sample interval can be substantially less than the planned level $C$ unless the sample is very large. We recommend using the plus four interval instead.

To get a more accurate confidence interval, add four imaginary observations, two successes and two failures, to your sample. Then use the same formula for the confidence interval. This is the plus four confidence interval. Use this interval in practice for confidence level 90% or higher and sample size $n$ at least 10.

The sample size needed to obtain a confidence interval with approximate margin of error $m$ for a population proportion is

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

where $\hat{p}^*$ is a guessed value for the sample proportion $\hat{p}$, and $z^*$ is the standard Normal critical point for the level of confidence you want. If you use $\hat{p}^* = 0.5$ in this formula, the margin of error of the interval will be less than or equal to $m$ no matter what the value of $\hat{p}$ is.
Significance 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 standard Normal distribution. Use this test when \( np_0 \geq 10 \) and \( n(1 - p_0) \geq 10 \).

18.21 **Information online.** A random sample of 1318 Internet users was asked where they will go for information the next time they need information about health or medicine; 606 said that they would use the Internet.\(^\text{15}\) Give a 99% confidence interval for the proportion of all Internet users who feel this way. Be sure to check that the conditions for use of your method are met.

18.22 **Seat belt use.** The proportion of drivers who use seat belts depends on things like age, gender, ethnicity, and local law. As part of a broader study, investigators observed a random sample of 117 female Hispanic drivers in Boston; 68 of these drivers were wearing seat belts.\(^\text{16}\) Give a 95% confidence interval for the proportion of all female Hispanic drivers in Boston who wear seat belts. Be sure to check that the conditions for use of your method are met.

18.23 **Attitudes toward nuclear power.** A Gallup Poll on energy use asked 512 randomly selected adults if they favored “increasing the use of nuclear power as a major source of energy.” Gallup reported that 225 said “Yes.” Does this poll give good evidence that fewer than half of all adults favor increased use of nuclear power? State hypotheses, give the test statistic, use Table C to approximate its \( P \)-value, and state your conclusion.

18.24 **Seat belt use, continued.** Do the data in Exercise 18.22 give good reason to conclude that more than half of Hispanic female drivers in Boston wear seat belts? State hypotheses, give the test statistic, use Table C to approximate its \( P \)-value, and state your conclusion.

18.25 **Student drinking.** The College Alcohol Study interviewed an SRS of 14,941 college students about their drinking habits. The sample was stratified using 140 colleges as strata, but the overall effect is close to an SRS of students. The response rate was between 60% and 70% at most colleges. This is quite good for a national sample, though nonresponse is as usual the biggest weakness of this survey. Of the
students in the sample, 10,010 supported cracking down on underage drinking.17 Give a 99% confidence interval for the proportion of all college students who feel this way.

18.26 Running red lights. A random-digit dialing telephone survey of 880 drivers asked, “Recalling the last ten traffic lights you drove through, how many of them were red when you entered the intersections?” Of the 880 respondents, 171 admitted that at least one light had been red.18
(a) Give a 95% confidence interval for the proportion of all drivers who ran one or more of the last ten red lights they met.
(b) Nonresponse is a practical problem for this survey—only 21.6% of calls that reached a live person were completed. Another practical problem is that people may not give truthful answers. What is the likely direction of the bias: do you think more or fewer than 171 of the 880 respondents really ran a red light? Why?

18.27 Detecting genetically modified soybeans. Most soybeans grown in the United States are genetically modified to, for example, resist pests and so reduce use of pesticides. Because some nations do not accept genetically modified (GM) foods, grain-handling facilities routinely test soybean shipments for the presence of GM beans. In a study of the accuracy of these tests, researchers submitted lots of soybeans containing 1% GM beans to 23 randomly selected facilities. Eighteen detected the GM beans.19
(a) Show that the conditions for the large-sample confidence interval are not met. Show that the conditions for the plus four interval are met.
(b) Use the plus four method to give a 90% confidence interval for the percent of all grain-handling facilities that will correctly detect 1% of GM beans in a shipment.

18.28 Equality for women? Have efforts to promote equality for women gone far enough in the United States? A poll on this issue by the cable network MSNBC contacted 1019 adults. A newspaper article about the poll said, “Results have a margin of sampling error of plus or minus 3 percentage points.”20
(a) Overall, 54% of the sample (550 of 1019 people) answered “Yes.” Find a 95% confidence interval for the proportion in the adult population who would say “Yes” if asked. Is the report’s claim about the margin of error roughly right? (Assume that the sample is an SRS.)
(b) The news article said that 65% of men, but only 43% of women, think that efforts to promote equality have gone far enough. Explain why we do not have enough information to give confidence intervals for men and women separately.
Chapter 18 Exercises

(c) Would a 95% confidence interval for women alone have a margin of error less than 0.03, about equal to 0.03, or greater than 0.03? Why? You see that the news article’s statement about the margin of error for poll results is a bit misleading.

18.29 The IRS plans an SRS. The Internal Revenue Service plans to examine an SRS of individual federal income tax returns from each state. One variable of interest is the proportion of returns claiming itemized deductions. The total number of tax returns in a state varies from more than 13 million in California to fewer than 220,000 in Wyoming.

(a) Will the sampling variability of the sample proportion change from state to state if an SRS of 2000 tax returns is selected in each state? Explain your answer.

(b) Will the sampling variability of the sample proportion change from state to state if an SRS of 1% of all tax returns is selected in each state? Explain your answer.

18.30 Condom usage. The National AIDS Behavioral Surveys (Example 18.1) also interviewed a sample of adults in the cities where AIDS is most common. This sample included 803 heterosexuals who reported having more than one sexual partner in the past year. We can consider this an SRS of size 803 from the population of all heterosexuals in high-risk cities who have multiple partners. These people risk infection with the AIDS virus. Yet 304 of the respondents said they never use condoms. Is this strong evidence that more than one-third of this population never use condoms?

18.31 Going to church. A Gallup Poll asked a sample of 1785 adults, “Did you, yourself, happen to attend church or synagogue in the last 7 days?” Of the respondents, 750 said “Yes.” Treat Gallup’s sample as an SRS of all American adults. Give a 99% confidence interval for the proportion of all adults who claim that they attended church or synagogue during the week preceding the poll. (The proportion who actually attended is no doubt lower—some people say “Yes” if they usually attend, often attend, or sometimes attend.)

18.32 Going to church, continued. Do the results of the poll in Exercise 18.31 provide good evidence that fewer than half of the population would claim to have attended church or synagogue?

18.33 Going to church, continued. How large a sample would be required to obtain a margin of error of 0.01 in a 99% confidence interval for the proportion who claim to have attended church or synagogue (see Exercise 18.31)? (Use the conservative guess $p^* = 0.5$, and explain why this method is reasonable in this situation.)

18.34 Small-business failures. A study of the survival of small businesses chose an SRS from the telephone directory’s Yellow Pages listings of
food-and-drink businesses in 12 counties in central Indiana. For various reasons, the study got no response from 45% of the businesses chosen. Interviews were completed with 148 businesses. Three years later, 22 of these businesses had failed.\(^{21}\)

(a) Give a 95% confidence interval for the percent of all small businesses in this class that fail within three years.

(b) Based on the results of this study, how large a sample would you need to reduce the margin of error to 0.04?

(c) The authors hope that their findings describe the population of all small businesses. What about the study makes this unlikely? What population do you think the study findings describe?

18.35 Matched pairs. One-sample procedures for proportions, like those for means, are used to analyze data from matched pairs designs. Here is an example.

Each of 50 subjects tastes two unmarked cups of coffee and says which he or she prefers. One cup in each pair contains instant coffee; the other, fresh-brewed coffee. Thirty-one of the subjects prefer the fresh-brewed coffee. Take \( p \) to be the proportion of the population who would prefer fresh-brewed coffee in a blind tasting.

(a) Test the claim that a majority of people prefer the taste of fresh-brewed coffee. State hypotheses and report the \( z \)-statistic and its \( P \)-value. Is your result significant at the 5% level? What is your practical conclusion?

(b) Find a 90% confidence interval for \( p \).

(c) When you do an experiment like this, in what order should you present the two cups of coffee to the subjects?

18.36 Customer satisfaction. An automobile manufacturer would like to know what proportion of its customers are not satisfied with the service provided by the local dealer. The customer relations department will survey a random sample of customers and compute a 99% confidence interval for the proportion who are not satisfied.

(a) Past studies suggest that this proportion will be about 0.2. Find the sample size needed if the margin of error of the confidence interval is to be about 0.015.

(b) When the sample is actually contacted, 10% of the sample say they are not satisfied. What is the margin of error of the 99% confidence interval?

18.37 Surveying students. You are planning a survey of students at a large university to determine what proportion favor an increase in student fees to support an expansion of the student newspaper. Using records provided by the registrar, you can select a random sample of students. You will ask each student in the sample whether he or she is in favor
of the proposed increase. Your budget will allow a sample of 100 students.

(a) For a sample of size 100, construct a table of the margins of error for 95% confidence intervals when \( \hat{p} \) takes the values 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9.

(b) A former editor of the student newspaper offers to provide funds for a sample of size 500. Repeat the margin of error calculations in (a) for the larger sample size. Then write a short thank-you note to the former editor describing how the larger sample size will improve the results of the survey.

18.38 Alternative medicine. A nationwide random survey of 1500 adults asked about attitudes toward “alternative medicine” such as acupuncture, massage therapy, and herbal therapy. Among the respondents, 660 said they would use alternative medicine if traditional medicine was not producing the results they wanted.

(a) Give a 95% confidence interval for the proportion of all adults who would use alternative medicine.

(b) Write a short paragraph for a news report based on the survey results.

Chapter 18 MEDIA EXERCISES

18.39 The weevils are coming. The imported longhorn weevil is a flightless insect that is a major pest of red clover. The ESEE story “Seasonal Weevil Migration” gives data from insect traps placed between a field of red clover and an adjacent woods. At one season (combining data from two years), the traps caught 55 weevils moving from the woods to the clover and 19 weevils moving from the clover to the woods. Use a confidence interval to estimate the proportion of migrating weevils in this season that move toward the clover.