EXAMPLE 22.2 Young Adults Living at Home

A surprising number of young adults (aged 19 to 25) still live at home with their parents. A random sample of 2253 men and 2629 women in this age group found that 44% of the men but only 35% of the women lived at home. Is this significant evidence that the proportions living at home differ in the populations of all young men and all young women? We want to *compare two population proportions*. This is the topic of Chapter 23. ■

To make inferences about a population mean μ, 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.

22.1 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 22.1, 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 had multiple sexual partners. The statistic that estimates the parameter p is the

sample proportion

sample proportion

$$\hat{p} = \frac{\text{number of successes in the sample}}{\text{total number of individuals in the sample}} = \frac{170}{2673} = 0.0636$$

Read the sample proportion $\hat{p}$ as "p-hat."

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.[2]

Sampling Distribution of a Sample Proportion

Draw an SRS of size n from a large population that contains proportion p of successes. Let $\hat{p}$ be the **sample proportion** of successes,

$$\hat{p} = \frac{\text{number of successes in the sample}}{n}$$

Then:

- The **mean** of the sampling distribution is p.
- The **standard deviation** of the sampling distribution is

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

- As the sample size increases, the sampling distribution of $\hat{p}$ becomes **approximately Normal**. That is, for large n, $\hat{p}$ has approximately the $N(p, \sqrt{p(1-p)/n})$ distribution.

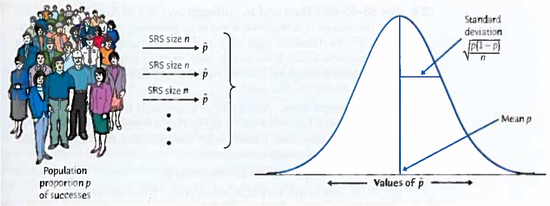


FIGURE 22.1
Select a large SRS from a population in which the proportion p are successes. The sampling distribution of the proportion $\hat{p}$ of successes in the sample is approximately Normal. The mean is p and the standard deviation is $\sqrt{p(1-p)/n}$.

Figure 22.1 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 $\bar{x}$, except that the distribution of $\hat{p}$ is only approximately Normal. The mean of the sampling distribution of $\hat{p}$ 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. As is the case for $\bar{x}$, 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.

EXAMPLE 22.3 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. In many such samples, the proportion $\hat{p}$ of the 2673 people in the sample who had more than one partner would vary according to (approximately) the Normal distribution with mean 0.06 and standard deviation

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.06)(0.94)}{2673}} = \sqrt{0.0000211} = 0.00459$$ ■

Apply Your Knowledge

22.1 Staph Infections. A study investigated ways to prevent staph infections in surgery patients. In a first step, the researchers examined the nasal secretions of a random sample of 6771 patients admitted to various hospitals for surgery. They found that 1251 of these patients tested positive for *Staphylococcus aureus*, a bacterium responsible for most staph infections.[3]

(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.

22.2 The 68–95–99.7 Rule and $\hat{p}$. Although over 50% of American adults believe the maxim that breakfast is the most important meal of the day, only about 30% eat breakfast daily.[4] A cereal manufacturer contacts an SRS of 1000 American adults. If the sample were repeated many times, what would be the range of sample proportions who eat breakfast daily according to the 95 part of the 68–95–99.7 rule?

22.3 Social Network Sites. About 90% of young adult Internet users (aged 18 to 29) use social network sites.[5] Suppose that a sample survey contacts an SRS of 1500 young adult Internet users and calculates the proportion $\hat{p}$ in this sample who use social network sites.

(a) What is the approximate distribution of $\hat{p}$?

(b) If the sample size were 6000 rather than 1500, what would be the approximate distribution of $\hat{p}$?

22.2 Large-sample confidence intervals for a proportion

We can follow the same path from sampling distribution to confidence interval as we did for $\bar{x}$ in Chapter 15. To obtain a level C confidence interval for p, we start by capturing the central probability C in the distribution of $\hat{p}$. To do this, go out z^* standard deviations from the mean p, where z^* is the critical value that captures the central area C under the standard Normal curve. Figure 22.2 shows the result. The confidence interval is

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

standard error of $\hat{p}$

This won't do, because we don't know the value of p. So we replace the standard deviation by the **standard error of $\hat{p}$**

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

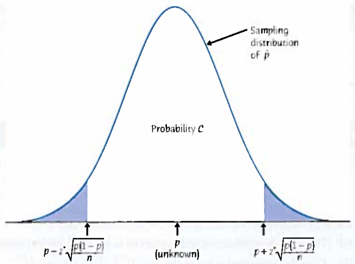


FIGURE 22.2
With probability C, $\hat{p}$ lies within $\pm z^*\sqrt{p(1-p)/n}$ of the unknown population proportion p. That is, in these samples p lies within $\pm z^*\sqrt{p(1-p)/n}$ of $\hat{p}$.

to get the confidence interval

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

As with previous confidence intervals, this interval has the familiar form

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

We can trust this confidence interval only for large samples. Because the number of successes must be a whole number, using a continuous Normal distribution to describe the behavior of $\hat{p}$ may not be accurate unless n is large. Because the approximation is least accurate for populations that are almost all successes or almost all failures, we require that the sample have both enough successes and enough failures rather than that the overall sample size be large. *Pay attention to both conditions for inference in the box below that summarizes the confidence interval: we must as usual be willing to regard the sample as an SRS from the population, and the sample must have both enough successes and enough failures. The condition on successes and failures ensures that the sample size is large enough to use the Normal approximation without knowing p.*

Large-Sample Confidence Interval for a Population Proportion

Draw an SRS of size n from a large population that contains an 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 numbers of successes and failures in the sample are both at least 15.[6]

Why not t? Notice that we *don't* change z^* to t^* when we replace the standard deviation by the standard error. When the sample mean $\bar{x}$ estimates the population mean μ, a separate parameter σ describes the variability of the distribution of $\bar{x}$. We separately estimate σ, and this leads to a t distribution. When the sample proportion $\hat{p}$ estimates the population proportion p, the variability depends on p, not on a separate parameter. There is no t distribution—we just make the Normal approximation a bit less accurate when we replace p in the standard deviation by $\hat{p}$.

EXAMPLE 22.4 Estimating Risky Behavior

The four-step process for any confidence interval is outlined on page 381.

STATE: The National AIDS Behavioral Surveys found that 170 of a sample of 2673 adult heterosexuals had had multiple partners. That is,

$$\hat{p} = \frac{170}{2673} = 0.0636$$

What can we say about the population of all adult heterosexuals?

PLAN: We will give a 99% confidence interval to estimate the proportion p of all adult heterosexuals who have multiple partners.

SOLVE: First verify the conditions 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 sample is large enough: the numbers of successes (170) and failures (2503) in the sample are both much larger than 15.

The sample size condition is easily satisfied. The condition that the sample be an SRS is only approximately met.

A 99% confidence interval for the proportion p of all adult heterosexuals with multiple partners uses the standard Normal critical value $z^* = 2.576$. The confidence interval is

$$\hat{p} \pm z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.0636 \pm 2.576\sqrt{\frac{(0.0636)(0.9364)}{2673}}$$

$$= 0.0636 \pm 0.0122$$

$$= 0.0514 \text{ to } 0.0758$$

CONCLUDE: We are 99% confident that the percent of adult heterosexuals who have had more than one sexual partner in the year prior to the survey lies between about 5.1% and 7.6%. ■

CAUTION *As usual, the practical problems of a large sample survey weaken our confidence in the AIDS survey's conclusions.* Only people in households with landline telephones could be reached. Although at the time of the survey about 89% of American households had landline telephones, as the number of cell phone only users increases, using a sample of households with landline phones is becoming less acceptable for surveys of the general population (see page 216). Additionally, some groups at high risk for AIDS, such as people who inject illegal drugs, often don't live in settled households and were therefore 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 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.*[7]

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

Statistics in Your World: WHO IS A SMOKER?

When estimating a proportion p, be sure you know what counts as a "success." The news says that 20% of adolescents smoke. Shocking. It turns out that this is the percent who smoked at least once in the past month. If we say that a smoker is someone who smoked on at least 20 of the past 30 days and smoked at least half a pack on those days, fewer than 4% of adolescents qualify.

LaunchPad Online Resources

- **The Snapshots video, *Inference for One Proportion*, provides the details of constructing a large-sample confidence interval for a proportion through an interesting example involving an opinion poll.**
- **The StatClips Examples video, *Confidence Intervals: Intervals for Proportions Example C*, illustrates the computation of a large-sample confidence interval for a proportion.**

Apply Your Knowledge

22.4 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.

22.5 Canadian Attitudes toward Guns. Canada has much stronger gun control laws than the United States, and Canadians support gun control more strongly than do Americans. A sample survey asked a random sample of 1505 adult Canadians, "Do you agree or disagree that all firearms should be registered?" Of the 1505 people in the sample, 1288 answered either "Agree strongly" or "Agree somewhat."[8]

(a) The survey dialed residential telephone numbers at random in all 10 Canadian provinces (omitting the sparsely populated northern territories). Based on what you know about sample surveys, what is likely to be the biggest weakness in this survey?

(b) Nonetheless, act as if we have an SRS from adults in the Canadian provinces. Give a 95% confidence interval for the proportion who support registration of all firearms.

22.6 Weightlifting Injuries. Resistance training is a popular form of conditioning aimed at enhancing sports performance and is widely used among high school, college, and professional athletes, although its use for younger athletes is controversial. A random sample of 4111 patients aged 8–30 admitted to U.S. emergency rooms with the Consumer Product Safety Commission code "weightlifting" were obtained. These injuries were classified as "accidental" if caused by dropped weight or improper equipment use. Of the 4111 weightlifting injuries, 1552 were classified as accidental.[9] Give a 90% confidence interval for the proportion of weightlifting injuries in this age group that were accidental. Follow the four-step process as illustrated in Example 22.4.

22.3 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 425) 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

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

NEW YORK, NEW YORK

New York City, they say, is bigger, richer, faster, and ruder. Maybe there's something to that. The sample survey firm Zogby International says that as a national average it takes five telephone calls to reach a live person. When calling to New York, it takes 12 calls. Survey firms assign their best interviewers to make calls to New York and often pay them bonuses to cope with the stress.

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 p^* based on a pilot study or on past experience with similar studies. You can do several calculations to cover the range of values of $\hat{p}$ you might get.
2. Use $p^* = 0.5$ as the guess. The margin of error m is largest when $p^* = 0.5$, 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 discussed in Section 22.5.

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 always 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. You can use the conservative guess $p^* = 0.5$ if you expect the true $\hat{p}$ to be roughly between 0.3 and 0.7. 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. 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 22.5 Planning a Poll

Colin Anderson/Brand X/Age fotostock

STATE: 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 intend to vote for Chavez. You will contact an SRS of registered voters in the city. You want to estimate the proportion p of Chavez voters with 95% confidence and a margin of error no greater than 3%, or 0.03. How large a sample do you need?

PLAN: Find the sample size n needed for margin of error $m = 0.03$ and 95% confidence. The winner's share in all but the most lopsided elections is between 30% and 70% of the vote. You can use the guess $p^* = 0.5$.

SOLVE: The sample size you need is

$$n = \left(\frac{1.96}{0.03}\right)^2 (0.5)(1 - 0.5) = 1067.1$$

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

CONCLUDE: An SRS of 1068 registered voters is adequate for margin of error ±3%. ■

If you want a 2.5% margin of error rather than 3%, then (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.

LaunchPad Online Resources

- The StatBoards video, *Computing a Sample Size for One Proportion*, provides an additional example of choosing the sample size to provide a given margin of error.

Apply Your Knowledge

22.7 Did You Use a Mobile Device? The Monterey Bay Aquarium, founded in 1984, is situated on the beautiful coast of Monterey Bay in the historic Cannery Row district. From 2009 to 2013, the aquarium conducted a survey of a random sample of visitors as they exited the museum. The survey includes visitor demographic information, use of social media, and opinions on their aquarium visit. In 2013, the survey included 165 visitors over 65, of which 42 used a mobile device such as an Android phone or iPad during their visit.[10]

(a) What is the margin of error of the large-sample 95% confidence interval for the proportion of visitors over 65 who used a mobile device during their visit?

(b) How large a sample is needed to get the common ±3 percentage point margin of error? Use the sample from part (a) as a pilot study to get p^*.

22.8 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.

(a) Starting with the 75% estimate for Italians, how large a sample must you collect in order to estimate the proportion of PTC tasters within ±0.04 with 90% confidence?

(b) Estimate the sample size required if you made no assumptions about the value of the proportion who could taste PTC. How much has the required sample size changed?

22.4 Significance tests for a proportion

The test statistic for the null hypothesis H_0: $p = p_0$ is the sample proportion $\hat{p}$ standardized using the value p_0 specified by H_0,

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

This z statistic has approximately the standard Normal distribution when H_0 is true. P-values therefore come from the standard Normal distribution. Unlike the confidence interval in which p is unknown and must be estimated by $\hat{p}$ when standardizing the estimate, in the test we can replace p by p_0 when standardizing as p_0 is specified by H_0. Additionally, because H_0 fixes a value of p when standardizing the estimate, the sample size conditions for use of the test are less stringent than for the large-sample confidence interval in which p must be estimated. Here is the procedure for tests.

Significance Tests for a Proportion

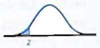
Draw an SRS of size n from a large population that contains an unknown proportion p of successes. To **test the hypothesis H_0: $p = p_0$**, compute the z statistic

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

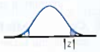
In terms of a variable Z having the standard Normal distribution, the approximate P-value for a test of H_0 against

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.[11]

EXAMPLE 22.6 Are Boys More Likely?

© Blaine Harrington III/Corbis

The four-step process for any significance test is outlined on page 401.

STATE: We hear that newborn babies are more likely to be boys than girls, presumably to compensate for higher mortality among boys in early life. Is this true? A random sample found 13,173 boys among 25,468 firstborn children.[12] The sample proportion of boys was

$$\hat{p} = \frac{13{,}173}{25{,}468} = 0.5172$$

Boys do make up more than half of the sample, but of course we don't expect a perfect 50–50 split in a random sample. Is this sample evidence that boys are more common than girls in the entire population?

PLAN: Take p to be the proportion of boys among all firstborn children of American mothers. (Biology says that this should be the same as the proportion among all children, but the survey data concern first births.)

We want to test the hypotheses

$$H_0: p = 0.5$$
$$H_a: p > 0.5$$

SOLVE: The conditions for inference require that we have a random sample and that $np_0 = (25{,}468)(0.5) = 12{,}734$ and $n(1 - p_0) = (25{,}468)(0.5) = 12{,}734$ are both greater than 10. Since the conditions for inference are met, we can go on to find the z test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \frac{0.5172 - 0.5}{\sqrt{\dfrac{(0.5)(0.5)}{25{,}468}}} = 5.49$$

The P-value is the area under the standard Normal curve to the right of $z = 5.49$. We know that this is very small; Table C shows that $P < 0.0005$. Minitab (Figure 22.3) says that P is 0 to three decimal places.

CONCLUDE: There is very strong evidence that more than half of firstborns are boys ($P < 0.001$). ■

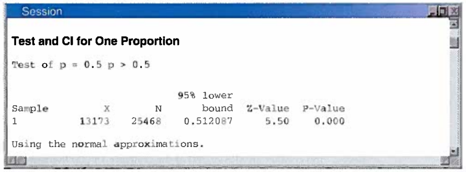


FIGURE 22.3 Minitab output for the significance test of Example 22.6. Roundoff error in Example 22.6 explains the small difference (5.49 versus 5.50) in the values of the z statistic.

EXAMPLE 22.7 Estimating the Chance of a Boy

With 13,173 successes in 25,468 trials, we have at least 15 successes and 15 failures in the sample. The conditions for the large-sample confidence interval are easily met. The 99% confidence interval is

$$\hat{p} \pm z^*\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 0.5172 \pm 2.576\sqrt{\frac{(0.5172)(0.4828)}{25{,}468}}$$
$$= 0.5172 \pm 0.0081$$
$$= 0.5091 \text{ to } 0.5253$$

We are 99% confident that between about 51% and 52.5% of first children are boys.

The confidence interval is more informative than the test in Example 22.6, which tells us only that more than half are boys. ■

LaunchPad Online Resources

- The **StatBoards video**, *Hypothesis Test for One Proportion*, illustrates the computation of a large-sample test for a proportion through an example.

Apply Your Knowledge

22.9 Spinning Euros. All euros have a national image on the "heads" side and a common design on the "tails" side. Spinning a coin, unlike tossing it, may not give heads and tails equal probabilities. Polish students spun the Belgian euro 250 times, with its portly king, Albert, displayed on the heads side. The result was 140 heads.[13] How significant is this evidence against equal probabilities? Follow the four-step process as illustrated in Example 22.6.

22.10 Vote for the Best Face? We often judge other people by their faces. It appears that some people judge candidates for elected office by their faces. Psychologists showed head-and-shoulders photos of the two main candidates in 32 races for the U.S. Senate to many subjects (dropping subjects who recognized one of the candidates) to see which candidate was rated "more competent" based on nothing but the photos. On election day, the candidates whose faces looked more competent won 22 of the 32 contests.[14] If faces don't influence voting, half of all races in the long run should be won by the candidate with the better face. Is there evidence that the candidate with the better face wins more than half the time? Follow the four-step process as illustrated in Example 22.6.

22.11 No Test. Explain whether we can 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 local candidate contacts an SRS of 900 of the registered voters in his district to see if there is evidence that more than half support the bill he is sponsoring.

(c) 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: $p = 0.99$.

22.5 Plus four confidence intervals for a proportion*

The large-sample confidence interval $\hat{p} \pm z^*\sqrt{\hat{p}(1-\hat{p})/n}$ for a sample proportion p is easy to calculate. It is also easy to understand because it rests directly on the approximately Normal distribution of $\hat{p}$. Unfortunately, confidence levels from this interval can be inaccurate, particularly with smaller sample sizes. 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 the sample size n and the true population proportion p.

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

*This section is optional.

because all you need to do is *add four imaginary observations: two successes and two failures.* With the added observations, the **plus four estimate** of p is

plus four estimate

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

The formula for the confidence interval is exactly as before, with the new sample size and number of successes.[15] You do not need software that offers the plus four interval—just enter the new sample size (actual size + 4) and number of successes (actual number + 2) into the large-sample procedure.

Plus Four Confidence Interval for a Proportion

Draw an SRS of size n from a large population that contains an unknown proportion p of successes. To get the **plus four confidence interval for *p***, add four imaginary observations—two successes and two failures. Then use the large-sample confidence interval with the new sample size ($n + 4$) and number of successes (actual number + 2).

Use this interval when the confidence level is at least 90% and the sample size n is at least 10, with any counts of successes and failures.

EXAMPLE 22.8 Cocaine Traces in Spanish Currency

STATE: Cocaine users commonly snort the powder up the nose through a rolled-up paper currency bill. Spain has a high rate of cocaine use, so it's not surprising that euro paper currency in Spain often bears traces of cocaine. Researchers collected 20 euro bills in each of several Spanish cities. In Madrid, 17 out of 20 bore traces of cocaine.[16] The researchers note that we can't tell whether the bills had been used to snort cocaine or had been contaminated in currency-sorting machines. Estimate the proportion of all euro bills in Madrid that have traces of cocaine.


© Andrea Matone/Alamy

PLAN: Take p to be the proportion of bills that show cocaine traces. That is, a "success" is a bill that shows cocaine traces. Give a 95% confidence interval for p.

SOLVE: It is not clear how the bills in the sample were selected, so we don't know if we have an SRS. We will act as though we have an SRS, but we proceed with caution. The conditions for use of the large-sample interval are not met because there are only three failures. To apply the plus four method, add two successes and two failures to the original data. The plus four estimate of p is

$$\tilde{p} = \frac{17 + 2}{20 + 4} = \frac{19}{24} = 0.7917$$

We calculate the plus four confidence interval in the same way as we do the large-sample interval, but we base it on 19 successes in 24 observations. Here it is:

$$\begin{aligned}\tilde{p} \pm z^*\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} &= 0.7917 \pm 1.960\sqrt{\frac{(0.7917)(0.2083)}{24}}\\ &= 0.7917 \pm 0.1625\\ &= 0.6292 \text{ to } 0.9542\end{aligned}$$

CONCLUDE: Assuming the sample can be regarded as an SRS, we estimate with 95% confidence that between about 63% and 95% of all euro bills in Madrid bear traces of cocaine. ■

For comparison, the ordinary sample proportion is

$$\hat{p} = \frac{17}{20} = 0.85$$

The plus four estimate $\tilde{p} = 0.7917$ in Example 22.8 is farther away from 1 than $\hat{p} = 0.85$. The plus four estimate gains its added accuracy by always moving toward 0.5 and away from 1 or 0, whichever is closer. This is particularly helpful when the sample contains only a few successes or a few failures. The numerical difference between a large-sample interval and the corresponding plus four interval is often small. Remember that the confidence level is the probability that the interval will catch the true population proportion *in very many uses*. Small differences every time can add up to more accurate confidence levels from plus four versus the large-sample interval.

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 samples of size n or larger. If $p = 0.1$, for example, the answer is $n = 646$ for the large-sample interval and $n = 11$ for the plus four interval.[17] The consensus of computational and theoretical studies is that plus four is better than the large-sample interval for many combinations of n and p.

Apply Your Knowledge


Jennifer Shields/Getty Images

22.12 Black Raspberries and Cancer. Sample surveys usually contact large samples, so we can use the large-sample confidence interval if the sample design is close to an SRS. Scientific studies often use smaller samples that require the plus four method. For example, Familial Adenomatous Polyposis (FAP) is a rare inherited disease characterized by the development of an extreme number of polyps early in life and colon cancer in virtually 100% of patients before the age of 40. A group of 14 people suffering from FAP being treated at the Cleveland Clinic drank black raspberry powder in a slurry of water every day for nine months. The numbers of polyps were reduced in 11 out of 14 of these patients.[18]

(a) Why can't we use the large-sample confidence interval for the proportion p of patients suffering from FAP that will have the number of polyps reduced after nine months of treatment?

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

(c) Give the plus four 90% confidence interval for the proportion of patients suffering from FAP who will have the number of polyps reduced after nine months of treatment.

22.13 Computer/Internet-Based Crime. With over 50% of adults spending more than an hour a day on the Internet, the number experiencing computer or Internet-based crime continues to rise. A survey in 2010 of a random sample of 1025 adults, aged 18 and older, reached by random digit dialing found 113 adults in the sample who said that they or a household member was a victim of a computer or Internet crime on their home computer in the past year.[19]

(a) Give the 95% large-sample confidence interval for the proportion p of all households that have experienced computer or Internet crime during the year before the survey was conducted. Be sure to verify that the sample size is large enough to use the large-sample confidence interval.

(b) Give the plus four 95% confidence interval for p. If you express the two intervals in percents, rounded to the nearest tenth of a percent, how do they differ? (The plus four interval always pulls the results away from 0% or 100%, whichever is closer.)

22.14 Cocaine Traces in Spanish Currency, Continued. The plus four method is particularly useful when there are *no* successes or *no* failures in the data. The study of Spanish currency described in Example 22.8 found that in Seville, all 20 of a sample of 20 euro bills had cocaine traces.

(a) What is the sample proportion $\hat{p}$ of contaminated bills? What is the large-sample 95% confidence interval for p? It's not plausible that *every* bill in Seville has cocaine traces, as this interval says.

(b) Find the plus four estimate $\tilde{p}$ and the plus four 95% confidence interval for p. These results are more reasonable in this situation.

CHAPTER 22 SUMMARY

Chapter Specifics

- 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^*. Use this interval only when *both* the number of successes and the number of failures in the sample are at least 15.

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

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

where 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 $p^* = 0.5$ 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 for 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 in practice when $np_0 \geq 10$ and $n(1-p_0) \geq 10$.

- **(Optional topic)** To get a more accurate confidence interval for smaller sample sizes, 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.