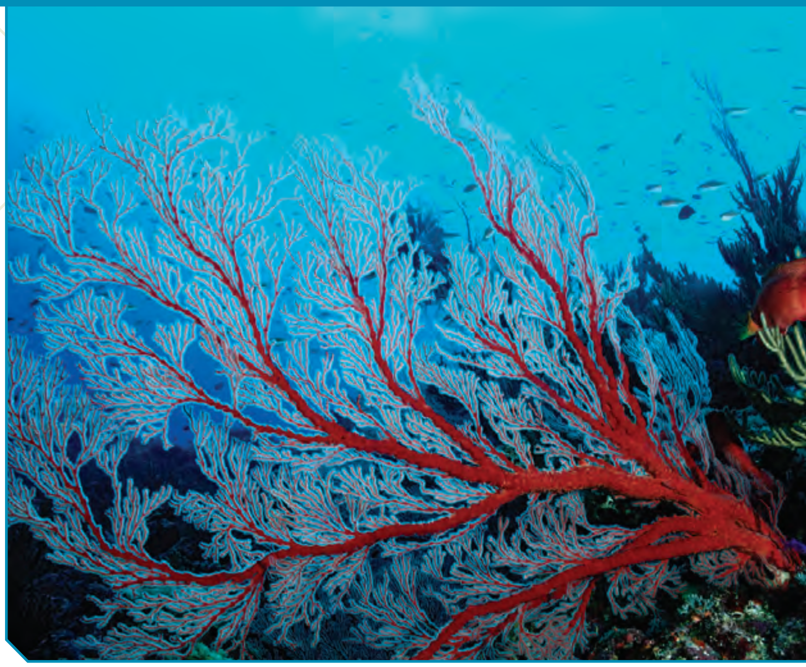
CHAPTER 19

Confidence Intervals for Proportions



WHO	Sea fans
WHAT	Percent infected
WHEN	June 2000
WHERE	Las Redes Reef, Akumal, Mexico, 40 feet deep
WHY	Research

Coral reef communities are home to one quarter of all marine plants and animals worldwide. These reefs support large fisheries by providing breeding grounds and safe havens for young fish of many species. Coral reefs are seawalls that protect shorelines against tides, storm surges, and hurricanes, and are sand "factories" that produce the limestone and sand of which beaches are made. Beyond the beach, these reefs are major tourist attractions for snorkelers and divers, driving a tourist industry worth tens of billions of dollars.

But marine scientists say that 10% of the world's reef systems have been destroyed in recent times. At current rates of loss, 70% of the reefs could be gone in 40 years. Pollution, global warming, outright destruction of reefs, and increasing acidification of the oceans are all likely factors in this loss.

Dr. Drew Harvell's lab studies corals and the diseases that affect them. They sampled sea fans[1] at 19 randomly selected reefs along the Yucatan peninsula and diagnosed whether the animals were affected by the disease *aspergillosis*.[2] In specimens collected at a depth of 40 feet at the Las Redes Reef in Akumal, Mexico, these scientists found that 54 of 104 sea fans sampled were infected with that disease.

Of course, we care about much more than these particular 104 sea fans. We care about the health of coral reef communities throughout the Caribbean. What can this study tell us about the prevalence of the disease among sea fans?

We have a sample proportion, which we write as $\hat{p}$, of 54/104, or 51.9%. Our first guess might be that this observed proportion is close to the population proportion, p. But we also know that because of natural sampling variability, if the researchers had drawn a second sample of 104 sea fans at roughly the same time, the proportion infected from that sample probably wouldn't have been exactly 51.9%.

[1] That's a sea fan in the picture. Although they look like trees, they are actually colonies of genetically identical animals.

[2] K. M. Mullen, C. D. Harvell, A. P. Alker, D. Dube, E. Jordán-Dahlgren, J. R. Ward, and L. E. Petes, "Host range and resistance to aspergillosis in three sea fan species from the Yucatan," *Marine Biology* (2006), Springer-Verlag.

What *can* we say about the population proportion, p? To start to answer this question, think about how different the sample proportion might have been if we'd taken another random sample from the same population. But wait. Remember—we aren't actually going to take more samples. We just want to *imagine* how the sample proportions might vary from sample to sample. In other words, we want to know about the *sampling distribution* of the sample proportion of infected sea fans.

A Confidence Interval

A S ***Activity:* Confidence Intervals and Sampling Distributions.** Simulate the sampling distribution, and see how it gives a confidence interval.

Let's look at our model for the sampling distribution. What do we know about it? We know it's approximately Normal (under certain assumptions, which we should be careful to check) and that its mean is the proportion of all infected sea fans on the Las Redes Reef. Is the infected proportion of *all* sea fans 51.9%? No, that's just $\hat{p}$, our estimate. We don't know the proportion, p, of all the infected sea fans; that's what we're trying to find out. We do know, though, that the sampling distribution model of $\hat{p}$ is centered at p, and we know that the standard deviation of the sampling distribution is $\sqrt{\frac{pq}{n}}$.

NOTATION ALERT:

Remember that $\hat{p}$ is our sample-based estimate of the true proportion p. Recall also that q is just shorthand for $1 - p$, and $\hat{q} = 1 - \hat{p}$.

When we use $\hat{p}$ to estimate the standard deviation of the sampling distribution model, we call that the **standard error** and write $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$.

Now we have a problem: Since we don't know p, we can't find the true standard deviation of the sampling distribution model. We do know the observed proportion, $\hat{p}$, so, of course we just use what we know, and we estimate. That may not seem like a big deal, but it gets a special name. Whenever we estimate the standard deviation of a sampling distribution, we call it a **standard error.**[3] For a sample proportion, $\hat{p}$, the standard error is

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

For the sea fans, then:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.519)(0.481)}{104}} = 0.049 = 4.9\%.$$

Now we know that the sampling model for $\hat{p}$ should look like this:

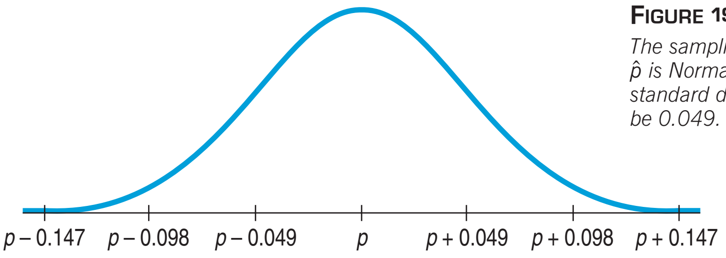


FIGURE 19.1

The sampling distribution model for $\hat{p}$ is Normal with a mean of p and a standard deviation we estimate to be 0.049.

Great. What does that tell us? Well, because it's Normal, it says that about 68% of all samples of 104 sea fans will have $\hat{p}$'s within 1 *SE*, 0.049, of p. And about 95% of all these samples will be within $p \pm 2$ *SE*s. But where is *our* sample proportion in this picture? And what value does p have? We still don't know!

We do know that for 95% of random samples, $\hat{p}$ will be no more than 2 *SE*s away from p. So let's look at this from $\hat{p}$'s point of view. If I'm $\hat{p}$, there's a 95%

[3] This isn't such a great name because it isn't standard and nobody made an error. But it's much shorter and more convenient than saying, "the estimated standard deviation of the sampling distribution of the sample statistic."

chance that p is no more than 2 *SE*s away from me. If I reach out 2 *SE*s, or 2×0.049, away from me on both sides, I'm 95% sure that p will be within my grasp. Now I've got him! Probably. Of course, even if my interval does catch p, I still don't know its true value. The best I can do is an interval, and even then I can't be positive it contains p.

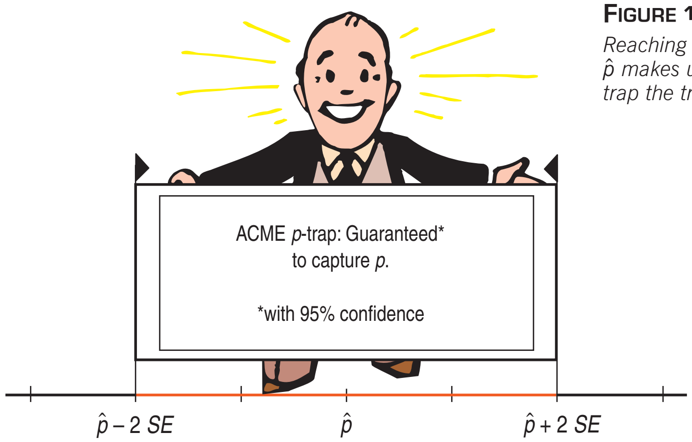


Figure 19.2

Reaching out 2 SEs on either side of $\hat{p}$ makes us 95% confident that we'll trap the true proportion, p.

So what can we really say about p? Here's a list of things we'd like to be able to say, in order of strongest to weakest and the reasons we can't say most of them:

A S ***Activity:* Can We Estimate a Parameter?** Consider these four interpretations of a confidence interval by simulating to see whether they could be right.

1. **"51.9% of *all* sea fans on the Las Redes Reef are infected."** It would be nice to be able to make absolute statements about population values with certainty, but we just don't have enough information to do that. There's no way to be sure that the population proportion is the same as the sample proportion; in fact, it almost certainly isn't. Observations vary. Another sample would yield a different sample proportion.
2. **"It is *probably* true that 51.9% of all sea fans on the Las Redes Reef are infected."** No. In fact, we can be pretty sure that whatever the true proportion is, it's not exactly 51.900%. So the statement is not true.
3. **"We don't know exactly what proportion of sea fans on the Las Redes Reef is infected, but we *know* that it's within the interval 51.9% ± 2 × 4.9%. That is, it's between 42.1% and 61.7%."** This is getting closer, but we still can't be certain. We can't know *for sure* that the true proportion is in this interval—or in any particular interval.

"Far better an approximate answer to the right question, . . . than an exact answer to the wrong question."

—John W. Tukey

4. **"We don't know exactly what proportion of sea fans on the Las Redes Reef is infected, but the interval from 42.1% to 61.7% *probably* contains the true proportion."** We've now fudged twice—first by giving an interval and second by admitting that we only think the interval "probably" contains the true value. And this statement is true.

That last statement may be true, but it's a bit wishy-washy. We can tighten it up a bit by quantifying what we mean by "probably." We saw that 95% of the time when we reach out 2 *SE*s from $\hat{p}$ we capture p, *so we can be 95% confident that this is one of those times*. After putting a number on the probability that this interval covers the true proportion, we've given our best guess of where the parameter is and how certain we are that it's within some range.

5. **"We are 95% confident that between 42.1% and 61.7% of Las Redes sea fans are infected."** Statements like these are called **confidence intervals.** They're the best we can do.

Each confidence interval discussed in the book has a name. You'll see many different kinds of confidence intervals in the following chapters. Some will be about more than *one* sample, some will be about statistics other than *proportions*, and some will use models other than the Normal. The interval calculated and interpreted here is sometimes called a **one-proportion *z*-interval.**[4]

JUST CHECKING

A Pew Research study regarding cell phones asked questions about cell phone experience. One growing concern is unsolicited advertising in the form of text messages. Pew asked cell phone owners, "Have you ever received unsolicited text messages on your cell phone from advertisers?" and 17% reported that they had. Pew estimates a 95% confidence interval to be 0.17 ± 0.04, or between 13% and 21%.

Are the following statements about people who have cell phones correct? Explain.

1. In Pew's sample, somewhere between 13% and 21% of respondents reported that they had received unsolicited advertising text messages.
2. We can be 95% confident that 17% of U.S. cell phone owners have received unsolicited advertising text messages.
3. We are 95% confident that between 13% and 21% of all U.S. cell phone owners have received unsolicited advertising text messages.
4. We know that between 13% and 21% of all U.S. cell phone owners have received unsolicited advertising text messages.
5. 95% of all U.S. cell phone owners have received unsolicited advertising text messages.

What Does "95% Confidence" Really Mean?

A S ***Activity:* Confidence Intervals for Proportions.** This new interactive tool makes it easy to construct and experiment with confidence intervals. We'll use this tool for the rest of the course—sure beats calculating by hand!

What do we mean when we say we have 95% confidence that our interval contains the true proportion? Formally, what we mean is that "95% of samples of this size will produce confidence intervals that capture the true proportion." This is correct, but a little long winded, so we sometimes say, "we are 95% confident that the true proportion lies in our interval." Our uncertainty is about whether the particular sample we have at hand is one of the successful ones or one of the 5% that fail to produce an interval that captures the true value.

Back in Chapter 18 we saw that proportions vary from sample to sample. If other researchers select their own samples of sea fans, they'll also find some infected by the disease, but each person's sample proportion will almost certainly differ from ours. When they each try to estimate the true rate of infection in the entire population, they'll center *their* confidence intervals at the proportions they observed in their own samples. Each of us will end up with a different interval.

Our interval guessed the true proportion of infected sea fans to be between about 42% and 62%. Another researcher whose sample contained more infected fans than ours did might guess between 46% and 66%. Still another who happened to collect fewer infected fans might estimate the true proportion to be between 23% and 43%. And so on. Every possible sample would produce yet another confidence interval. Although wide intervals like these can't pin down the actual rate of infection very precisely, we expect that most of them should be winners, capturing the true value. Nonetheless, some will be duds, missing the population proportion entirely.

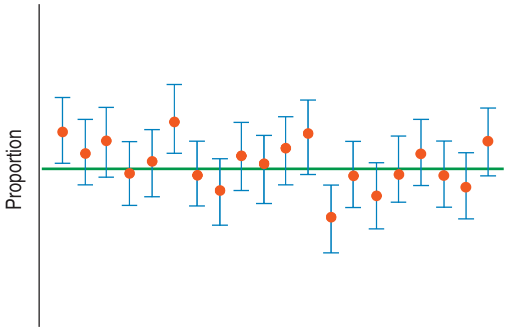
On the next page you'll see confidence intervals produced by simulating 20 different random samples. The red dots are the proportions of infected fans in

[4] In fact, this confidence interval is so standard for a single proportion that you may see it simply called a "confidence interval for the proportion."

each sample, and the blue segments show the confidence intervals found for each. The green line represents the true rate of infection in the population, so you can see that most of the intervals caught it—but a few missed. (And notice again that it is the *intervals* that vary from sample to sample; the green line doesn't move.)

TI-*nspire*

Confidence intervals. Generate confidence intervals from many samples to see how often they capture the true proportion.



The horizontal green line shows the true percentage of all sea fans that are infected. Most of the 20 simulated samples produced confidence intervals that captured the true value, but a few missed.

Of course, there's a huge number of possible samples that *could* be drawn, each with its own sample proportion. These are just some of them. Each sample proportion can be used to make a confidence interval. That's a large pile of possible confidence intervals, and ours is just one of those in the pile. Did *our* confidence interval "work"? We can never be sure, because we'll never know the true proportion of all the sea fans that are infected. However, the Central Limit Theorem assures us that 95% of the intervals in the pile are winners, covering the true value, and only 5% are duds. *That's* why we're 95% confident that our interval is a winner!

FOR EXAMPLE Polls and margin of error

On January 30–31, 2007, Fox News/Opinion Dynamics polled 900 registered voters nationwide.[5] When asked, "Do you believe global warming exists?" 82% said "Yes". Fox reported their margin of error to be $\pm 3\%$.

Question: It is standard among pollsters to use a 95% confidence level unless otherwise stated. Given that, what does Fox News mean by claiming a margin of error of $\pm 3\%$ in this context?

If this polling were done repeatedly, 95% of all random samples would yield estimates that come within $\pm 3\%$ of the true proportion of all registered voters who believe that global warming exists.

Margin of Error: Certainty vs. Precision

We've just claimed that with a certain confidence we've captured the true proportion of all infected sea fans. Our confidence interval had the form

$$\hat{p} \pm 2\, SE(\hat{p}).$$

The extent of the interval on either side of $\hat{p}$ is called the **margin of error** (*ME*). We'll want to use the same approach for many other situations besides estimating proportions. In general, confidence intervals look like this:

$$Estimate \pm ME.$$

[5] www.foxnews.com, "Fox News Poll: Most Americans Believe in Global Warming," Feb 7, 2007.

The margin of error for our 95% confidence interval was 2 *SE*. What if we wanted to be more confident? To be more confident, we'll need to capture p more often, and to do that we'll need to make the interval wider. For example, if we want to be 99.7% confident, the margin of error will have to be 3 *SE*.

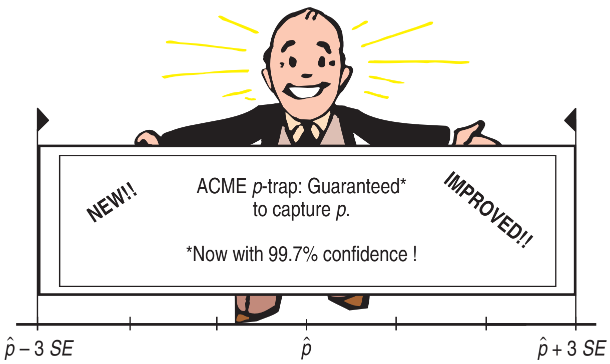
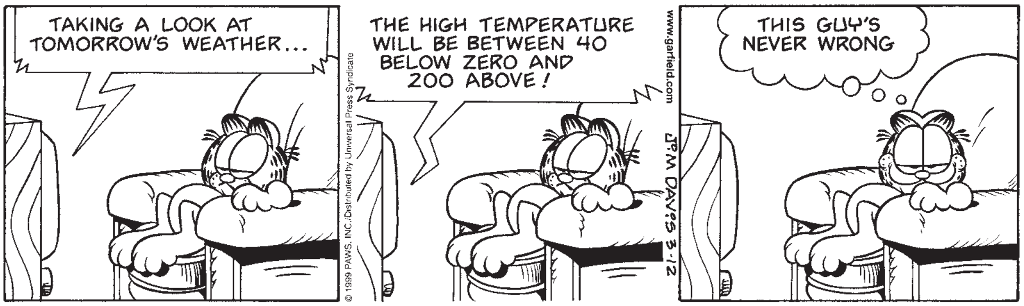


FIGURE 19.3
Reaching out 3 SEs on either side of $\hat{p}$ makes us 99.7% confident we'll trap the true proportion p. Compare with Figure 19.2.

A S ***Activity:*** **Balancing Precision and Certainty.** What percent of parents expect their kids to pay for college with a student loan? Investigate the balance between the precision and the certainty of a confidence interval.

The more confident we want to be, the larger the margin of error must be. We can be 100% confident that the proportion of infected sea fans is between 0% and 100%, but this isn't likely to be very useful. On the other hand, we could give a confidence interval from 51.8% to 52.0%, but we can't be very confident about a precise statement like this. Every confidence interval is a balance between certainty and precision.

The tension between certainty and precision is always there. Fortunately, in most cases we can be both sufficiently certain and sufficiently precise to make useful statements. There is no simple answer to the conflict. You must choose a confidence level yourself. The data can't do it for you. The choice of confidence level is somewhat arbitrary. The most commonly chosen confidence levels are 90%, 95%, and 99%, but any percentage can be used. (In practice, though, using something like 92.9% or 97.2% is likely to make people think you're up to something.)



FOR EXAMPLE **Finding the margin of error (Take 1)**

Recap: A January 2007 Fox poll of 900 registered voters reported a margin of error of ±3%. It is a convention among pollsters to use a 95% confidence level and to report the "worst case" margin of error, based on $p = 0.5$.

Question: How did Fox calculate their margin of error?

Assuming $p = 0.5$, for random samples of $n = 900$, $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.5)(0.5)}{900}} = 0.0167$

For a 95% confidence level, $ME = 2(0.0167) = 0.033$, so Fox's margin of error is just a bit over ±3%.

Critical Values

NOTATION ALERT:

We'll put an asterisk on a letter to indicate a critical value, so z^* is always a critical value from a Normal model.

In our sea fans example we used $2SE$ to give us a 95% confidence interval. To change the confidence level, we'd need to change the *number* of SEs so that the size of the margin of error corresponds to the new level. This number of SEs is called the **critical value.** Here it's based on the Normal model, so we denote it z^*. For any confidence level, we can find the corresponding critical value from a computer, a calculator, or a Normal probability table, such as Table Z.

For a 95% confidence interval, you'll find the precise critical value is $z^* = 1.96$. That is, 95% of a Normal model is found within ± 1.96 standard deviations of the mean. We've been using $z^* = 2$ from the 68–95–99.7 Rule because it's easy to remember.

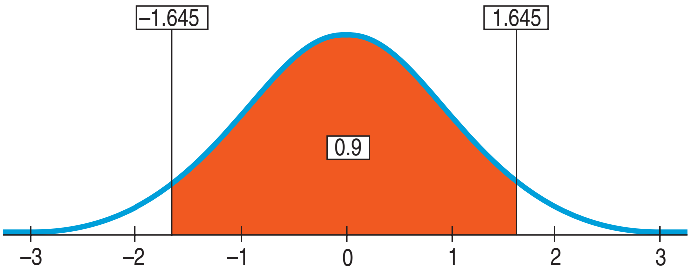


FIGURE 19.4
For a 90% confidence interval, the critical value is 1.645, because, for a Normal model, 90% of the values are within 1.645 standard deviations from the mean.

FOR EXAMPLE Finding the margin of error (Take 2)

Recap: In January 2007 a Fox News poll of 900 registered voters found that 82% of the respondents believed that global warming exists. Fox reported a 95% confidence interval with a margin of error of $\pm 3\%$.

Questions: Using the critical value of z and the standard error based on the observed proportion, what would be the margin of error for a 90% confidence interval? What's good and bad about this change?

With $n = 900$ and $\hat{p} = 0.82$, $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.82)(0.18)}{900}} = 0.0128$

For a 90% confidence level, $z^* = 1.645$, so $ME = 1.645(0.0128) = 0.021$

Now the margin of error is only about $\pm 2\%$, producing a narrower interval. That makes for a more precise estimate of voter belief, but provides less certainty that the interval actually contains the true proportion of voters believing in global warming.

JUST CHECKING

Think some more about the 95% confidence interval Fox News created for the proportion of registered voters who believe that global warming exists.

6. If Fox wanted to be 98% confident, would their confidence interval need to be wider or narrower?

7. Fox's margin of error was about $\pm 3\%$. If they reduced it to $\pm 2\%$, would their level of confidence be higher or lower?

8. If Fox News had polled more people, would the interval's margin of error have been larger or smaller?

Assumptions and Conditions

We've just made some pretty sweeping statements about sea fans. Those statements were possible because we used a Normal model for the sampling distribution. But is that model appropriate?

As we've seen, all statistical models make assumptions. Different models make different assumptions. If those assumptions are not true, the model might be inappropriate and our conclusions based on it may be wrong. Because the confidence interval is built on the Normal model for the sampling distribution, the assumptions and conditions are the same as those we discussed in Chapter 18. But, because they are so important, we'll go over them again.

We can never be certain that an assumption is true, but we can decide intelligently whether it is reasonable. When we have data, we can often decide whether an assumption is plausible by checking a related condition. However, we want to make a statement about the world at large, not just about the data we collected. So the assumptions we make are not just about how our data look, but about how representative they are.

Independence Assumption

A S *Activity:* **Assumptions and Conditions.** Here's an animated review of the assumptions and conditions.

Independence Assumption: We first need to *Think* about whether the independence assumption is plausible. We often look for reasons to suspect that it fails. We wonder whether there is any reason to believe that the data values somehow affect each other. (For example, might the disease in sea fans be contagious?) Whether you decide that the **Independence Assumption** is plausible depends on your knowledge of the situation. It's not one you can check by looking at the data.

However, now that we have data, there are two conditions that we can check:

Randomization Condition: Were the data sampled at random or generated from a properly randomized experiment? Proper randomization can help ensure independence.

10% Condition: Samples are almost always drawn without replacement. Usually, of course, we'd like to have as large a sample as we can. But when the population itself is small we have another concern. When we sample from small populations, the probability of success may be different for the last few individuals we draw than it was for the first few. For example, if most of the women have already been sampled, the chance of drawing a woman from the remaining population is lower. If the sample exceeds 10% of the population, the probability of a success changes so much during the sampling that our Normal model may no longer be appropriate. But if less than 10% of the population is sampled, the effect on independence is negligible.

Sample Size Assumption

The model we use for inference is based on the Central Limit Theorem. The **Sample Size Assumption** addresses the question of whether the sample is large enough to make the sampling model for the sample proportions approximately Normal. It turns out that we need more data as the proportion gets closer and closer to either extreme (0 or 1). We can check this assumption with the:

Success/Failure Condition: We must expect at least 10 "successes" and at least 10 "failures." Recall that by tradition we arbitrarily label one alternative (usually the outcome being counted) as a "success" even if it's something bad (like a sick sea fan). The other alternative is, of course, then a "failure."

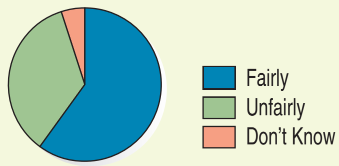
A S ***Activity:* A Confidence Interval for *p*.** View the video story of pollution in Chesapeake Bay, and make a confidence interval for the analysis with the interactive tool.

ONE-PROPORTION *z*-INTERVAL

When the conditions are met, we are ready to find the confidence interval for the population proportion, p. The confidence interval is $\hat{p} \pm z^* \times SE(\hat{p})$ where the standard deviation of the proportion is estimated by $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$.

STEP-BY-STEP EXAMPLE — A Confidence Interval for a Proportion

In May 2006, the Gallup Poll[6] asked 510 randomly sampled adults the question "Generally speaking, do you believe the death penalty is applied fairly or unfairly in this country today?" Of these, 60% answered "Fairly," 35% said "Unfairly," and 4% said they didn't know.



WHO	Adults in the United States
WHAT	Response to a question about the death penalty
WHEN	May 2006
WHERE	United States
HOW	510 adults were randomly sampled and asked by the Gallup Poll
WHY	Public opinion research

Question: From this survey, what can we conclude about the opinions of *all* adults?

To answer this question, we'll build a confidence interval for the proportion of all U.S. adults who believe the death penalty is applied fairly. There are four steps to building a confidence interval for proportions: Plan, Model, Mechanics, and Conclusion.

THINK

Plan State the problem and the W's.

Identify the *parameter* you wish to estimate.

Identify the *population* about which you wish to make statements.

Choose and state a confidence level.

I want to find an interval that is likely, with 95% confidence, to contain the true proportion, p, of U.S. adults who think the death penalty is applied fairly. I have a random sample of 510 U.S. adults.

Model Think about the assumptions and check the conditions.

✔ **Independence Assumption:** Gallup phoned a random sample of U.S. adults. It is very unlikely that any of their respondents influenced each other.

✔ **Randomization Condition:** Gallup drew a random sample from all U.S. adults. I don't have details of their randomization but assume that I can trust it.

✔ **10% Condition:** Although sampling was necessarily without replacement, there are many more U.S. adults than were sampled. The sample is certainly less than 10% of the population.

[6] www.gallup.com

✔ **Success/Failure Condition:**
$n\hat{p} = 510(60\%) = 306 \geq 10$ and
$n\hat{q} = 510(40\%) = 204 \geq 10$,
so the sample appears to be large enough to use the Normal model.

State the sampling distribution model for the statistic.

Choose your method.

The conditions are satisfied, so I can use a Normal model to find a **one-proportion z-interval.**

SHOW

Mechanics Construct the confidence interval.

$n = 510$, $\hat{p} = 0.60$, so

First find the standard error. (Remember: It's called the "standard error" because we don't know p and have to use $\hat{p}$ instead.)

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.60)(0.40)}{510}} = 0.022$$

Next find the margin of error. We could informally use 2 for our critical value, but 1.96 is more accurate.

Because the sampling model is Normal, for a 95% confidence interval, the critical value $z^* = 1.96$.

The margin of error is

$$ME = z^* \times SE(\hat{p}) = 1.96(0.022) = 0.043$$

Write the confidence interval (CI).

So the 95% confidence interval is

$$0.60 \pm 0.043 \text{ or } (0.557, 0.643)$$

REALITY CHECK

The CI is centered at the sample proportion and about as wide as we might expect for a sample of 500.

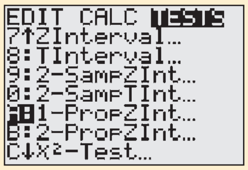
TELL

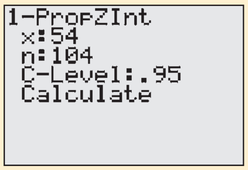
Conclusion Interpret the confidence interval in the proper context. We're 95% confident that our interval captured the true proportion.

I am 95% confident that between 55.7% and 64.3% of all U.S. adults think that the death penalty is applied fairly.

TI Tips

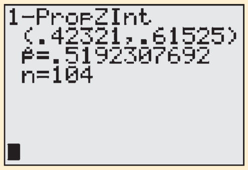
Finding confidence intervals





It will come as no surprise that your TI can calculate a confidence interval for a population proportion. Remember the sea fans? Of 104 sea fans, 54 were diseased. To find the resulting confidence interval, we first take a look at a whole new menu.

- Under `STAT` go to the `TESTS` menu. Quite a list! Commands are found here for the inference procedures you will learn through the coming chapters.
- We're using a Normal model to find a confidence interval for a proportion based on one sample. Scroll down the list and select `A:1-PropZInt`.
- Enter the number of successes observed and the sample size.
- Specify a confidence level and then `Calculate`.



And there it is! Note that the TI calculates the sample proportion for you, but the important result is the interval itself, 42% to 62%. The calculator did the easy part—just Show. Tell is harder. It's your job to interpret that interval correctly.

Beware: You may run into a problem. When you enter the value of `x`, you need a *count*, not a percentage. Suppose the marine scientists had reported that 52% of the 104 sea fans were infected. You can enter `x:.52*104`, and the calculator will evaluate that as 54.08. Wrong. Unless you fix that result, you'll get an error message. Think about it—the number of infected sea fans must have been a whole number, evidently 54. When the scientists reported the results, they rounded off the actual percentage ($54 \div 104 = 51.923\%$) to 52%. Simply change the value of `x` to 54 and you should be able to `Calculate` the correct interval.

Choosing Your Sample Size

The question of how large a sample to take is an important step in planning any study. We weren't ready to make that calculation when we first looked at study design in Chapter 12, but now we can—and we always should.

Suppose a candidate is planning a poll and wants to estimate voter support within 3% with 95% confidence. How large a sample does she need?

Let's look at the margin of error:

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.03 = 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

We want to find n, the sample size. To find n we need a value for $\hat{p}$. We don't know $\hat{p}$ because we don't have a sample yet, but we can probably guess a value. The worst case—the value that makes $\hat{p}\hat{q}$ (and therefore n) largest—is 0.50, so if we use that value for $\hat{p}$, we'll certainly be safe. Our candidate probably expects to be near 50% anyway.

Our equation, then, is

$$0.03 = 1.96\sqrt{\frac{(0.5)(0.5)}{n}}.$$

To solve for n, we first multiply both sides of the equation by $\sqrt{n}$ and then divide by 0.03:

$$0.03\sqrt{n} = 1.96\sqrt{(0.5)(0.5)}$$

$$\sqrt{n} = \frac{1.96\sqrt{(0.5)(0.5)}}{0.03} \approx 32.67$$

Notice that evaluating this expression tells us the *square root* of the sample size. We need to square that result to find n:

$$n \approx (32.67)^2 \approx 1067.1$$

To be safe, we round up and conclude that we need at least 1068 respondents to keep the margin of error as small as 3% with a confidence level of 95%.

What do I use instead of $\hat{p}$?
Often we have an estimate of the population proportion based on experience or perhaps a previous study. If so, use that value as $\hat{p}$ in calculating what size sample you need. If not, the cautious approach is to use $p = 0.5$ in the sample size calculation; that will determine the largest sample necessary regardless of the true proportion.

FOR EXAMPLE **Choosing a sample size**

Recap: The Fox News poll which estimated that 82% of all voters believed global warming exists had a margin of error of $\pm 3\%$. Suppose an environmental group planning a follow-up survey of voters' opinions on global warming wants to determine a 95% confidence interval with a margin of error of no more than $\pm 2\%$.

Question: How large a sample do they need? Use the Fox News estimate as the basis for your calculation.

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.02 = 1.96\sqrt{\frac{(0.82)(0.18)}{n}}$$

$$\sqrt{n} = \frac{1.96\sqrt{(0.82)(0.18)}}{0.02} \approx 37.65$$

$$n = 37.65^2 = 1{,}417.55$$

The environmental group's survey will need about 1,418 respondents.

Public opinion polls often sample 1000 people, which gives an ME of 3% when $p = 0.5$. But businesses and nonprofit organizations typically use much larger samples to estimate the proportion who will accept a direct mail offer. Why? Because that proportion is very low—often far below 5%. An ME of 3% wouldn't be precise enough. An ME like 0.1% would be more useful, and that requires a very large sample size.

Unfortunately, bigger samples cost more money and more effort. Because the standard error declines only with the *square root* of the sample size, to cut the standard error (and thus the ME) in half, we must *quadruple* the sample size.

Generally a margin of error of 5% or less is acceptable, but different circumstances call for different standards. For a pilot study, a margin of error of 10% may be fine, so a sample of 100 will do quite well. In a close election, a polling organization might want to get the margin of error down to 2%. Drawing a large sample to get a smaller ME, however, can run into trouble. It takes time to survey 2400 people, and a survey that extends over a week or more may be trying to hit a target that moves during the time of the survey. An important event can change public opinion in the middle of the survey process.

Keep in mind that the sample size for a survey is the number of respondents, not the number of people to whom questionnaires were sent or whose phone numbers were dialed. And keep in mind that a low response rate turns any study essentially into a voluntary response study, which is of little value for inferring population values. It's almost always better to spend resources on increasing the response rate than on surveying a larger group. A full or nearly full response by a modest-size sample can yield useful results.

Surveys are not the only place where proportions pop up. Banks sample huge mailing lists to estimate what proportion of people will accept a credit card offer. Even pilot studies may mail offers to over 50,000 customers. Most don't respond; that doesn't make the sample smaller—they simply said "No thanks". Those who do respond want the card. To the bank, the response rate[7] is $\hat{p}$. With a typical success rate around 0.5%, the bank needs a very small margin of error—often as low as 0.1%—to make a sound business decision. That calls for a large sample, and the bank must take care in estimating the size needed. For our election poll calculation we used $p = 0.5$, both because it's safe and because we honestly believed p to be near 0.5. If the bank used 0.5, they'd get an absurd answer. Instead, they base their calculation on a proportion closer to the one they expect to find.

[7] In marketing studies every mailing yields a response—"yes" or "no"—and "response rate" means the proportion of customers who accept an offer. That's not the way we use the term for survey response.

FOR EXAMPLE Sample size revisited

A credit card company is about to send out a mailing to test the market for a new credit card. From that sample, they want to estimate the true proportion of people who will sign up for the card nationwide. A pilot study suggests that about 0.5% of the people receiving the offer will accept it.

Question: To be within a tenth of a percentage point (0.001) of the true rate with 95% confidence, how big does the test mailing have to be?

Using the estimate $\hat{p} = 0.5\%$: $ME = 0.001 = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{(0.005)(0.995)}{n}}$

$$(0.001)^2 = 1.96^2 \frac{(0.005)(0.995)}{n} \Rightarrow n = \frac{1.96^2(0.005)(0.995)}{(0.001)^2}$$
$$= 19{,}111.96 \text{ or } 19{,}112$$

That's a lot, but it's actually a reasonable size for a trial mailing such as this. Note, however, that if they had assumed 0.50 for the value of p, they would have found

$$ME = 0.001 = z^* \sqrt{\frac{pq}{n}} = 1.96\sqrt{\frac{(0.5)(0.5)}{n}}$$

$$(0.001)^2 = 1.96^2 \frac{(0.5)(0.5)}{n} \Rightarrow n = \frac{1.96^2(0.5)(0.5)}{(0.001)^2} = 960{,}400.$$

Quite a different (and unreasonable) result.

WHAT CAN GO WRONG?

Confidence intervals are powerful tools. Not only do they tell what we know about the parameter value, but—more important—they also tell what we *don't* know. In order to use confidence intervals effectively, you must be clear about what you say about them.

Don't Misstate What the Interval Means

- **Don't suggest that the parameter varies.** A statement like "There is a 95% chance that the true proportion is between 42.7% and 51.3%" sounds as though you think the population proportion wanders around and sometimes happens to fall between 42.7% and 51.3%. When you interpret a confidence interval, make it clear that *you* know that the population parameter is fixed and that it is the interval that varies from sample to sample.
- **Don't claim that other samples will agree with yours.** Keep in mind that the confidence interval makes a statement about the true population proportion. An interpretation such as "In 95% of samples of U.S. adults, the proportion who think marijuana should be decriminalized will be between 42.7% and 51.3%" is just wrong. The interval isn't about sample proportions but about the population proportion.
- **Don't be certain about the parameter.** Saying "Between 42.1% and 61.7% of sea fans are infected" asserts that the population proportion cannot be outside that interval. Of course, we can't be absolutely certain of that. (Just pretty sure.)
- **Don't forget: It's about the parameter.** Don't say, "I'm 95% confident that $\hat{p}$ is between 42.1% and 61.7%." Of course you are—in fact, we calculated that $\hat{p} = 51.9\%$ of the

(*continued*)

What *Can* I Say?

Confidence intervals are based on random samples, so the interval is random, too. The CLT tells us that 95% of the random samples will yield intervals that capture the true value. That's what we mean by being 95% confident.

Technically, we should say, "I am 95% confident that the interval from 42.1% to 61.7% captures the true proportion of infected sea fans." That formal phrasing emphasizes that *our confidence (and our uncertainty) is about the interval, not the true proportion.* But you may choose a more casual phrasing like "I am 95% confident that between 42.1% and 61.7% of the Las Redes fans are infected." Because you've made it clear that the uncertainty is yours and you didn't suggest that the randomness is in the true proportion, this is OK. Keep in mind that it's the interval that's random and is the focus of both our confidence and doubt.

fans in our sample were infected. So we already *know* the sample proportion. The confidence interval is about the (unknown) population parameter, p.

- **Don't claim to know too much.** Don't say, "I'm 95% confident that between 42.1% and 61.7% of all the sea fans in the world are infected." You didn't sample from all 500 species of sea fans found in coral reefs around the world. Just those of this type on the Las Redes Reef.
- **Do take responsibility.** Confidence intervals are about *un*certainty. *You* are the one who is uncertain, not the parameter. You have to accept the responsibility and consequences of the fact that not all the intervals you compute will capture the true value. In fact, about 5% of the 95% confidence intervals you find will fail to capture the true value of the parameter. You *can* say, "I am 95% confident that between 42.1% and 61.7% of the sea fans on the Las Redes Reef are infected."[8]
- **Do treat the whole interval equally.** Although a confidence interval is a set of plausible values for the parameter, don't think that the values in the middle of a confidence interval are somehow "more plausible" than the values near the edges. Your interval provides no information about where in your current interval (if at all) the parameter value is most likely to be hiding.

Margin of Error Too Large to Be Useful

We know we can't be exact, but how precise do we need to be? A confidence interval that says that the percentage of infected sea fans is between 10% and 90% wouldn't be of much use. Most likely, you have some sense of how large a margin of error you can tolerate. What can you do?

One way to make the margin of error smaller is to reduce your level of confidence. But that may not be a useful solution. It's a rare study that reports confidence levels lower than 80%. Levels of 95% or 99% are more common.

The time to think about whether your margin of error is small enough to be useful is when you design your study. Don't wait until you compute your confidence interval. To get a narrower interval without giving up confidence, you need to have less variability in your sample proportion. How can you do that? Choose a larger sample.

Violations of Assumptions

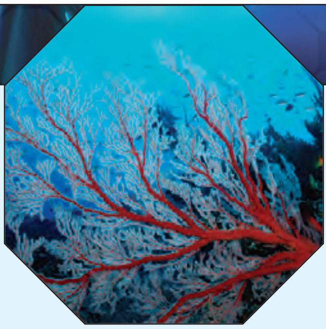
Confidence intervals and margins of error are often reported along with poll results and other analyses. But it's easy to misuse them and wise to be aware of the ways things can go wrong.

- **Watch out for biased sampling.** Don't forget about the potential sources of bias in surveys that we discussed in Chapter 12. Just because we have more statistical machinery now doesn't mean we can forget what we've already learned. A questionnaire that finds that 85% of people enjoy filling out surveys still suffers from nonresponse bias even though now we're able to put confidence intervals around this (biased) estimate.
- **Think about independence.** The assumption that the values in our sample are mutually independent is one that we usually cannot check. It always pays to think about it, though. For example, the disease affecting the sea fans might be contagious, so that fans growing near a diseased fan are more likely themselves to be diseased. Such contagion would violate the Independence Assumption and could severely affect our sample proportion. It could be that the proportion of infected sea fans on the entire reef is actually quite small, and the researchers just happened to find an infected area. To avoid this, the researchers should be careful to sample sites far enough apart to make contagion unlikely.

[8] When we are being very careful we say, "95% of samples of this size will produce confidence intervals that capture the true proportion of infected sea fans on the Las Redes Reef."

CONNECTIONS

Now we can see a practical application of sampling distributions. To find a confidence interval, we lay out an interval measured in standard deviations. We're using the standard deviation as a ruler again. But now the standard deviation we need is the standard deviation of the sampling distribution. That's the one that tells how much the proportion varies. (And when we estimate it from the data, we call it a standard error.)



WHAT HAVE WE LEARNED?

The first 10 chapters of the book explored graphical and numerical ways of summarizing and presenting sample data. We've learned (at last!) to use the sample we have at hand to say something about the *world at large*. This process, called statistical inference, is based on our understanding of sampling models and will be our focus for the rest of the book.

As our first step in statistical inference, we've learned to use our sample to make a *confidence interval* that estimates what proportion of a population has a certain characteristic.

We've learned that:

- Our best estimate of the true population proportion is the proportion we observed in the sample, so we center our confidence interval there.
- Samples don't represent the population perfectly, so we create our interval with a *margin of error.*
- This method successfully captures the true population proportion most of the time, providing us with a level of confidence in our interval.
- The higher the level of confidence we want, the *wider* our confidence interval becomes.
- The larger the sample size we have, the *narrower* our confidence interval can be.
- When designing a study, we can calculate the sample size we'll need to be able to reach conclusions that have a desired degree of precision and level of confidence.
- There are important assumptions and conditions we must check before using this (or any) statistical inference procedure.

We've learned to interpret a confidence interval by *Telling* what we believe is true in the entire population from which we took our random sample. Of course, we can't be *certain.* We've learned not to overstate or misinterpret what the confidence interval says.

Terms

Standard error

440. When we estimate the standard deviation of a sampling distribution using statistics found from the data, the estimate is called a standard error.

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

Confidence interval

441. A level C confidence interval for a model parameter is an interval of values usually of the form

$$estimate \pm margin\ of\ error$$

found from data in such a way that C% of all random samples will yield intervals that capture the true parameter value.

One-proportion z-interval

442–444. A confidence interval for the true value of a proportion. The confidence interval is

$$\hat{p} \pm z^* SE(\hat{p}),$$

where z^* is a critical value from the Standard Normal model corresponding to the specified confidence level.

Margin of error 443. In a confidence interval, the extent of the interval on either side of the observed statistic value is called the margin of error. A margin of error is typically the product of a critical value from the sampling distribution and a standard error from the data. A small margin of error corresponds to a confidence interval that pins down the parameter precisely. A large margin of error corresponds to a confidence interval that gives relatively little information about the estimated parameter. For a proportion,

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Critical value 445. The number of standard errors to move away from the mean of the sampling distribution to correspond to the specified level of confidence. The critical value, denoted z^*, is usually found from a table or with technology.

Skills

THINK

- Understand confidence intervals as a balance between the precision and the certainty of a statement about a model parameter.
- Understand that the margin of error of a confidence interval for a proportion changes with the sample size and the level of confidence.
- Know how to examine your data for violations of conditions that would make inference about a population proportion unwise or invalid.

SHOW

- Be able to construct a one-proportion z-interval.

TELL

- Be able to interpret a one-proportion z-interval in a simple sentence or two. Write such an interpretation so that it does not state or suggest that the parameter of interest is itself random, but rather that the bounds of the confidence interval are the random quantities about which we state our degree of confidence.

CONFIDENCE INTERVALS FOR PROPORTIONS ON THE COMPUTER

Confidence intervals for proportions are so easy and natural that many statistics packages don't offer special commands for them. Most statistics programs want the "raw data" for computations. For proportions, the raw data are the "success" and "failure" status for each case. Usually, these are given as 1 or 0, but they might be category names like "yes" and "no." Often we just know the proportion of successes, $\hat{p}$, and the total count, n. Computer packages don't usually deal with summary data like this easily, but the statistics routines found on many graphing calculators allow you to create confidence intervals from summaries of the data—usually all you need to enter are the number of successes and the sample size.

In some programs you can reconstruct variables of 0's and 1's with the given proportions. But even when you have (or can reconstruct) the raw data values, you may not get *exactly* the same margin of error from a computer package as you would find working by hand. The reason is that some packages make approximations or use other methods. The result is very close but not exactly the same. Fortunately, Statistics means never having to say you're certain, so the approximate result is good enough.