



## Marketing Credit Cards: The MBNA Story

When Delaware substantially raised its interest rate ceiling in 1981, banks and other lending institutions rushed to establish corporate headquarters there. One of these was the Maryland Bank National Association, which established a credit card branch in Delaware using the acronym MBNA. Starting in 1982 with 250 employees in a vacant supermarket in Ogletown, Delaware, MBNA grew explosively in the next two decades.

One of the reasons for this growth was MBNA's use of affinity groups—issuing cards endorsed by alumni associations, sports teams, interest groups, and labor unions, among others. MBNA sold the idea to these groups by letting them share a small percentage of the profit. By 2006, MBNA had become Delaware's largest private employer.

At its peak, MBNA had more than 50 million cardholders and had outstanding credit card loans of \$82.1 billion, making MBNA the third-largest U.S. credit card bank.



“In American corporate history, I doubt there are many companies that burned as brightly, for such a short period of time, as MBNA,” said Rep. Mike Castle, R-Del.<sup>1</sup> MBNA was bought by Bank of America in 2005 for \$35 billion. Bank of America kept the brand briefly before issuing all cards under its own name in 2007.

Unlike the early days of the credit card industry when MBNA established itself, the environment today is intensely competitive, with companies constantly looking for ways to attract new customers and to maximize the profitability of the customers they already have. Many of the large companies have millions of customers, so instead of trying out a new idea with all their customers, they almost always conduct a pilot study or trial first, conducting a survey or an experiment on a sample of their customers.

Credit card companies make money on their cards in three ways: they earn a percentage of every transaction, they charge interest on balances that are not paid in full, and they collect fees (yearly fees, late fees, etc.). To generate all three types of revenue, the marketing departments of credit card banks constantly seek ways to encourage customers to increase the use of their cards.

A marketing specialist at one company had an idea of offering double air miles to their customers with an airline-affiliated card if they increased their spending by at least \$800 in the month following the offer. To forecast the cost and revenue of the offer, the finance department needed to know what percent of customers would actually qualify for the double miles. The marketer decided to send the offer to a random sample of 1000 customers to find out. In that sample, she found that 211 (21.1%) of the cardholders increased their spending by more than the required \$800. But, another analyst drew a different sample of 1000 customers of whom 202 (20.2%) of the cardholders exceeded \$800.

The two samples don't agree. We know that observations vary, but how much variability among samples should we expect to see?

Why do sample proportions vary at all? How can two samples of the same population measuring the same quantity get different results? The answer is fundamental to statistical inference. Each proportion is based on a *different* sample of cardholders. The proportions vary from sample to sample because the samples are comprised of different people.

**WHO** Cardholders of a bank's credit card  
**WHAT** Whether cardholders increased their spending by at least \$800 in the subsequent month  
**WHEN** February 2008  
**WHERE** United States  
**WHY** To predict costs and benefits of a program offer

## 10.1 The Distribution of Sample Proportions

We'd like to know how much proportions can vary from sample to sample. We've talked about *Plan*, *Do*, and *Report*, but to learn more about the variability, we have to add *Imagine*. When we sample, we see only the results from the actual sample that we draw, but we can *imagine* what we might have seen had we drawn *all* other possible random samples. What would the histogram of all those sample proportions look like?

If we could take many random samples of 1000 cardholders, we would find the proportion of each sample who spent more than \$800 and collect all of those proportions into a histogram. Where would you expect the center of that histogram to be? Of course, we don't *know* the answer, but it is reasonable to think that it will be at the true proportion in the population. We probably will never know the value of the true proportion. But it is important to us, so we'll give it a label,  $p$  for “true proportion.”

### Imagine

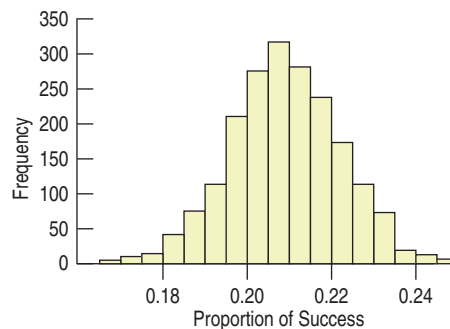
We see only the sample we actually drew, but if we *imagine* the results of all the other possible samples we could have drawn (by modeling or simulating them), we can learn more.

<sup>1</sup>Delaware News Online, January 1, 2006.

In fact, we can do better than just imagining. We can *simulate*. We can't really take all those different random samples of size 1000, but we can use a computer to pretend to draw random samples of 1000 individuals from some population of values over and over. In this way, we can study the process of drawing many samples from a real population. A *simulation* can help us understand how sample proportions vary due to random sampling.

When we have only two possible outcomes for an event, the convention in Statistics is to arbitrarily label one of them “success” and the other “failure.” Here, a “success” would be that a customer increases card charges by at least \$800, and a “failure” would be that the customer didn't. In the simulation, we'll set the true proportion of successes to a known value, draw random samples, and then record the sample proportion of successes, which we'll denote by  $\hat{p}$ , for each sample.

The proportion of successes in each of our simulated samples will vary from one sample to the next, but the *way* in which the proportions vary shows us the proportions of real samples would vary. Because we can specify the true proportion of successes, we can see how close each sample comes to estimating that true value. Here's a histogram of the proportions of cardholders who increased spending by at least \$800 in 2000 independent samples of 1000 cardholders, when the true proportion is  $p = 0.21$ . (We know this is the true value of  $p$  because in a simulation we can control it.)



**Figure 10.1** The distribution of 2000 sample values of  $\hat{p}$ , from simulated samples of size 1000 drawn from a population in which the true  $p$  is 0.21.

It should be no surprise that we don't get the same proportion for each sample we draw, even though the underlying true value,  $p$ , stays the same at  $p = 0.21$ . Since each  $\hat{p}$  comes from a random sample, we don't expect them to all be equal to  $p$ . And since each comes from a *different* independent random sample, we don't expect them to be equal to each other, either. The remarkable thing is that even though the  $\hat{p}$ 's vary from sample to sample, they do so in a way that we can model and understand.

## For Example

### The distribution of a sample proportion

A supermarket has installed “self-checkout” stations that allow customers to scan and bag their own groceries. These are popular, but because customers occasionally encounter a problem, a staff member must be available to help out. The manager wants to estimate what proportion of customers need help so that he can optimize the number of self-check stations per staff member. He collects data from the stations for 30 days, recording the proportion of customers on each day that need help and makes a histogram of the observed proportions.

#### Questions:

1. If the proportion needing help is independent from day to day, what shape would you expect his histogram to follow?
2. Is the assumption of independence reasonable?

#### Answers:

1. Normal, centered at the true proportion.
2. Possibly not. For example, shoppers on weekends might be less experienced than regular weekday shoppers and would then need more help.

## 10.2 Sampling Distribution for Proportions

The collection of  $\hat{p}$ 's may be better behaved than you expected. The histogram in Figure 10.1 is unimodal and symmetric. It is also bell-shaped—and that means that the Normal may be an appropriate model. It is one of the early discoveries and successes of statistics that this distribution of sample proportions can be modeled by the Normal.

The distribution that we displayed in Figure 10.1 is just one simulation. The more proportions we simulate, the more the distribution settles down to a smooth bell shape. The distribution of proportions over all possible independent samples from the same population is called the **sampling distribution** of the proportions.

In fact, we can use the Normal model to describe the behavior of proportions. With the Normal model, we can find the percentage of values falling between any two values. But to make that work, we need to know the mean and standard deviation.

We know already that a sampling distribution of a sample proportion is centered around the true proportion,  $p$ . An amazing fact about proportions gives us the appropriate standard deviation to use as well. It turns out that once we know the mean,  $p$ , and the sample size,  $n$ , we also know the standard deviation of the sampling distribution as you can see from its formula:

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

If the true proportion of credit cardholders who increased their spending by more than \$800 is 0.21, then for samples of size 1000, we expect the distribution of sample proportions to have a standard deviation of:

$$SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.21(1-0.21)}{1000}} = 0.0129, \text{ or about } 1.3\%.$$

Remember that the two samples of size 1000 had proportions of 21.1% and 20.2%. Since the standard deviation of proportions is 1.3%, these two proportions are not even a full standard deviation apart. In other words, the two samples don't really disagree. Proportions of 21.1% and 20.2% from samples of 1000 are both *consistent* with a true proportion of 21%. We know from Chapter 3 that this difference between sample proportions is referred to as **sampling error**. But it's not really an *error*. It's just the *variability* you'd expect to see from one sample to another. A better term might be *sampling variability*.

Look back at Figure 10.1 to see how well the model worked in our simulation. If  $p = 0.21$ , we now know that the standard deviation should be about 0.013. The 68–95–99.7 Rule from the Normal model says that 68% of the samples will have proportions within 1 SD of the mean of 0.21. How closely does our simulation match the predictions? The actual standard deviation of our 2000 *sample* proportions is 0.0129 or 1.29%. And, of the 2000 simulated samples, 1346 of them had proportions between 0.197 and .223 (one standard deviation on either side of 0.21). The 68–95–99.7 Rule predicts 68%—the actual number is 1346/2000 or 67.3%.

Now we know everything we need to know to model the sampling distribution. We know the mean and standard deviation of the sampling distribution of proportions: they're  $p$ , the true population proportion, and  $\sqrt{\frac{pq}{n}}$ . So the particular Normal model,  $N\left(p, \sqrt{\frac{pq}{n}}\right)$ , is a **sampling distribution model for the sample proportion**.

We saw this worked well in a simulation, but can we rely on it in all situations? It turns out that this model can be justified theoretically with just a little mathematics.

### Notation Alert!

We use  $p$  for the proportion in the population and  $\hat{p}$  for the observed proportion in a sample. We'll also use  $q$  for the proportion of failures ( $q = 1 - p$ ), and  $\hat{q}$  for its observed value, just to simplify some formulas.

We have now answered the question raised at the start of the chapter. To discover how variable a sample proportion is, we need to know the proportion and the size of the sample. That's all.

### Effect of Sample Size

Because  $n$  is in the denominator of  $SD(\hat{p})$ , the larger the sample, the smaller the standard deviation. We need a small standard deviation to make sound business decisions, but larger samples cost more. That tension is a fundamental issue in statistics.

It won't work for *all* situations, but it works for most situations that you'll encounter in practice. We'll provide conditions to check so you'll know when the model is useful.

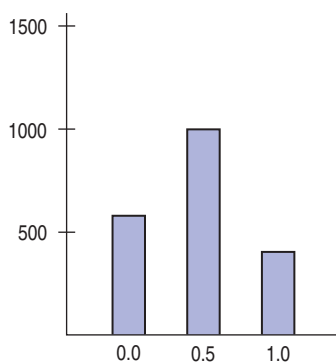
### The sampling distribution model for a proportion

Provided that the sampled values are independent and the sample size is large enough, the sampling distribution of  $\hat{p}$  is modeled by a Normal model with mean  $\mu(\hat{p}) = p$  and standard deviation  $SD(\hat{p}) = \sqrt{\frac{pq}{n}}$ .

### Just Checking

- 1 You want to poll a random sample of 100 shopping mall customers about whether they like the proposed location for the new coffee shop on the third floor, with a panoramic view of the food court. Of course, you'll get just one number, your sample proportion,  $\hat{p}$ . But if you imagined all the possible samples of 100 customers you could draw and imagined the histogram of all the sample proportions from these samples, what shape would it have?
- 2 Where would the center of that histogram be?
- 3 If you think that about half the customers are in favor of the plan, what would the standard deviation of the sample proportions be?

The sampling distribution model for  $\hat{p}$  is valuable for a number of reasons. First, because it is known from mathematics to be a good model (and one that gets better and better as the sample size gets larger), we don't need to actually draw many samples and accumulate all those sample proportions, or even to simulate them. The Normal sampling distribution model tells us what the distribution of sample proportions would look like. Second, because the Normal model is a mathematical model, we can calculate what fraction of the distribution will be found in any region. You can find the fraction of the distribution in *any* interval of values using Table Z at the back of the book or with technology.



**Figure 10.2** Proportions from samples of size 2 can take on only three possible values. A Normal model does not work well here.

### How Good Is the Normal Model?

We've seen that the simulated proportions follow the 68–95–99.7 Rule well. But do all sample proportions really work like this? Stop and think for a minute about what we're claiming. We've said that if we draw repeated random samples of the same size,  $n$ , from some population and measure the proportion,  $\hat{p}$ , we get for each sample, then the collection of these proportions will pile up around the underlying population proportion,  $p$ , in such a way that a histogram of the sample proportions can be modeled well by a Normal model.

There must be a catch. Suppose the samples were of size 2, for example. Then the only possible numbers of successes could be 0, 1, or 2, and the proportion values would be 0, 0.5, and 1. There's no way the histogram could ever look like a Normal model with only three possible values for the variable (Figure 10.2).

Well, there *is* a catch. The claim is only approximately true. (But, that's fine. Models are *supposed* to be only approximately true.) And the model becomes a better and better representation of the distribution of the sample proportions as the sample size gets bigger.<sup>2</sup> Samples of size 1 or 2 just aren't going to work very well, but the distributions of proportions of many larger samples do have histograms that are remarkably close to a Normal model.

<sup>2</sup>Formally, we say the claim is true in the limit as the sample size ( $n$ ) grows.

## For Example

## Sampling distribution for proportions

Time-Warner provides cable, phone, and Internet services to customers, some of whom subscribe to “packages” including several services. Nationwide, suppose that 30% of their customers are “package subscribers” and subscribe to all three types of service. A local representative in Phoenix, Arizona, wonders if the proportion in his region is the same as the national proportion.

**Questions:** If the same proportion holds in his region and he takes a survey of 100 customers at random from his subscriber list:

1. What proportion of customers would you expect to be package subscribers?
2. What is the standard deviation of the sample proportion?
3. What shape would you expect the sampling distribution of the proportion to have?
4. Would you be surprised to find out that in a sample of 100, 49 of the customers are package subscribers? Explain. What might account for this high percentage?

**Answers:**

1. Because 30% of customers nationwide are package subscribers, we would expect the same for the sample proportion.
2. The standard deviation is  $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.3)(0.7)}{100}} = 0.046$ .
3. Normal.
4. 49 customers results in a sample proportion of 0.49. The mean is 0.30 with a standard deviation of 0.046. This sample proportion is more than 4 standard deviations higher than the mean:  $\frac{(0.49 - 0.30)}{0.046} = 4.13$ . It would be very unusual to find such a large proportion in a random sample. Either it is a very unusual sample, or the proportion in his region is not the same as the national average.

## Assumptions and Conditions

Most models are useful only when specific assumptions are true. In the case of the model for the distribution of sample proportions, there are two assumptions:

**Independence Assumption:** The sampled values must be *independent* of each other.

**Sample Size Assumption:** The sample size,  $n$ , must be *large* enough.

Of course, the best we can do with assumptions is to think about whether they are likely to be true, and we should do so. However, we often can check corresponding *conditions* that provide information about the assumptions as well. Think about the Independence Assumption and check the following corresponding conditions before using the Normal model to model the distribution of sample proportions:

**Randomization Condition:** If your data come from an experiment, subjects should have been randomly assigned to treatments. If you have a survey, your sample should be a simple random sample of the population. If some other sampling design was used, be sure the sampling method was not biased and that the data are representative of the population.

**10% Condition:** If sampling has not been made with replacement (that is, returning each sampled individual to the population before drawing the next individual), then the sample size,  $n$ , must be no larger than 10% of the population. If it is, you must adjust the size of the confidence interval with methods more advanced than those found in this book.

**Success/Failure Condition:** The Success/Failure condition says that the sample size must be big enough so that both the number of “successes,”  $np$ , and the number of “failures,”  $nq$ , are expected to be at least 10.<sup>3</sup> Expressed

<sup>3</sup>We saw where the 10 came from in the Math Box on page 262.

without the symbols, this condition just says that we need to expect at least 10 successes and at least 10 failures to have enough data for sound conclusions. For the bank's credit card promotion example, we labeled as a "success" a cardholder who increases monthly spending by at least \$800 during the trial. The bank observed 211 successes and 789 failures. Both are at least 10, so there are certainly enough successes and enough failures for the condition to be satisfied.<sup>4</sup>

These two conditions seem to contradict each other. The Success/Failure condition wants a big sample size. How big depends on  $p$ . If  $p$  is near 0.5, we need a sample of only 20 or so. If  $p$  is only 0.01, however, we'd need 1000. But the 10% condition says that the sample size can't be too large a fraction of the population. Fortunately, the tension between them isn't usually a problem in practice. Often, as in polls that sample from all U.S. adults, or industrial samples from a day's production, the populations are much larger than 10 times the sample size.

## For Example

### Assumptions and conditions for sample proportions

The analyst conducting the Time-Warner survey says that, unfortunately, only 20 of the customers he tried to contact actually responded, but that of those 20, 8 are package subscribers.

#### Questions:

1. If the proportion of package subscribers in his region is 0.30, how many package subscribers, on average, would you expect in a sample of 20?
2. Would you expect the shape of the sampling distribution of the proportion to be Normal? Explain.

#### Answers:

1. You would expect  $0.30 \cdot 20 = 6$  package subscribers.
2. No. Because 6 is less than 10, we should be cautious in using the Normal as a model for the sampling distribution of proportions. (The number of *observed* successes, 8, is also less than 10.)

## Guided Example Foreclosures



An analyst at a home loan lender was looking at a package of 90 mortgages that the company had recently purchased in central California. The analyst was aware that in that region about 13% of the homeowners with current mortgages will

default on their loans in the next year and the house will go into foreclosure. In deciding to buy the collection of mortgages, the finance department assumed that no more than 15 of the mortgages would go into default. Any amount above that will result in losses for the company. In the package of 90 mortgages, what's the probability that there will be more than 15 foreclosures?

### PLAN

**Setup** State the objective of the study.

We want to find the probability that in a group of 90 mortgages, more than 15 will default. Since 15 out of 90 is 16.7%, we need the probability of finding more than 16.7% defaults out of a sample of 90, if the proportion of defaults is 13%.

(continued)

<sup>4</sup>The Success/Failure condition is about the number of successes and failures we *expect*, but if the number of successes and failures that *occurred* is  $\geq 10$ , then you can use that.

**Model** Check the conditions.

State the parameters and the sampling distribution model.

**Plot** Make a picture. Sketch the model and shade the area we're interested in, in this case the area to the right of 16.7%.

- ✓ **Independence Assumption** If the mortgages come from a wide geographical area, one homeowner defaulting should not affect the probability that another does. However, if the mortgages come from the same neighborhood(s), the independence assumption may fail and our estimates of the default probabilities may be wrong.
- ✓ **Randomization Condition.** The 90 mortgages in the package can be considered as a random sample of mortgages in the region.
- ✓ **10% Condition.** The 90 mortgages are less than 10% of the population.
- ✓ **Success/Failure Condition**

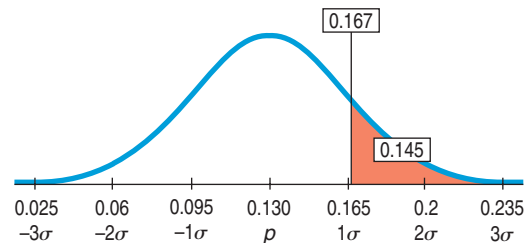
$$np = 90(0.13) = 11.7 \geq 10$$

$$nq = 90(0.87) = 78.3 \geq 10$$

The population proportion is  $p = 0.13$ . The conditions are satisfied, so we'll model the sampling distribution of  $\hat{p}$  with a Normal model, with mean 0.13 and standard deviation

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.13)(0.87)}{90}} \approx 0.035.$$

Our model for  $\hat{p}$  is  $N(0.13, 0.035)$ . We want to find  $P(\hat{p} > 0.167)$ .



**DO** **Mechanics** Use the standard deviation as a ruler to find the  $z$ -score of the cutoff proportion. Find the resulting probability from a table, a computer program, or a calculator.

$$z = \frac{\hat{p} - p}{SD(\hat{p})} = \frac{0.167 - 0.13}{0.035} = 1.06$$

$$P(\hat{p} > 0.167) = P(z > 1.06) = 0.1446$$

**REPORT** **Conclusion** Interpret the probability in the context of the question.

**MEMO**

**Re: Mortgage Defaults**

Assuming that the 90 mortgages we recently purchased are a random sample of mortgages in this region, there is about a 14.5% chance that we will exceed the 15 foreclosures that Finance has determined as the break-even point.

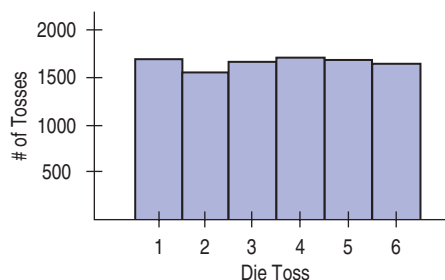


## 10.3 The Central Limit Theorem

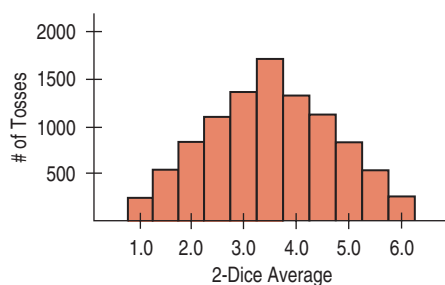
Proportions summarize categorical variables. When we sample at random, the results we get will vary from sample to sample. The Normal model seems an incredibly simple way to summarize all that variation. Could something that simple work for means? We won't keep you in suspense. It turns out that means also have a sampling distribution that we can model with a Normal model. And it turns out that there's a theoretical result that proves it to be so. As we did with proportions, we can get some insight from a simulation.

### Simulating the Sampling Distribution of a Mean

Here's a simple simulation with a quantitative variable. Let's start with one fair die. If we toss this die 10,000 times, what should the histogram of the numbers on the face of the die look like? Here are the results of a simulated 10,000 tosses:

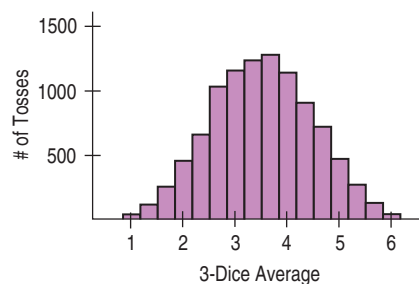


That's called the *uniform distribution*, and it's certainly not Normal. Now let's toss a *pair* of dice and record the average of the two. If we repeat this (or at least simulate repeating it) 10,000 times, recording the average of each pair, what will the histogram of these 10,000 averages look like? Before you look, think a minute. Is getting an average of 1 on *two* dice as likely as getting an average of 3 or 3.5? Let's see:



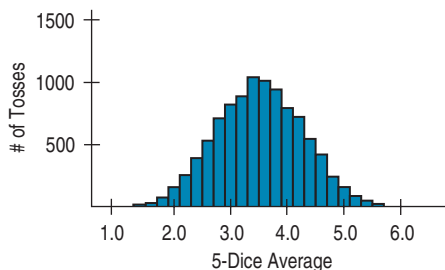
We're much more likely to get an average near 3.5 than we are to get one near 1 or 6. Without calculating those probabilities exactly, it's fairly easy to see that the *only* way to get an average of 1 is to get two 1s. To get a total of 7 (for an average of 3.5), though, there are many more possibilities. This distribution even has a name—the *triangular distribution*.

What if we average three dice? We'll simulate 10,000 tosses of three dice and take their average.



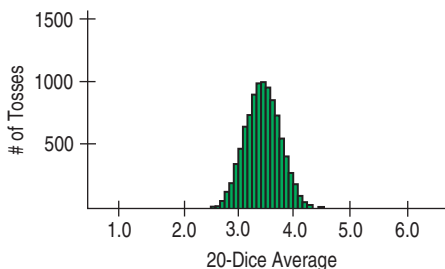
What's happening? First notice that it's getting harder to have averages near the ends. Getting an average of 1 or 6 with three dice requires all three to come up 1 or 6, respectively. That's less likely than for two dice to come up both 1 or both 6. The distribution is being pushed toward the middle. But what's happening to the shape?

Let's continue this simulation to see what happens with larger samples. Here's a histogram of the averages for 10,000 tosses of five dice.



The pattern is becoming clearer. Two things are happening. The first fact we knew already from the Law of Large Numbers, which we saw in Chapter 7. It says that as the sample size (number of dice) gets larger, each sample average tends to become closer to the population mean. So we see the shape continuing to tighten around 3.5. But the shape of the distribution is the surprising part. It's becoming bell-shaped. In fact, it's approaching the Normal model.

Are you convinced? Let's skip ahead and try 20 dice. The histogram of averages for throws 10,000 of 20 dice looks like this.



Now we see the Normal shape again (and notice how much smaller the spread is). But can we count on this happening for situations other than dice throws? What kinds of sample means have sampling distributions that we can model with a Normal model? It turns out that Normal models work well amazingly often.



| Pierre-Simon Laplace, 1749–1827.

*“The theory of probabilities is at bottom nothing but common sense reduced to calculus.”*

—LAPLACE, IN *THÉORIE ANALYTIQUE DES PROBABILITIÉS*, 1812

## The Central Limit Theorem

The dice simulation may look like a special situation. But it turns out that what we saw with dice is true for means of repeated samples for almost every situation. When we looked at the sampling distribution of a proportion, we had to check only a few conditions. For means, the result is even more remarkable. There are almost no conditions at all.

Let's say that again: **The sampling distribution of *any* mean becomes Normal as the sample size grows.** All we need is for the observations to be independent and collected with randomization. We don't even care about the shape of the population distribution!<sup>5</sup> This surprising fact was proved in a fairly general form in 1810 by Pierre-Simon Laplace, and caused quite a stir (at least in mathematics circles)

<sup>5</sup>Technically, the data must come from a population with a finite variance.

**Pierre-Simon Laplace**

Laplace was one of the greatest scientists and mathematicians of his time. In addition to his contributions to probability and statistics, he published many new results in mathematics, physics, and astronomy (where his nebular theory was one of the first to describe the formation of the solar system in much the way it is understood today). He also played a leading role in establishing the metric system of measurement.

His brilliance, though, sometimes got him into trouble. A visitor to the Académie des Sciences in Paris reported that Laplace let it be known widely that he considered himself the best mathematician in France. The effect of this on his colleagues was not eased by the fact that Laplace was right.

because it is so unintuitive. Laplace's result is called the **Central Limit Theorem**<sup>6</sup> (CLT).

Not only does the distribution of means of many random samples get closer and closer to a Normal model as the sample size grows, but *this is true regardless of the shape of the population distribution!* Even if we sample from a skewed or bimodal population, the Central Limit Theorem tells us that means of repeated random samples will tend to follow a Normal model as the sample size grows. Of course, you won't be surprised to learn that it works better and faster the closer the population distribution is to a Normal model. And it works better for larger samples. If the data come from a population that's exactly Normal to start with, then the observations themselves are Normal. If we take samples of size 1, their "means" are just the observations—so, of course, they have a Normal sampling distribution. But now suppose the population distribution is very skewed (like the CEO data from Chapter 5, for example). The CLT works, although it may take a sample size of dozens or even hundreds of observations for the Normal model to work well.

For example, think about a real bimodal population, one that consists of only 0s and 1s. The CLT says that even means of samples from this population will follow a Normal sampling distribution model. But wait. Suppose we have a categorical variable and we assign a 1 to each individual in the category and a 0 to each individual not in the category. Then we find the mean of these 0s and 1s. That's the same as counting the number of individuals who are in the category and dividing by  $n$ . That mean will be the *sample proportion*,  $\hat{p}$ , of individuals who are in the category (a "success"). So maybe it wasn't so surprising after all that proportions, like means, have Normal sampling distribution models; proportions are actually just a special case of Laplace's remarkable theorem. Of course, for such an extremely bimodal population, we need a reasonably large sample size—and that's where the Success/Failure condition for proportions comes in.

**The Central Limit Theorem (CLT)**

The mean of a random sample has a sampling distribution whose shape can be approximated by a Normal model. The larger the sample, the better the approximation will be.

Be careful. We have been slipping smoothly between the real world, in which we draw random samples of data, and a magical mathematical-model world, in which we describe how the sample means and proportions we observe in the real world might behave if we could see the results from every random sample that we might have drawn. Now we have *two* distributions to deal with. The first is the real-world distribution of the sample, which we might display with a histogram (for quantitative data) or with a bar chart or table (for categorical data). The second is the math-world *sampling distribution* of the statistic, which we model with a Normal model based on the Central Limit Theorem. Don't confuse the two.

For example, don't mistakenly think the CLT says that the *data* are Normally distributed as long as the sample is large enough. In fact, as samples get larger, we expect the distribution of the data to look more and more like the distribution of the population from which it is drawn—skewed, bimodal, whatever—but not necessarily Normal. You can collect a sample of CEO salaries for the next 1000 years, but the histogram will never look Normal. It will be skewed to the right.

<sup>6</sup>The word "central" in the name of the theorem means "fundamental." It doesn't refer to the center of a distribution.

The Central Limit Theorem doesn't talk about the distribution of the data from the sample. It talks about the sample *means* and sample *proportions* of many different random samples drawn from the same population. Of course, we never actually draw all those samples, so the CLT is talking about an imaginary distribution—the sampling distribution model.

The CLT does require that the sample be big enough when the population shape is not unimodal and symmetric. But it is still a very surprising and powerful result.

## For Example

### The Central Limit Theorem

The supermarket manager in the example on page 279 also examines the amount spent by customers using the self-checkout stations. He finds that the distribution of these amounts is unimodal but skewed to the high end because some customers make unusually expensive purchases. He finds the mean spent on each of the 30 days studied and makes a histogram of those values.

#### Questions:

1. What shape would you expect for this histogram?
2. If, instead of averaging all customers on each day, he selects the first 10 for each day and just averages those, how would you expect his histogram of the means to differ from the one in (1)?

#### Answers:

1. Normal. It doesn't matter that the sample is drawn from a skewed distribution; the CLT tells us that the means will follow a Normal model.
2. The CLT requires large samples. Samples of 10 are not large enough.

## 10.4 The Sampling Distribution of the Mean

The CLT says that the sampling distribution of any mean or proportion is approximately Normal. But which Normal? We know that any Normal model is specified by its mean and standard deviation. For proportions, the sampling distribution is centered at the population proportion. For means, it's centered at the population mean. What else would we expect?

What about the standard deviations? We noticed in our dice simulation that the histograms got narrower as the number of dice we averaged increased. This shouldn't be surprising. Means vary less than the individual observations. Think about it for a minute. Which would be more surprising, having *one* person in your Statistics class who is over 6'9" tall or having the *mean* of 100 students taking the course be over 6'9"? The first event is fairly rare.<sup>7</sup> You may have seen somebody this tall in one of your classes sometime. But finding a class of 100 whose mean height is over 6'9" tall just won't happen. Why? *Means have smaller standard deviations than individuals.*

"The *n*'s justify the means."

—APOCRYPHAL STATISTICAL SAYING

That is, the Normal model for the sampling distribution of the mean has a standard deviation equal to  $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$  where  $\sigma$  is the standard deviation of the population. To emphasize that this is a standard deviation *parameter* of the sampling distribution model for the sample mean,  $\bar{y}$ , we write  $SD(\bar{y})$  or  $\sigma(\bar{y})$ .

<sup>7</sup>If students are a random sample of adults, fewer than 1 out of 10,000 should be taller than 6'9". Why might college students not really be a random sample with respect to height? Even if they're not a perfectly random sample, a college student over 6'9" tall is still rare.

### The sampling distribution model for a mean

When a random sample is drawn from any population with mean  $\mu$  and standard deviation  $\sigma$ , its sample mean,  $\bar{y}$ , has a sampling distribution with the same mean  $\mu$  but whose standard deviation is  $\frac{\sigma}{\sqrt{n}}$ , and we write  $\sigma(\bar{y}) = SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$ . No matter what population the random sample comes from, the shape of the sampling distribution is approximately Normal as long as the sample size is large enough. The larger the sample used, the more closely the Normal approximates the sampling distribution model for the mean.

We now have two closely related sampling distribution models. Which one we use depends on which kind of data we have.

- When we have categorical data, we calculate a sample proportion,  $\hat{p}$ . Its sampling distribution follows a Normal model with a mean at the population proportion,  $p$ , and a standard deviation  $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \frac{\sqrt{pq}}{\sqrt{n}}$ .
- When we have quantitative data, we calculate a sample mean,  $\bar{y}$ . Its sampling distribution has a Normal model with a mean at the population mean,  $\mu$ , and a standard deviation  $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$ .

The means of these models are easy to remember, so all you need to be careful about is the standard deviations. Remember that these are standard deviations of the *statistics*  $\hat{p}$  and  $\bar{y}$ . They both have a square root of  $n$  in the denominator. That tells us that the larger the sample, the less either statistic will vary. The only difference is in the numerator. If you just start by writing  $SD(\bar{y})$  for quantitative data and  $SD(\hat{p})$  for categorical data, you'll be able to remember which formula to use.

## Assumptions and Conditions for the Sampling Distribution of the Mean

The CLT requires essentially the same assumptions as we saw for modeling proportions:

**Independence Assumption:** The sampled values must be independent of each other.

**Randomization Condition:** The data values must be sampled randomly, or the concept of a sampling distribution makes no sense.

**Sample Size Assumption:** The sample size must be sufficiently large. We can't check these directly, but we can think about whether the Independence Assumption is plausible. We can also check some related conditions:

**10% Condition:** When the sample is drawn without replacement (as is usually the case), the sample size,  $n$ , should be no more than 10% of the population.

**Large Enough Sample Condition:** The CLT doesn't tell us how large a sample we need. The truth is, it depends; there's no one-size-fits-all rule. If the population is unimodal and symmetric, even a fairly small sample is okay. You may hear that 30 or 50 observations is always enough to guarantee Normality, but in truth, it depends on the shape of the original data distribution. For highly skewed distributions, it may require samples of several hundred for the sampling distribution of means to be approximately Normal. Always plot the data to check.

## Sample Size—Diminishing Returns

The standard deviation of the sampling distribution declines only with the square root of the sample size. The mean of a random sample of 4 has half  $\left(\frac{1}{\sqrt{4}} = \frac{1}{2}\right)$  the standard deviation of an individual data value. To cut it in half again, we'd need a sample of 16, and a sample of 64 to halve it once more. In practice, random sampling works well, and means have smaller standard deviations than the individual data values that were averaged. This is the power of averaging.

If only we could afford a much larger sample, we could get the standard deviation of the sampling distribution *really* under control so that the sample mean could tell us still more about the unknown population mean. As we shall see, that square root limits how much we can make a sample tell about the population. This is an example of something that's known as the **Law of Diminishing Returns**.

### For Example

#### Working with the sampling distribution of the mean

Suppose that the weights of boxes shipped by a company follow a unimodal, symmetric distribution with a mean of 12 lbs and a standard deviation of 4 lbs. Boxes are shipped in palettes of 10 boxes. The shipper has a limit of 150 lbs for such shipments.

**Question:** What's the probability that a palette will exceed that limit?

**Answer:** Asking the probability that the total weight of a sample of 10 boxes exceeds 150 lbs is the same as asking the probability that the *mean* weight exceeds 15 lbs. First we'll check the conditions. We will assume that the 10 boxes on the palette are a random sample from the population of boxes and that their weights are mutually independent. We are told that the underlying distribution of weights is unimodal and symmetric, so a sample of 10 boxes should be large enough. And 10 boxes is surely less than 10% of the population of boxes shipped by the company.

Under these conditions, the CLT says that the sampling distribution of  $\bar{y}$  has a Normal model with mean 12 and standard deviation

$$SD(\bar{y}) = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{10}} = 1.26 \text{ and } z = \frac{\bar{y} - \mu}{SD(\bar{y})} = \frac{15 - 12}{1.26} = 2.38$$

$$P(\bar{y} > 150) = P(z > 2.38) = 0.0087$$

So the chance that the shipper will reject a palette is only .0087—less than 1%.

## 10.5 How Sampling Distribution Models Work

Both of the sampling distributions we've looked at are Normal. We know for proportions,  $SD(\hat{p}) = \sqrt{\frac{pq}{n}}$ , and for means,  $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$ . These are great if we know, or can pretend that we know,  $p$  or  $\sigma$ , and sometimes we'll do that.

Often we know only the observed proportion,  $\hat{p}$ , or the sample standard deviation,  $s$ . 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 (SE).**

For a sample proportion,  $\hat{p}$ , the standard error is:

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

For the sample mean,  $\bar{y}$ , the standard error is:

$$SE(\bar{y}) = \frac{s}{\sqrt{n}}$$

You may see a “standard error” reported by a computer program in a summary or offered by a calculator. It’s safe to assume that if no statistic is specified, what was meant is  $SE(\bar{y})$ , the standard error of the mean.

### Just Checking

- 4 The entrance exam for business schools, the GMAT, given to 100 students had a mean of 520 and a standard deviation of 120. What was the standard error for the mean of this sample of students?
- 5 As the sample size increases, what happens to the standard error, assuming the standard deviation remains constant?
- 6 If the sample size is doubled, what is the impact on the standard error?

To keep track of how the concepts we’ve seen combine, we can draw a diagram relating them. At the heart is the idea that *the statistic itself (the proportion or the mean) is a random quantity*. We can’t know what our statistic will be because it comes from a random sample. A different random sample would have given a different result. This sample-to-sample variability is what generates the sampling distribution, the distribution of all the possible values that the statistic could have had.

We could simulate that distribution by pretending to take lots of samples. Fortunately, for the mean and the proportion, the CLT tells us that we can model their sampling distribution directly with a Normal model.

The two basic truths about sampling distributions are:

1. Sampling distributions arise because samples vary. Each random sample will contain different cases and, so, a different value of the statistic.
2. Although we can always simulate a sampling distribution, the Central Limit Theorem saves us the trouble for means and proportions.

When we don’t know  $\sigma$ , we estimate it with the standard deviation of the one real sample. That gives us the standard error,  $SE(\bar{y}) = \frac{s}{\sqrt{n}}$ .

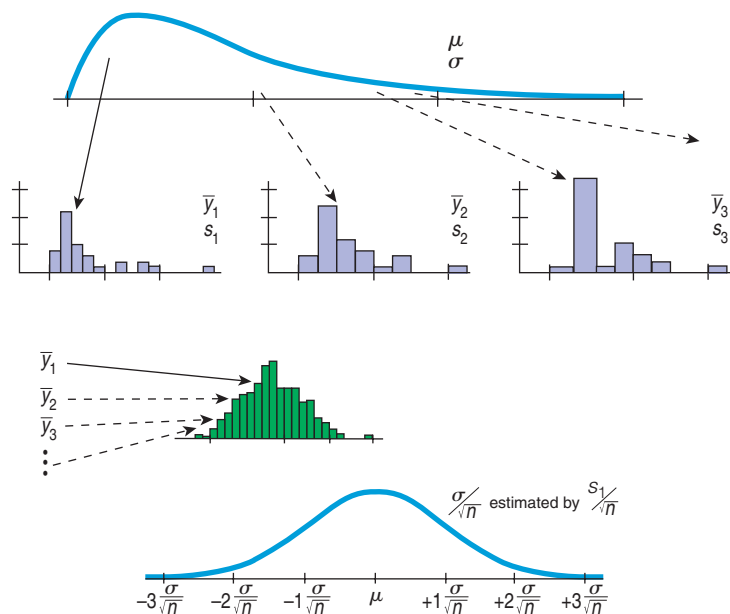
Figure 10.3 diagrams the process.

**Figure 10.3** We start with a population model, which can have any shape. It can even be bimodal or skewed (as this one is). We label the mean of this model  $\mu$  and its standard deviation,  $\sigma$ .

We draw one real sample (solid line) of size  $n$  and show its histogram and summary statistics. We *imagine* (or simulate) drawing many other samples (dotted lines), which have their own histograms and summary statistics.

We (imagine) gathering all the means into a histogram.

The CLT tells us we can model the shape of this histogram with a Normal model. The mean of this Normal is  $\mu$ , and the standard deviation is  $SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$ .



## What Can Go Wrong?

- **Don't confuse the sampling distribution with the distribution of the sample.** When you take a sample, you always look at the distribution of the values, usually with a histogram, and you may calculate summary statistics. Examining the distribution of the sample like this is wise. But that's not the sampling distribution. The sampling distribution is an imaginary collection of the values that a statistic might have taken for all the random samples—the one you got and the ones that you didn't get. Use the sampling distribution model to make statements about how the statistic varies.
- **Beware of observations that are not independent.** The CLT depends crucially on the assumption of independence. Unfortunately, this isn't something you can check in your data. You have to think about how the data were gathered. Good sampling practice and well-designed randomized experiments ensure independence.
- **Watch out for small samples from skewed populations.** The CLT assures us that the sampling distribution model is Normal if  $n$  is large enough. If the population is nearly Normal, even small samples may work. If the population is very skewed, then  $n$  will have to be large before the Normal model will work well. If we sampled 15 or even 20 CEOs and used  $\bar{y}$  to make a statement about the mean of all CEOs' compensation, we'd likely get into trouble because the underlying data distribution is so skewed. Unfortunately, there's no good rule to handle this.<sup>8</sup> It just depends on how skewed the data distribution is. Always plot the data to check.

## Ethics in Action

**H**ome Illusions, a national retailer of contemporary furniture and home décor has recently experienced customer complaints about the delivery of its products. This retailer uses different carriers depending on the order destination. Its policy with regard to most items it sells and ships is to simply deliver to the customer's doorstep. However, its policy with regard to furniture is to "deliver, unpack, and place furniture in the intended area of the home." Most of their recent complaints have been from customers in the northeastern region of the United States who were dissatisfied because their furniture deliveries were not unpacked and placed in their homes. Since the retailer uses different carriers, it is important for them to label their packages correctly so the delivery company can distinguish between furniture and nonfurniture deliveries. Home Illusions sets as a target "1% or less" for incorrect labeling of packages. Joe Zangard, V.P. Logistics, was asked to look into the problem. The retailer's largest warehouse in the northeast prepares about 1000 items per week for shipping. Joe's initial attention was directed at this facility, not only because of its large volume, but

also because he had some reservations about the newly hired warehouse manager, Brent Mossir. Packages at the warehouse were randomly selected and examined over a period of several weeks. Out of 1000 packages, 13 were labeled incorrectly. Since Joe had expected the count to be 10 or fewer, he was confident that he had now pinpointed the problem. His next step was to set up a meeting with Brent in order to discuss the ways in which he can improve the labeling process at his warehouse.

**ETHICAL ISSUE** *Joe is treating the sample proportion as if it were the true fixed value. By not recognizing that this sample proportion varies from sample to sample, he has unfairly judged the labeling process at Brent's warehouse. This is consistent with his initial misgivings about Brent being hired as warehouse manager (related to Item A, ASA Ethical Guidelines).*

**ETHICAL SOLUTION** *Joe Zangard needs to use the normal distribution to model the sampling distribution for the sample proportion. In this way, he would realize that the sample proportion observed is less than one standard deviation away from 1% (the upper limit of the target) and thus not conclusively larger than the limit.*

<sup>8</sup>For proportions, there is a rule: the Success/Failure condition. That works for proportions because the standard deviation of a proportion is linked to its mean. You may hear that 30 or 50 observations is enough to guarantee Normality, but it really depends on the skewness of the original data distribution.



## What Have We Learned?

### Learning Objectives

- Model the variation in statistics from sample to sample with a sampling distribution.
  - The Central Limit Theorem tells us that the sampling distribution of both the sample proportion and the sample mean are Normal.
- Understand that, usually, the mean of a sampling distribution is the value of the parameter estimated.
  - For the sampling distribution of  $\hat{p}$ , the mean is  $p$ .
  - For the sampling distribution of  $\bar{y}$  the mean is  $\mu$ .
- Interpret the standard deviation of a sampling distribution.
  - The standard deviation of a sampling model is the most important information about it.
  - The standard deviation of the sampling distribution of a proportion is  $\sqrt{\frac{pq}{n}}$ , where  $q = 1 - p$ .
  - The standard deviation of the sampling distribution of a mean is  $\frac{\sigma}{\sqrt{n}}$ , where  $\sigma$  is the population standard deviation.
- Understand that the Central Limit Theorem is a *limit* theorem.
  - The sampling distribution of the mean is Normal, *no matter what the underlying distribution of the data is*.
  - The CLT says that this happens in the limit, as the sample size grows. The Normal model applies sooner when sampling from a unimodal, symmetric population and more gradually when the population is very non-Normal.

### Terms

Central Limit Theorem	The Central Limit Theorem (CLT) states that the sampling distribution model of the sample mean (and proportion) is approximately Normal for large $n$ , regardless of the distribution of the population, as long as the observations are independent.
Sampling distribution	The distribution of a statistic over many independent samples of the same size from the same population.
Sampling distribution model for a mean	If the independence assumption and randomization condition are met and the sample size is large enough, the sampling distribution of the sample mean is well modeled by a Normal model with a mean equal to the population mean, $\mu$ , and a standard deviation equal to $\frac{\sigma}{\sqrt{n}}$ .
Sampling distribution model for a proportion	If the independence assumption and randomization condition are met and we expect at least 10 successes and 10 failures, then the sampling distribution of a proportion is well modeled by a Normal model with a mean equal to the true proportion value, $p$ , and a standard deviation equal to $\sqrt{\frac{pq}{n}}$ .
Sampling error	The variability we expect to see from sample to sample is often called the sampling error, although sampling variability is a better term.
Standard error	When the standard deviation of the sampling distribution of a statistic is estimated from the data, the resulting statistic is called a standard error (SE).

## Brief CASE



### Real Estate Simulation

Many variables important to the real estate market are skewed, limited to only a few values or considered as categorical variables. Yet, marketing and business decisions are often made based on means and proportions calculated over many homes. One reason these statistics are useful is the Central Limit Theorem.

Data on 1063 houses sold recently in the Saratoga, New York area, are in the file **Saratoga\_Real\_Estate** on your disk. Let's investigate how the CLT guarantees that the sampling distribution of proportions approaches the Normal and that the same is true for means of a quantitative variable even when samples are drawn from populations that are far from Normal.

#### Part 1: Proportions

The variable *Fireplace* is a dichotomous variable where 1 = *has a fireplace* and 0 = *does not have a fireplace*.

- Calculate the proportion of homes that have fireplaces for all 1063 homes. Using this value, calculate what the standard error of the sample proportion would be for a sample of size 50.
- Using the software of your choice, draw 100 samples of size 50 from this population of homes, find the proportion of homes with fireplaces in each of these samples, and make a histogram of these proportions.
- Compare the mean and standard deviation of this (sampling) distribution to what you previously calculated.

#### Part 2: Means

- Select one of the quantitative variables and make a histogram of the entire population of 1063 homes. Describe the distribution (including its mean and SD).
- Using the software of your choice, draw 100 samples of size 50 from this population of homes, find the means of these samples, and make a histogram of these means.
- Compare the (sampling) distribution of the means to the distribution of the population.
- Repeat the exercise with samples of sizes 10 and of 30. What do you notice about the effect of the sample size?

Some statistics packages make it easier than others to draw many samples and find means. Your instructor can provide advice on the path to follow for your package.

An alternative approach is to have each member of the class draw one sample to find the proportion and mean and then combine the statistics for the entire class.

## Exercises

### SECTION 10.1

1. An investment website can tell what devices are used to access the site. The site managers wonder whether they should enhance the facilities for trading via “smart phones” so they want to estimate the proportion of users who access the site that way (even if they also use their computers sometimes). They draw a random sample of 200 investors from their customers. Suppose that the true proportion of smart phone users is 36%.

- What would you expect the shape of the sampling distribution for the sample proportion to be?
- What would be the mean of this sampling distribution?
- If the sample size were increased to 500, would your answers change? Explain.

2. The proportion of adult women in the United States is approximately 51%. A marketing survey telephones 400 people at random.

- What proportion of women in the sample of 400 would you expect to see?
- How many women, on average, would you expect to find in a sample of that size? (*Hint:* Multiply the expected proportion by the sample size.)

### SECTION 10.2

3. The investment website of Exercise 1 draws a random sample of 200 investors from their customers. Suppose that the true proportion of smart phone users is 36%.

- What would the standard deviation of the sampling distribution of the proportion of smart phone users be?
- What is the probability that the sample proportion of smart phone users is greater than 0.36?
- What is the probability that the sample proportion is between 0.30 and 0.40?
- What is the probability that the sample proportion is less than 0.28?
- What is the probability that the sample proportion is greater than 0.42?

4. The proportion of adult women in the United States is approximately 51%. A marketing survey telephones 400 people at random.

- What is the sampling distribution of the observed proportion that are women?
- What is the standard deviation of that proportion?
- Would you be surprised to find 53% women in a sample of size 400? Explain.
- Would you be surprised to find 41% women in a sample of size 400? Explain.
- Would you be surprised to find that there were fewer than 160 women in the sample? Explain.

5. A real estate agent wants to know how many owners of homes worth over \$1,000,000 might be considering putting their home on the market in the next 12 months. He surveys 40 of them and finds that 10 of them are considering such a move. Are all the assumptions and conditions for finding the sampling distribution of the proportion satisfied? Explain briefly.

6. A tourist agency wants to know what proportion of visitors to the Eiffel Tower are from the Far East. To find out they survey 100 people in the line to purchase tickets to the top of the tower one Sunday afternoon in May. Are all the assumptions and conditions for finding the sampling distribution of the proportion satisfied? Explain briefly.

### SECTION 10.3

7. A sample of 40 games sold for the iPad has prices that have a distribution that is skewed to the high end with a mean of \$3.48 and a standard deviation of \$2.23. Teens who own iPads typically own about 20 games. Using the 68–95–99.7 Rule, draw and label an appropriate sampling model for the average amount a teen would spend per game if they had 20 games, assuming those games were a representative sample of available games.

8. Statistics for the closing price of the USAA Aggressive Growth Fund for the year 2009 indicate that the average closing price was \$23.90, with a standard deviation of \$3.00. Using the 68–95–99.7 Rule, draw and label an appropriate sampling model for the mean closing price of 36 days’ closing prices selected at random. What (if anything) do you need to assume about the distribution of prices? Are those assumptions reasonable?

### SECTION 10.4

9. According to the Gallup poll, 27% of U.S. adults have high levels of cholesterol. They report that such elevated levels “could be financially devastating to the U.S. health-care system” and are a major concern to health insurance providers. According to recent studies, cholesterol levels in healthy U.S. adults average about 215 mg/dL with a standard deviation of about 30 mg/dL and are roughly Normally distributed. If the cholesterol levels of a sample of 42 healthy U.S. adults is taken,

- What shape should the sampling distribution of the mean have?
- What would the mean of the sampling distribution be?
- What would its standard deviation be?
- If the sample size were increased to 100, how would your answers to parts a–c change?

10. As in Exercise 9, cholesterol levels in healthy U.S. adults average about 215 mg/dL with a standard deviation of

about 30 mg/dL and are roughly Normally distributed. If the cholesterol levels of a sample of 42 healthy US adults is taken, what is the probability that the mean cholesterol level of the sample

- a) Will be no more than 215?
- b) Will be between 205 and 225?
- c) Will be less than 200?
- d) Will be greater than 220?

**SECTION 10.5**

**11.** A marketing researcher for a phone company surveys 100 people and finds that that proportion of clients who are likely to switch providers when their contract expires is 0.15.

- a) What is the standard deviation of the sampling distribution of the proportion?
- b) If she wants to reduce the standard deviation by half, how large a sample would she need?

**12.** A market researcher for a provider of iPod accessories wants to know the proportion of customers who own cars to assess the market for a new iPod car charger. A survey of 500 customers indicates that 76% own cars.

- a) What is the standard deviation of the sampling distribution of the proportion?
- b) How large would the standard deviation have been if he had surveyed only 125 customers (assuming the proportion is about the same)?

**13.** Organizers of a fishing tournament believe that the lake holds a sizable population of largemouth bass. They assume that the weights of these fish have a model that is skewed to the right with a mean of 3.5 pounds and a standard deviation of 2.32 pounds.

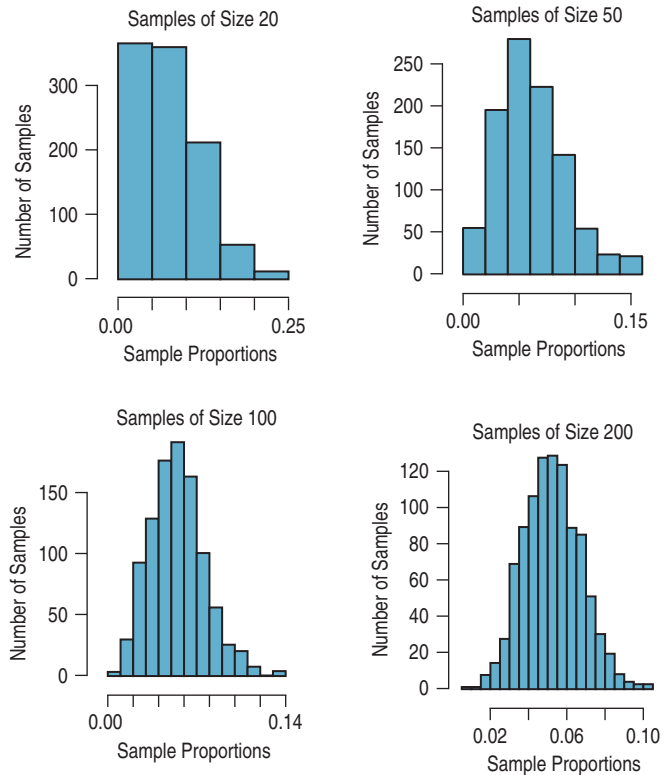
- a) Explain why a skewed model makes sense here.
- b) Explain why you cannot determine the probability that a largemouth bass randomly selected (“caught”) from the lake weighs over 3 pounds.
- c) Each contestant catches 5 fish each day. Can you determine the probability that someone’s catch averages over 3 pounds? Explain.
- d) The 12 contestants competing each caught the limit of 5 fish. What’s the standard deviation of the mean weight of the 60 fish caught?
- e) Would you be surprised if the mean weight of the 60 fish caught in the competition was more than 4.5 pounds? Use the 68–95–99.7 Rule.

**14.** In 2008 and 2009, Systemax bought two failing electronics stores, Circuit City and CompUSA. They have kept both the names active and customers can purchase products from either website. If they take a random sample of a mixture of recent purchases from the two websites, the distribution of the amounts purchased will be bimodal.

- a) As their sample size increases, what’s the expected shape of the distribution of amounts purchased in the sample?
- b) As the sample size increases, what’s the expected shape of the sampling model for the mean amount purchased of the sample?

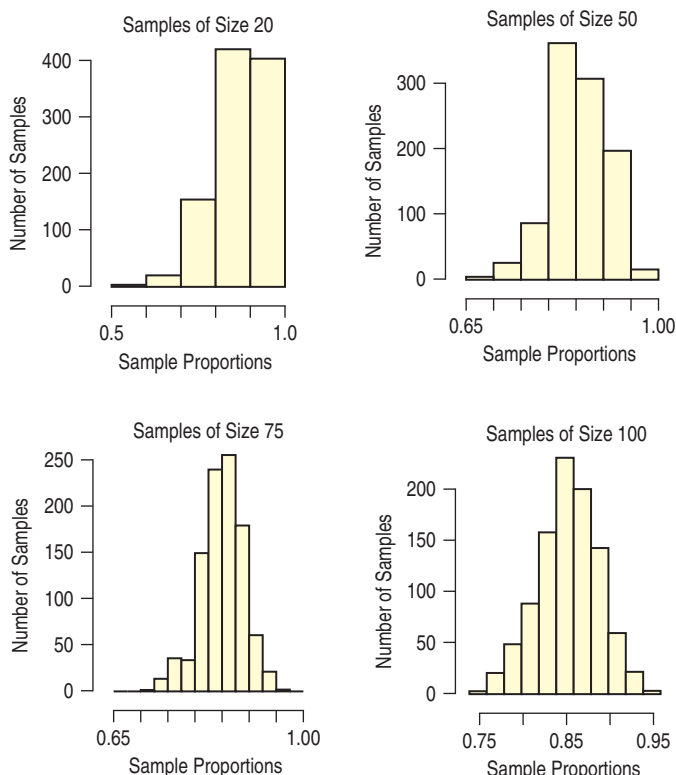
**CHAPTER EXERCISES**

**15. Send money.** When they send out their fundraising letter, a philanthropic organization typically gets a return from about 5% of the people on their mailing list. To see what the response rate might be for future appeals, they did a simulation using samples of size 20, 50, 100, and 200. For each sample size, they simulated 1000 mailings with success rate  $p = 0.05$  and constructed the histogram of the 1000 sample proportions, shown below. Explain how these histograms demonstrate what the Central Limit Theorem says about the sampling distribution model for sample proportions. Be sure to talk about shape, center, and spread.



**16. Character recognition.** An automatic character recognition device can successfully read about 85% of handwritten credit card applications. To estimate what might happen when this device reads a stack of applications, the company did a simulation using samples of size 20, 50, 75, and 100. For each sample size, they simulated 1000 samples with success rate  $p = 0.85$  and constructed the histogram of the 1000 sample proportions, shown here. Explain how these histograms demonstrate what the Central Limit Theorem

says about the sampling distribution model for sample proportions. Be sure to talk about shape, center, and spread.



**17. Send money, again.** The philanthropic organization in Exercise 15 expects about a 5% success rate when they send fundraising letters to the people on their mailing list. In Exercise 15 you looked at the histograms showing distributions of sample proportions from 1000 simulated mailings for samples of size 20, 50, 100, and 200. The sample statistics from each simulation were as follows:

$n$	mean	st. dev.
20	0.0497	0.0479
50	0.0516	0.0309
100	0.0497	0.0215
200	0.0501	0.0152

- According to the Central Limit Theorem, what should the theoretical mean and standard deviations be for these sample sizes?
- How close are those theoretical values to what was observed in these simulations?
- Looking at the histograms in Exercise 15, at what sample size would you be comfortable using the Normal model as an approximation for the sampling distribution?
- What does the Success/Failure Condition say about the choice you made in part c?

**18. Character recognition, again.** The automatic character recognition device discussed in Exercise 16 successfully reads about 85% of handwritten credit card applications. In Exercise 16 you looked at the histograms showing distributions of sample proportions from 1000 simulated samples of size 20, 50, 75, and 100. The sample statistics from each simulation were as follows:

$n$	mean	st. dev.
20	0.8481	0.0803
50	0.8507	0.0509
75	0.8481	0.0406
100	0.8488	0.0354

- According to the Central Limit Theorem, what should the theoretical mean and standard deviations be for these sample sizes?
- How close are those theoretical values to what was observed in these simulations?
- Looking at the histograms in Exercise 16, at what sample size would you be comfortable using the Normal model as an approximation for the sampling distribution?
- What does the Success/Failure Condition say about the choice you made in part c?

**19. Stock picking.** In a large Business Statistics class, the professor has each person select stocks by throwing 16 darts at pages of the *Wall Street Journal*. They then check to see whether their stock picks rose or fell the next day and report their proportion of “successes.” As a lesson, the professor has selected pages of the *Journal* for which exactly half the publicly traded stocks went up and half went down. The professor then makes a histogram of the reported proportions.

- What shape would you expect this histogram to be? Why?
- Where do you expect the histogram to be centered?
- How much variability would you expect among these proportions?
- Explain why a Normal model should not be used here.

**20. Quality management.** Manufacturing companies strive to maintain production consistency, but it is often difficult for outsiders to tell whether they have succeeded. Sometimes, however, we can find a simple example. The candy company that makes M&M’s candies claims that 10% of the candies it produces are green and that bags are packed randomly. We can check on their production controls by sampling bags of candies. Suppose we open bags containing about 50 M&M’s and record the proportion of green candies.

- If we plot a histogram showing the proportions of green candies in the various bags, what shape would you expect it to have?
- Can that histogram be approximated by a Normal model? Explain.

- c) Where should the center of the histogram be?  
 d) What should the standard deviation of the proportion be?

**21. Bigger portfolio.** The class in Exercise 19 expands its stock-picking experiment.

- a) The students use computer-generated random numbers to choose 25 stocks each. Use the 68–95–99.7 Rule to describe the sampling distribution model.  
 b) Confirm that you can use a Normal model here.  
 c) They increase the number of stocks picked to 64 each. Draw and label the appropriate sampling distribution model. Check the appropriate conditions to justify your model.  
 d) Explain how the sampling distribution model changes as the number of stocks picked increases.

**22. More quality.** Would a bigger sample help us to assess manufacturing consistency? Suppose instead of the 50-candy bags of Exercise 20, we work with bags that contain 200 M&M's each. Again we calculate the proportion of green candies found.

- a) Explain why it's appropriate to use a Normal model to describe the distribution of the proportion of green M&M's they might expect.  
 b) Use the 68–95–99.7 Rule to describe how this proportion might vary from bag to bag.  
 c) How would this model change if the bags contained even more candies?

**23. A winning investment strategy?** One student in the class of Exercise 19 claims to have found a winning strategy. He watches a cable news show about investing and *during the show* throws his darts at the pages of the *Journal*. He claims that of 200 stocks picked in this manner, 58% were winners.

- a) What do you think of his claim? Explain.  
 b) If there are 100 students in the class, are you surprised that one was this successful? Explain.

**24. Even more quality.** In a really large bag of M&M's, we found 12% of 500 candies were green. Is this evidence that the manufacturing process is out of control and has made too many greens? Explain.

**25. Speeding.** State police believe that 70% of the drivers traveling on a major interstate highway exceed the speed limit. They plan to set up a radar trap and check the speeds of 80 cars.

- a) Using the 68–95–99.7 Rule, draw and label the distribution of the proportion of these cars the police will observe speeding.  
 b) Do you think the appropriate conditions necessary for your analysis are met? Explain.

**26. Smoking, 2008.** Public health statistics indicate that 20.6% of American adults smoke cigarettes. Using the 68–95–99.7 Rule, describe the sampling distribution model

for the proportion of smokers among a randomly selected group of 50 adults. Be sure to discuss your assumptions and conditions.

**27. Vision.** It is generally believed that nearsightedness affects about 12% of all children. A school district has registered 170 incoming kindergarten children.

- a) Can you apply the Central Limit Theorem to describe the sampling distribution model for the sample proportion of children who are nearsighted? Check the conditions and discuss any assumptions you need to make.  
 b) Sketch and clearly label the sampling model, based on the 68–95–99.7 Rule.  
 c) How many of the incoming students might the school expect to be nearsighted? Explain.

**28. Mortgages.** In early 2007 the Mortgage Lenders Association reported that homeowners, hit hard by rising interest rates on adjustable-rate mortgages, were defaulting in record numbers. The foreclosure rate of 1.6% meant that millions of families were in jeopardy of losing their homes. Suppose a large bank holds 1731 adjustable-rate mortgages.

- a) Can you use the Normal model to describe the sampling distribution model for the sample proportion of foreclosures? Check the conditions and discuss any assumptions you need to make.  
 b) Sketch and clearly label the sampling model, based on the 68–95–99.7 Rule.  
 c) How many of these homeowners might the bank expect will default on their mortgages? Explain.

**29. Loans.** Based on past experience, a bank believes that 7% of the people who receive loans will not make payments on time. The bank has recently approved 200 loans.

- a) What are the mean and standard deviation of the proportion of clients in this group who may not make timely payments?  
 b) What assumptions underlie your model? Are the conditions met? Explain.  
 c) What's the probability that over 10% of these clients will not make timely payments?

**30. Contacts.** The campus representative for Lens.com wants to know what percentage of students at a university currently wear contact lens. Suppose the true proportion is 30%.

- a) We randomly pick 100 students. Let  $\hat{p}$  represent the proportion of students in this sample who wear contacts. What's the appropriate model for the distribution of  $\hat{p}$ ? Specify the name of the distribution, the mean, and the standard deviation. Be sure to verify that the conditions are met.  
 b) What's the approximate probability that more than one third of this sample wear contacts?

**31. Back to school?** Best known for its testing program, ACT, Inc., also compiles data on a variety of issues in

education. In 2004 the company reported that the national college freshman-to-sophomore retention rate held steady at 74% over the previous four years. Consider colleges with freshman classes of 400 students. Use the 68–95–99.7 Rule to describe the sampling distribution model for the percentage of those students we expect to return to that school for their sophomore years. Do you think the appropriate conditions are met?

**32. Binge drinking.** A national study found that 44% of college students engage in binge drinking (5 drinks at a sitting for men, 4 for women). Use the 68–95–99.7 Rule to describe the sampling distribution model for the proportion of students in a randomly selected group of 200 college students who engage in binge drinking. Do you think the appropriate conditions are met?

**33. Back to school, again.** Based on the 74% national retention rate described in Exercise 31, does a college where 522 of the 603 freshman returned the next year as sophomores have a right to brag that it has an unusually high retention rate? Explain.

**34. Binge sample.** After hearing of the national result that 44% of students engage in binge drinking (5 drinks at a sitting for men, 4 for women), a professor surveyed a random sample of 244 students at his college and found that 96 of them admitted to binge drinking in the past week. Should he be surprised at this result? Explain.

**35. Polling.** Just before a referendum on a school budget, a local newspaper polls 400 voters in an attempt to predict whether the budget will pass. Suppose that the budget actually has the support of 52% of the voters. What's the probability the newspaper's sample will lead them to predict defeat? Be sure to verify that the assumptions and conditions necessary for your analysis are met.

**36. Seeds.** Information on a packet of seeds claims that the germination rate is 92%. What's the probability that more than 95% of the 160 seeds in the packet will germinate? Be sure to discuss your assumptions and check the conditions that support your model.

**37. Apples.** When a truckload of apples arrives at a packing plant, a random sample of 150 is selected and examined for bruises, discoloration, and other defects. The whole truckload will be rejected if more than 5% of the sample is unsatisfactory. Suppose that in fact 8% of the apples on the truck do not meet the desired standard. What's the probability that the shipment will be accepted anyway?

**38. Genetic defect.** It's believed that 4% of children have a gene that may be linked to juvenile diabetes. Researchers hoping to track 20 of these children for several years test

732 newborns for the presence of this gene. What's the probability that they find enough subjects for their study?

**39. Nonsmokers.** While some nonsmokers do not mind being seated in a smoking section of a restaurant, about 60% of the customers demand a smoke-free area. A new restaurant with 120 seats is being planned. How many seats should be in the nonsmoking area in order to be very sure of having enough seating there? Comment on the assumptions and conditions that support your model, and explain what "very sure" means to you.

**40. Meals.** A restaurateur anticipates serving about 180 people on a Friday evening, and believes that about 20% of the patrons will order the chef's steak special. How many of those meals should he plan on serving in order to be pretty sure of having enough steaks on hand to meet customer demand? Justify your answer, including an explanation of what "pretty sure" means to you.

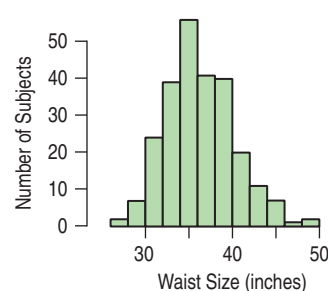
**41. Sampling.** A sample is chosen randomly from a population that can be described by a Normal model.

- What's the sampling distribution model for the sample mean? Describe shape, center, and spread.
- If we choose a larger sample, what's the effect on this sampling distribution model?

**42. Sampling, part II.** A sample is chosen randomly from a population that was strongly skewed to the left.

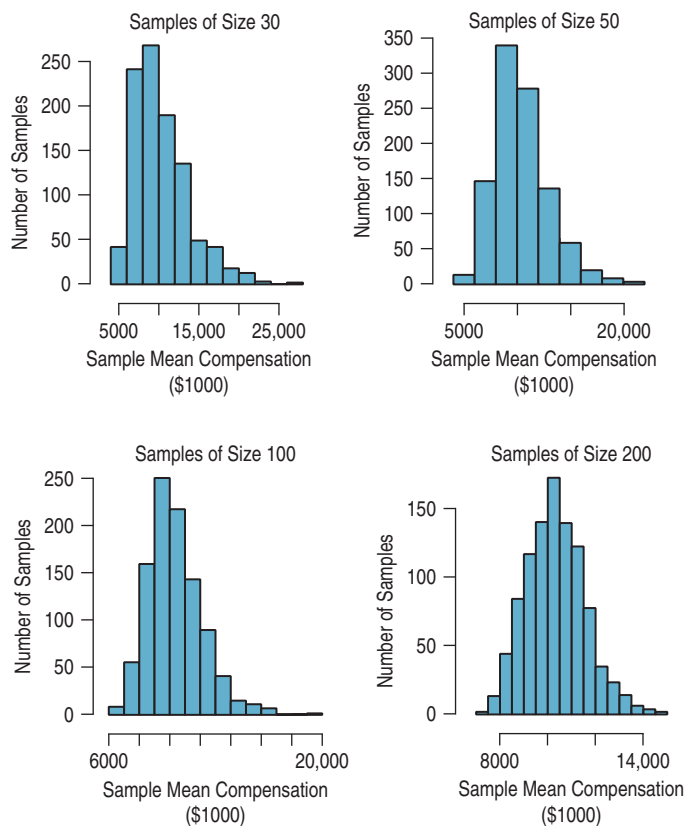
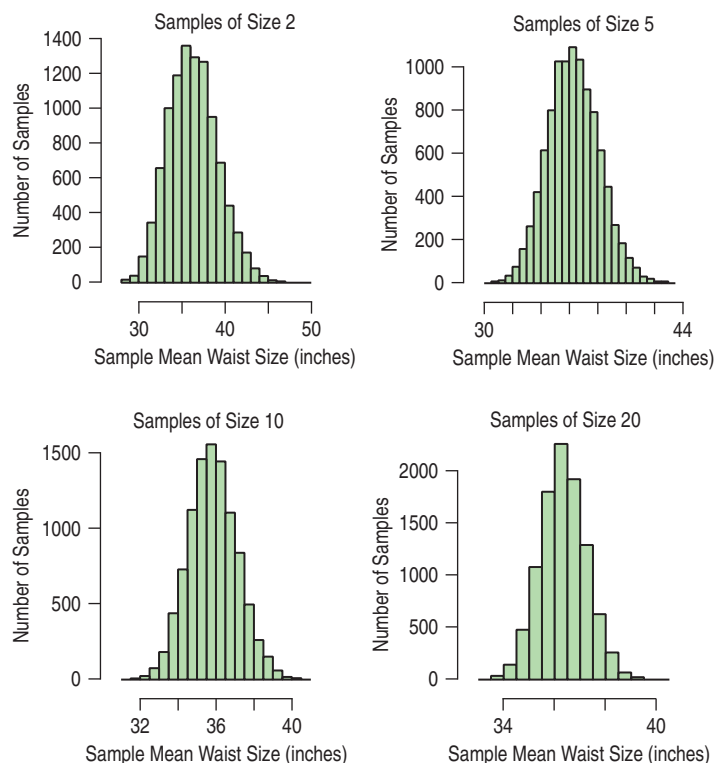
- Describe the sampling distribution model for the sample mean if the sample size is small.
- If we make the sample larger, what happens to the sampling distribution model's shape, center, and spread?
- As we make the sample larger, what happens to the expected distribution of the data in the sample?

**43. Waist size.** A study commissioned by a clothing manufacturer measured the *Waist Size* of 250 men, finding a mean of 36.33 inches and a standard deviation of 4.02 inches. Here is a histogram of these measurements:

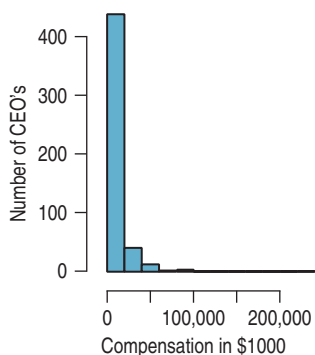


- Describe the histogram of *Waist Size*.
- To explore how the mean might vary from sample to sample, they simulated by drawing many samples of size 2, 5, 10, and 20, with replacement, from the 250 measurements.

Here are histograms of the sample means for each simulation. Explain how these histograms demonstrate what the Central Limit Theorem says about the sampling distribution model for sample means.



**44. CEO compensation.** The average total annual compensation for CEOs of the 800 largest U.S. companies (in \$1000) is 10,307.31 and the standard deviation is 17964.62. Here is a histogram of their annual compensations (in \$1000):



c) Comment on the “rule of thumb” that “With a sample size of at least 30, the sampling distribution of the mean is Normal.”

**45. Waist size revisited.** A study commissioned by a clothing manufacturer measured the *Waist Sizes* of a random sample of 250 men. The mean and standard deviation of the *Waist Sizes* for all 250 men are 36.33 in and 4.019 inches, respectively. In Exercise 43 you looked at the histograms of simulations that drew samples of sizes 2, 5, 10, and 20 (with replacement). The summary statistics for these simulations were as follows:

<i>n</i>	mean	st. dev.
2	36.314	2.855
5	36.314	1.805
10	36.341	1.276
20	36.339	0.895

- Describe the histogram of *Total Compensation*. A research organization simulated sample means by drawing samples of 30, 50, 100, and 200, with replacement, from the 800 CEOs. The histograms show the distributions of means for many samples of each size.
- Explain how these histograms demonstrate what the Central Limit Theorem says about the sampling distribution model for sample means. Be sure to talk about shape, center, and spread.

- According to the Central Limit Theorem, what should the theoretical mean and standard deviation be for each of these sample sizes?
- How close are the theoretical values to what was observed in the simulation?
- Looking at the histograms in Exercise 43, at what sample size would you be comfortable using the Normal model as an approximation for the sampling distribution?
- What about the shape of the distribution of *Waist Size* explains your choice of sample size in part c?



**46. CEOs revisited.** In Exercise 44 you looked at the annual compensation for 800 CEOs, for which the true mean and standard deviation were (in thousands of dollars) 10,307.31 and 17,964.62, respectively. A simulation drew samples of sizes 30, 50, 100, and 200 (with replacement) from the total annual compensations of the Fortune 800 CEOs. The summary statistics for these simulations were as follows:

<i>n</i>	mean	st. dev.
30	10,251.73	3359.64
50	10,343.93	2483.84
100	10,329.94	1779.18
200	10,340.37	1230.79

- According to the Central Limit Theorem, what should the theoretical mean and standard deviation be for each of these sample sizes?
- How close are the theoretical values to what was observed from the simulation?
- Looking at the histograms in Exercise 44, at what sample size would you be comfortable using the Normal model as an approximation for the sampling distribution?
- What about the shape of the distribution of *Total Compensation* explains your answer in part c?

**47. GPAs.** A college's data about the incoming freshmen indicates that the mean of their high school GPAs was 3.4, with a standard deviation of 0.35; the distribution was roughly mound-shaped and only slightly skewed. The students are randomly assigned to freshman writing seminars in groups of 25. What might the mean GPA of one of these seminar groups be? Describe the appropriate sampling distribution model—shape, center, and spread—with attention to assumptions and conditions. Make a sketch using the 68–95–99.7 Rule.

**48. Home values.** Assessment records indicate that the value of homes in a small city is skewed right, with a mean of \$140,000 and standard deviation of \$60,000. To check the accuracy of the assessment data, officials plan to conduct a detailed appraisal of 100 homes selected at random. Using the 68–95–99.7 Rule, draw and label an appropriate sampling model for the mean value of the homes selected.

**49. The trial of the pyx.** In 1150, it was recognized in England that coins should have a standard weight of precious metal as the basis for their value. A guinea, for example, was supposed to contain 128 grains of gold. (There are

360 grains in an ounce.) In the “trial of the pyx,” coins minted under contract to the crown were weighed and compared to standard coins (which were kept in a wooden box called the pyx). Coins were allowed to deviate by no more than 0.28 grains—roughly equivalent to specifying that the standard deviation should be no greater than 0.09 grains (although they didn't know what a standard deviation was in 1150). In fact, the trial was performed by weighing 100 coins at a time and requiring the *sum* to deviate by no more than  $100 \times 0.28 = 28$  or 28 grains—equivalent to the sum having a standard deviation of about 9 grains.

- In effect, the trial of the pyx required that the mean weight of the sample of 100 coins have a standard deviation of 0.09 grains. Explain what was wrong with performing the trial in this manner.
- What should the limit have been on the standard deviation of the mean?

*Note:* Because of this error, the crown was exposed to being cheated by private mints that could mint coins with greater variation and then, after their coins passed the trial, select out the heaviest ones and recast them at the proper weight, retaining the excess gold for themselves. The error persisted for over 600 years, until sampling distributions became better understood.

**50. Safe cities.** Allstate Insurance Company identified the 10 safest and 10 least-safe U.S. cities from among the 200 largest cities in the United States, based on the mean number of years drivers went between automobile accidents. The cities on both lists were all smaller than the 10 largest cities. Using facts about the sampling distribution model of the mean, explain why this is not surprising.

### Just Checking Answers

- A Normal model (approximately).
- At the actual proportion of all customers who like the new location.
- $SD(\hat{p}) = \sqrt{\frac{(0.5)(0.5)}{100}} = 0.05$
- $SE(\bar{y}) = 120/\sqrt{100} = 12$
- Decreases.
- The standard error decreases by  $1/\sqrt{2}$ .

*This page intentionally left blank*

## Investigating the Central Limit Theorem

In the data set you investigated for Case Study Part I, were four variables: Age, Own Home?, Time Between Gifts, and Largest Gift. (The square root of Largest Gift was also included.) The Central Limit Theorem (CLT) says that for distributions that are nearly Normal, the sampling distribution of the mean will be approximately Normal even for small samples. For moderately skewed distributions, one rule of thumb says that sample sizes of 30 are large enough for the CLT to be approximately true. After looking at the distribution of each quantitative variable, think about how large a sample you think you would need for the mean (or proportion) to have a sampling distribution that is approximately Normal. The rest of this case study will investigate how soon the Central Limit Theorem actually starts to work for a variety of distributions.

Before starting, suppose that the organization tells us that no valid donor is under 21 years old. Exclude any such cases from the rest of the analysis (exclude the entire row for all donors under 21 years old).

For each quantitative variable, investigate the relationship between skewness and sample size by following the steps below.

- a) Describe the distribution of each quantitative variable for all valid cases, paying special attention to the shape.
- b) What is the mean of this variable? From that, what is the standard error of the mean?
- c) How large do you think a sample would have to be for the sampling distribution of the mean to be approximately Normal?
- d) Shuffle the valid cases, randomizing the order of the rows. (Note: if you are unable to generate different random samples from each variable, the means of a simulation are available on the book's website. In that case, skip steps d and e, use those files, and proceed to step f.)
- e) Split 2500 of the valid cases into 100 samples of size 25, find the mean of each sample, and save those values.
- f) Plot the histogram of the 100 means. Describe the distribution. Does the CLT seem to be working?
- g) From the standard error in step b, according to the Normal distribution, what percent of the sample means should be farther from the mean Age than two standard errors?
- h) From the 100 means that you found, what percent are *actually* more than two standard errors from the mean? What does that tell you about the CLT for samples of size 25 for this variable?
- i) If the distribution is not sufficiently Normal for this sample size, increase the sample size to 50 and repeat steps e–h.
- j) If the distribution is still not Normal enough for  $n = 50$ , try  $n = 100$  or even  $n = 250$  and repeat steps e–h until the CLT seems to be in effect (or use the sample means from various sample sizes in the files on the book's website).

For the categorical variable Own Home?, a histogram is inappropriate. Instead, follow these steps:

- k) What proportion of the valid cases own their own home?
- l) How large do you think a sample would have to be for the sampling distribution of the proportion to be approximately Normal?
- m) Shuffle the valid cases, randomizing the order of the rows.
- n) Split 2500 of the valid cases into 100 samples of size 25, find the proportion that own their own homes in each sample, and save those values. Given your answer to step a, what should the standard error of the sample proportions be?

- o) Plot the histogram of the 100 proportions. Describe the distribution. Is the distribution approximately Normal?
- p) From the standard error in step b, if the distribution is Normal, what percent of the sample proportions should be farther from the proportion you found in step a than two standard errors?
- q) From the 100 proportions that you found, what percent are actually more than two standard errors from the true proportion? What does that tell you about the CLT for samples of size 25 for this variable?
- r) If the distribution is not sufficiently Normal for this sample size, increase the sample size to 50 and repeat steps e–h.
- s) If the distribution is still not sufficiently Normal for  $n = 50$ , try  $n = 100$  or even  $n = 250$  and repeat steps d–g until the CLT seems to be in effect.

Write a short report including a summary of what you found out about the relationship between the skewness of the distribution of a quantitative variable and the sample size needed for the sampling distribution of the mean to be approximately Normal. For the categorical variable, how large a sample seems to be needed for the CLT to work? For Largest Gift, what does this say about the wisdom of using the square root of Largest Gift instead of the original variable?

# Confidence Intervals for Proportions

C H A P T E R

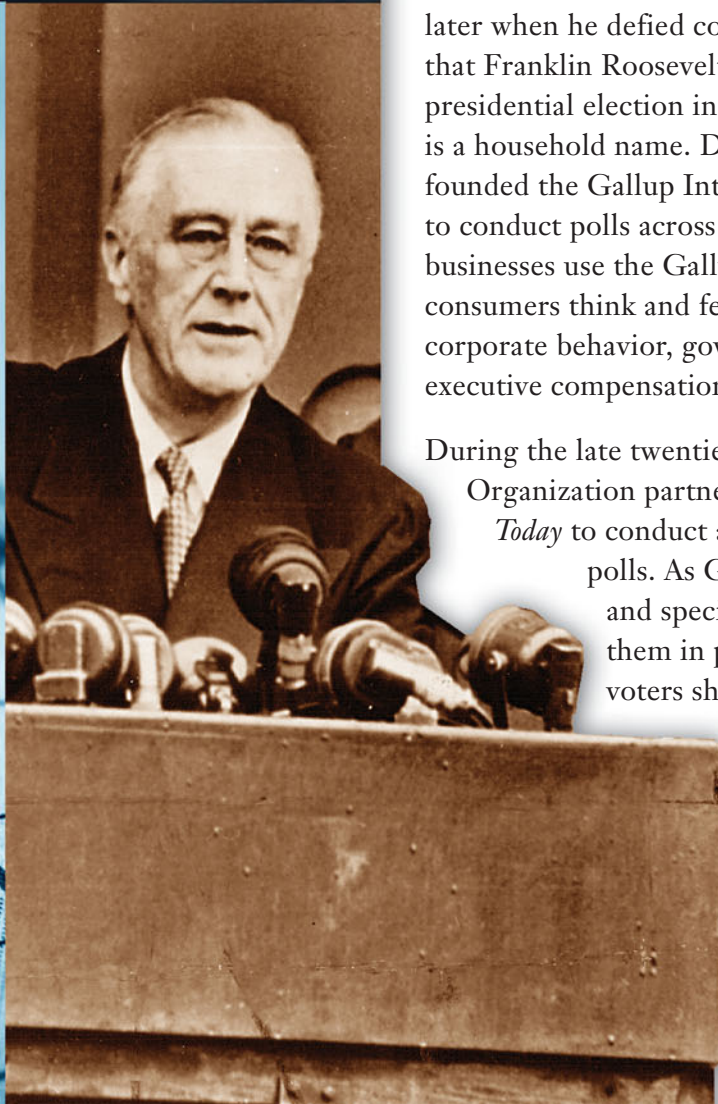
# 11



## The Gallup Organization

Dr. George Gallup was working as a market research director at an advertising agency in the 1930s when he founded the Gallup Organization to measure and track the public's attitudes toward political, social, and economic issues. He gained notoriety a few years later when he defied common wisdom and predicted that Franklin Roosevelt would win the U.S. presidential election in 1936. Today, the Gallup Poll is a household name. During the late 1930s, he founded the Gallup International Research Institute to conduct polls across the globe. International businesses use the Gallup polls to track how consumers think and feel about such issues as corporate behavior, government policies, and executive compensation.

During the late twentieth century, the Gallup Organization partnered with CNN and *USA Today* to conduct and publish public opinion polls. As Gallup once said, "If politicians and special interests have polls to guide them in pursuing their interests, the voters should have polls as well."<sup>1</sup>



<sup>1</sup>Source: The Gallup Organization, Princeton, NJ, [www.gallup.com](http://www.gallup.com).

Gallup's Web-based data storage system now holds data from polls taken over the last 65 years on a variety of topics, including consumer confidence, household savings, stock market investment, and unemployment.

**WHO** U.S. adults  
**WHAT** Proportion who think economy is getting better  
**WHEN** March 2010  
**WHY** To measure expectations about the economy

To plan their inventory and production needs, businesses use a variety of forecasts about the economy. One important attribute is consumer confidence in the overall economy. Tracking changes in consumer confidence over time can help businesses gauge whether the demand for their products is on an upswing or about to experience a downturn. The Gallup Poll periodically asks a random sample of U.S. adults whether they think economic conditions are getting better, getting worse, or staying about the same. When they polled 2976 respondents in March 2010, only 1012 thought economic conditions in the United States were getting better—a sample proportion of  $\hat{p} = 1012/2976 = 34.0\%$ .<sup>2</sup> We (and Gallup) hope that this observed proportion is close to the population proportion,  $p$ , but we know that a second sample of 2976 adults wouldn't have a sample proportion of exactly 34.0%. In fact, Gallup did sample another group of adults just a few days later and found a sample proportion of 38.0%.

From Chapter 10, we know it isn't surprising that two random samples give slightly different results. We'd like to say something, not about different random *samples*, but about the proportion of *all* adults who thought that economic conditions in the United States were getting better in March 2010. The sampling distribution will be the key to our ability to generalize from our sample to the population.

## 11.1 A Confidence Interval

### Notation Alert!

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

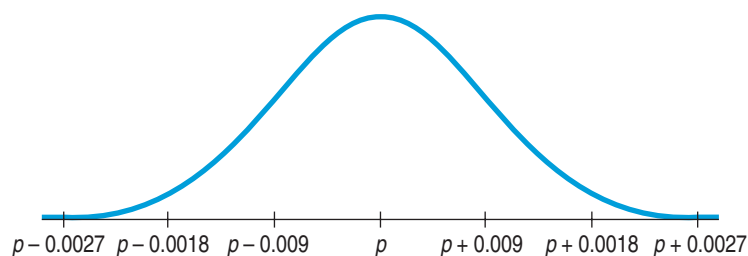
What do we know about our sampling distribution model? We know that it's centered at the true proportion,  $p$ , of all U.S. adults who think the economy is improving. But we don't know  $p$ . It isn't 34.0%. That's the  $\hat{p}$  from our sample. What we do know is that the sampling distribution model of  $\hat{p}$  is centered at  $p$ , and we know that the standard deviation of the sampling distribution is  $\sqrt{\frac{pq}{n}}$ . We also know, from the Central Limit Theorem, that the shape of the sampling distribution is approximately Normal, when the sample is large enough.

We don't know  $p$ , so we can't find the true standard deviation of the sampling distribution model. But we'll use  $\hat{p}$  and find the standard error:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.34)(1 - 0.34)}{2976}} = 0.009$$

Because the Gallup sample of 2976 is large, we know that the sampling distribution model for  $\hat{p}$  should look approximately like the one shown in Figure 11.1.

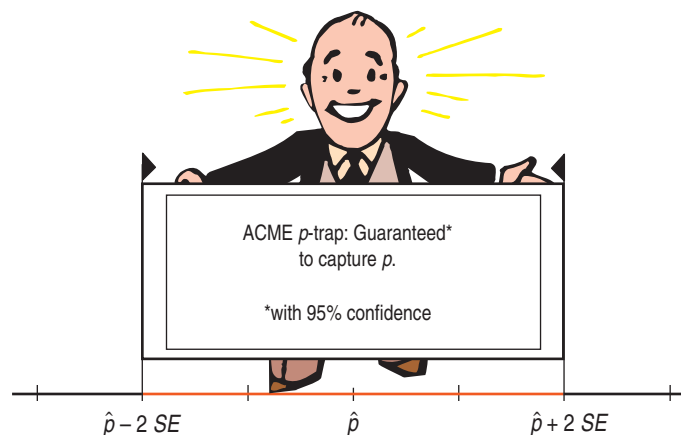
<sup>2</sup>A proportion is a *number* between 0 and 1. In business it's usually reported as a percentage. You may see it written either way.



**Figure 11.1** The sampling distribution of sample proportions is centered at the true proportion,  $p$ , with a standard deviation of 0.009.

The sampling distribution model for  $\hat{p}$  is Normal with a mean of  $p$  and a standard deviation we estimate to be  $\sqrt{\frac{\hat{p}\hat{q}}{n}}$ . Because the distribution is Normal, we'd expect that about 68% of all samples of 2976 U.S. adults taken in March 2010 would have had sample proportions within 1 standard deviation of  $p$ . And about 95% of all these samples will have proportions within  $p \pm 2$  SEs. 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 SEs away from  $p$ . So let's reverse it and look at it from  $\hat{p}$ 's point of view. If I'm  $\hat{p}$ , there's a 95% chance that  $p$  is no more than 2 SEs away from me. If I reach out 2 SEs, or  $2 \times 0.009$ , away from me on both sides, I'm 95% sure that  $p$  will be within my grasp. Of course, I won't know, and even if my interval does catch  $p$ , I still don't know its true value. The best I can do is state a probability that I've covered the true value in the interval.



**Figure 11.2** Reaching out 2 SEs on either side of  $\hat{p}$  makes us 95% confident we'll trap the true proportion,  $p$ .

## What Can We Say about a Proportion?

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

1. **"34.0% of all U.S. adults thought the economy was improving."** 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 34.0% of all U.S. adults thought the economy was improving."** No. In fact, we can be pretty sure that whatever the true proportion is, it's not exactly 34.0%, so the statement is not true.

3. **“We don’t know exactly what proportion of U.S. adults thought the economy was improving, but we know that it’s within the interval  $34.0\% \pm 2 \times 0.9\%$ . That is, it’s between 32.2% and 35.8%.”** 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 range.
4. **“We don’t know exactly what proportion of U.S. adults thought the economy was improving, but the interval from 32.2% to 35.8% *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.

*“Far better an approximate answer to the right question, . . . than an exact answer to the wrong question.”*

—JOHN W. TUKEY

That last statement is true, but it’s a bit wishy-washy. We can tighten it up by quantifying what we mean by “probably.” We saw that 95% of the time when we reach out 2 SEs 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 32.2% and 35.8% of U.S. adults thought the economy was improving.”** This is now an appropriate interpretation of our confidence intervals. It’s not perfect, but it’s about 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 an example of a one-proportion z-interval.**<sup>3</sup> We’ll lay out the formal definition in the next few pages.

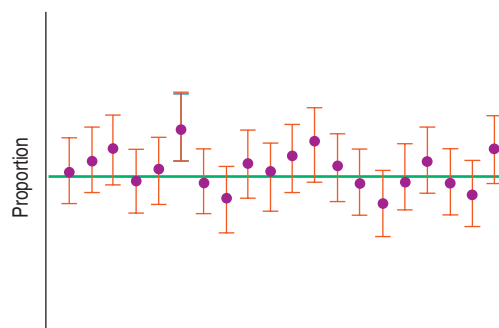
## What Does “95% Confidence” Really Mean?

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. In Chapter 10, we saw how proportions vary from sample to sample. If other pollsters had selected their own samples of adults, they would have found some who thought the economy was getting better, but each sample proportion would almost certainly differ from ours. When they each tried to estimate the true proportion, they’d center their confidence intervals at the proportions they observed in their own samples. Each would have ended up with a different interval.

Figure 11.3 shows the confidence intervals produced by simulating 20 samples. The purple dots are the simulated proportions of adults in each sample who thought the economy was improving, and the orange segments show the confidence intervals found for each simulated sample. The green line represents the true percentage of adults who thought the economy was improving. You can see that most of the simulated confidence intervals include the true value—but one missed. (Note that it is the *intervals* that vary from sample to sample; the green line doesn’t move.)

<sup>3</sup>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.”





**Figure 11.3** The horizontal green line shows the true proportion of people in March 2010 who thought the economy was improving. Most of the 20 simulated samples shown here produced 95% confidence intervals that captured the true value, but one missed.

Of course, a huge number of possible samples *could* be drawn, each with its own sample proportion. This simulation approximates just some of them. Each sample 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 U.S. adults who thought in March 2010 that the economy was improving. However, the Central Limit Theorem assures us that 95% of the intervals in the pile are winners, covering the true value, and only 5%, on average, miss the target. That’s why we’re 95% *confident* that our interval is a winner.

## For Example

### Finding a 95% confidence interval for a proportion

The Chamber of Commerce of a mid-sized city has supported a proposal to change the zoning laws for a new part of town. The new regulations would allow for mixed commercial and residential development. The vote on the measure is scheduled for three weeks from today, and the president of the Chamber of Commerce is concerned that they may not have the majority of votes that they will need to pass the measure. She commissions a survey that asks likely voters if they plan to vote for the measure. Of the 516 people selected at random from likely voters, 289 said they would likely vote for the measure.

#### Questions:

- Find a 95% confidence interval for the true proportion of voters who will vote for the measure. (Use the 68–95–99.7% Rule.)
- What would you report to the president of the Chamber of Commerce?

#### Answer:

$$\text{a. } \hat{p} = \frac{289}{516} = 0.56 \quad \text{So, } SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.56)(0.44)}{516}} = 0.022$$

A 95% confidence interval for  $p$  can be found from  $\hat{p} \pm 2 SE(\hat{p}) = 0.56 \pm 2(0.022) = (0.516, 0.604)$  or 51.6% to 60.4%.

- We are 95% confident that the true proportion of voters who plan to vote for the measure is between 51.6% and 60.4%. This assumes that the sample we have is representative of all likely voters.

## 11.2 Margin of Error: Certainty vs. Precision

We’ve just claimed that at a certain confidence level we’ve captured the true proportion of all U.S. adults who thought the economy was improving in March 2010. Our confidence interval stretched out the same distance on either side of the estimated proportion with the form:

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

**Confidence Intervals**

We'll see many confidence intervals in this book. All have the form:

$$\text{estimate} \pm \text{ME.}$$

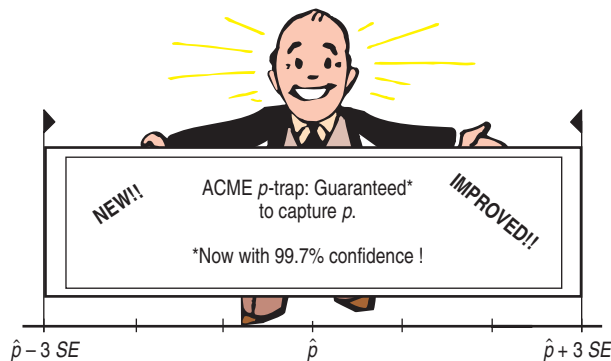
For proportions at 95% confidence:

$$\text{ME} \approx 2 SE(\hat{p}).$$

The *extent* of that interval on either side of  $\hat{p}$  is called the **margin of error (ME)**. In general, confidence intervals look like this:

$$\text{estimate} \pm \text{ME.}$$

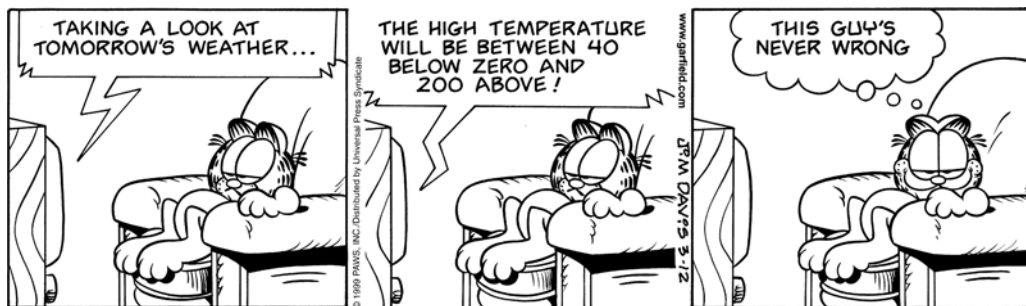
The margin of error for our 95% confidence interval was 2 SEs. What if we wanted to be more confident? To be more confident, we'd need to capture  $p$  more often, and to do that, we'd need to make the interval wider. For example, if we want to be 99.7% confident, the margin of error will have to be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



**Figure 11.4** Reaching out 3 SEs on either side of  $\hat{p}$  makes us 99.7% confident we'll trap the true proportion  $p$ . Compare the width of this interval with the interval in Figure 11.2.

The more confident we want to be, the larger the margin of error must be. We can be 100% confident that any proportion is between 0% and 100%, but that's not very useful. Or we could give a narrow confidence interval, say, from 33.98% to 34.02%. But we couldn't be very confident about a statement this precise. Every confidence interval is a balance between certainty and precision.

The tension between certainty and precision is always there. There is no simple answer to the conflict. Fortunately, in most cases we can be both sufficiently certain and sufficiently precise to make useful statements. The choice of confidence level is somewhat arbitrary, but you must choose the level yourself. The data can't do it for you. 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% might be viewed with suspicion.)



Garfield © 1999 Paws, Inc. Reprinted with permission of UNIVERSAL PRESS SYNDICATE. All rights reserved.

## Critical Values

In our opening example, our margin of error was 2 SEs, which produced a 95% confidence interval. To change the confidence level, we'll need to change the *number* of SEs to correspond to the new level. A wider confidence interval means more confidence. For any confidence level the number of SEs we must stretch out on either side of  $\hat{p}$  is called the **critical value**. Because it is based on the Normal

**Notation Alert!**

We put an asterisk on a letter to indicate a critical value. We usually use “ $z$ ” when we talk about Normal models, so  $z^*$  is always a critical value from a Normal model.

Some common confidence levels and their associated critical values:

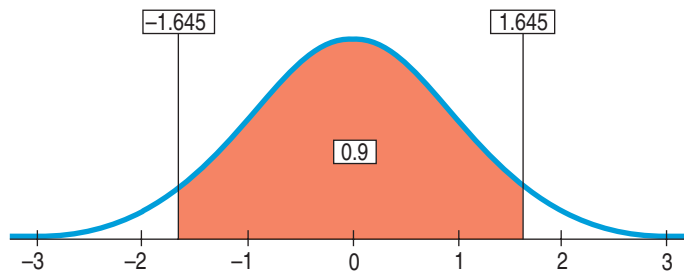
CI	$z^*$
90%	1.645
95%	1.960
99%	2.576

model, 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 in the back of the book.

For a 95% confidence interval, the precise critical value is  $z^* = 1.96$ . That is, 95% of a Normal model is found within  $\pm 1.96$  standard deviations of the mean. We’ve been using  $z^* = 2$  from the 68–95–99.7 Rule because 2 is very close to 1.96 and is easier to remember. Usually, the difference is negligible, but if you want to be precise, use 1.96.<sup>4</sup>

Suppose we could be satisfied with 90% confidence. What critical value would we need? We can use a smaller margin of error. Our greater precision is offset by our acceptance of being wrong more often (that is, having a confidence interval that misses the true value). Specifically, for a 90% confidence interval, the critical value is only 1.645 because for a Normal model, 90% of the values are within 1.645 standard deviations from the mean. By contrast, suppose your boss demands more confidence. If she wants an interval in which she can have 99% confidence, she’ll need to include values within 2.576 standard deviations, creating a wider confidence interval.

**Figure 11.5** For a 90% confidence interval, the critical value is 1.645 because for a Normal model, 90% of the values fall within 1.645 standard deviations of the mean.

**For Example****Finding confidence intervals for proportions with different levels of confidence**

The president of the Chamber of Commerce is worried that 95% confidence is too low and wants a 99% confidence interval.

**Question:** Find a 99% confidence interval. Would you reassure her that the measure will pass? Explain.

**Answer:** In the example on page 309, we used 2 as the value of  $z^*$  for 95% confidence. A more precise value would be 1.96 for 95% confidence. For 99% confidence, the critical  $z$ -value is 2.756. So, a 99% confidence interval for the true proportion is

$$\hat{p} \pm 2.576 SE(\hat{p}) = 0.56 \pm 2.576(0.022) = (0.503, 0.617)$$

The confidence interval is now wider: 50.3% to 61.7%.

The Chamber of Commerce needs at least 50% for the vote to pass. At a 99% confidence level, it looks now as if the measure will pass. However, we must assume that the sample is representative of the voters in the actual election and that people vote in the election as they said they will when they took the survey.

## 11.3 Assumptions and Conditions

The statements we made about what all U.S. adults thought about the economy 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. If those assumptions are not true, the model might be inappropriate, and our conclusions based on it may be

<sup>4</sup>It’s been suggested that since 1.96 is both an unusual value and so important in Statistics, you can recognize someone who’s had a Statistics course by just saying “1.96” and seeing whether they react.

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 10. But, because they are so important, we'll go over them again.

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

Here are the assumptions and the corresponding conditions to check before creating (or believing) a confidence interval about a proportion.

## Independence Assumption

You first need to think about whether the independence assumption is plausible. You can look for reasons to suspect that it fails. You might wonder whether there is any reason to believe that the data values somehow affect each other. (For example, might any of the adults in the sample be related?) This condition depends on your knowledge of the situation. It's not one you can check by looking at the data. However, now that you have data, there are two conditions that you 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, you'd like to have as large a sample as you can. But if you sample from a small population, the probability of success may be different for the last few individuals you 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 a Normal model may no longer be appropriate. But if less than 10% of the population is sampled, it is safe to assume to have independence.

## Sample Size Assumption

The model we use for inference is based on the Central Limit Theorem. So, the sample must be large enough for the Normal sampling model to be appropriate. It turns out that we need more data when the proportion is close to either extreme (0 or 1). This requirement is easy to check with the following condition:

- **Success/Failure Condition:** We must expect our sample to contain 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. The other alternative is then a "failure." So we check that both  $n\hat{p} \geq 10$  and  $n\hat{q} \geq 10$ .

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

## For Example

## Assumptions and conditions for a confidence interval for proportions

We previously reported a confidence interval to the president of the Chamber of Commerce.

**Question:** Were the assumptions and conditions for making this interval satisfied?

**Answer:** Because the sample was randomized, we assume that the responses of the people surveyed were independent so the randomization condition is met. We assume that 516 people represent fewer than 10% of the likely voters in the town so the 10% condition is met. Because 289 people said they were likely to vote for the measure and thus 227 said they were not, both are much larger than 10 so the Success/Failure condition is also met.

All the conditions to make a confidence interval for the proportion appear to have been satisfied.

## Guided Example Public Opinion



In the film *Up in the Air*, George Clooney portrays a man whose job it is to tell workers that they have been fired. The reactions to such news took a

somewhat odd turn in France in the spring of 2009 when workers at Sony France took the boss hostage for a night and barricaded their factory entrance with a tree trunk. He was freed only after he agreed to reopen talks on their severance packages. Similar incidents occurred at 3M and Caterpillar plants in France. A poll taken by *Le Parisien* in April 2009 found 45% of the French “supportive” of such action. A similar poll taken by *Paris Match*, April 2–3, 2009, found 30% “approving” and 63% were “understanding” or “sympathetic” of the action. Only 7% condemned the practice of “bossnapping.”

The *Paris Match* poll was based on a random representative sample of 1010 adults. What can we conclude about the proportion of all French adults who sympathize with (without supporting outright) the practice of bossnapping?

To answer this question, we’ll build a confidence interval for the proportion of all French adults who sympathize with the practice of bossnapping. As with other procedures, there are three steps to building and summarizing a confidence interval for proportions: Plan, Do, and Report.

<b>WHO</b>	Adults in France
<b>WHAT</b>	Proportion who sympathize with the practice of bossnapping
<b>WHEN</b>	April 2–3, 2009
<b>WHERE</b>	France
<b>HOW</b>	1010 adults were randomly sampled by the French Institute of Public Opinion (l’Ifop) for the magazine <i>Paris Match</i>
<b>WHY</b>	To investigate public opinion of bossnapping

## PLAN

**Setup** State the context of the question.

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

Choose and state a confidence level.

**Model** Think about the assumptions and check the conditions to decide whether we can use the Normal model.

We want to find an interval that is likely with 95% confidence to contain the true proportion,  $p$ , of French adults who sympathize with the practice of bossnapping. We have a random sample of 1010 French adults, with a sample proportion of 63%.

- ✓ **Independence Assumption:** A French polling agency, l’Ifop, phoned a random sample of French adults. It is unlikely that any respondent influenced another.
- ✓ **Randomization Condition:** l’Ifop drew a random sample from all French adults. We don’t have details of their randomization but assume that we can trust it.
- ✓ **10% Condition:** Although sampling was necessarily without replacement, there are many more French adults than were sampled. The sample is certainly less than 10% of the population.

(continued)

State the sampling distribution model for the statistic. Choose your method.

✓ **Success/Failure Condition:**

$$n\hat{p} = 1010 \times 0.63 = 636 \geq 10 \text{ and}$$

$$n\hat{q} = 1010 \times 0.37 = 374 \geq 10,$$

so the sample is large enough.

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

**DO**

**Mechanics** Construct the confidence interval. 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.)

$$n = 1010, \hat{p} = 0.63, \text{ so}$$

$$SE(\hat{p}) = \sqrt{\frac{0.63 \times 0.37}{1010}} = 0.015$$

Next, find the margin of error. We could informally use 2 for our critical value, but 1.96 is more accurate.<sup>5</sup>

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

The margin of error is:

$$ME = z^* \times SE(\hat{p}) = 1.96 \times 0.015 = 0.029$$

Write the confidence interval.

So the 95% confidence interval is:

$$0.63 \pm 0.029 \text{ or } (0.601, 0.659).$$

**REALITY CHECK**

Check that the interval is plausible. We may not have a strong expectation for the center, but the width of the interval depends primarily on the sample size—especially when the estimated proportion is near 0.5.

The confidence interval covers a range of about plus or minus 3%. That's about the width we might expect for a sample size of about 1000 (when  $\hat{p}$  is reasonably close to 0.5).

**REPORT**

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

**MEMO**

**Re: Bossnapping Survey**

The polling agency Ifop surveyed 1010 French adults and asked whether they approved, were sympathetic to or disapproved of recent bossnapping actions. Although we can't know the true proportion of French adults who were sympathetic (without supporting outright), based on this survey we can be 95% confident that between 60.1% and 65.9% of all French adults were. Because this is an ongoing concern for public safety, we may want to repeat the survey to obtain more current data. We may also want to keep these results in mind for future corporate public relations.

**Just Checking**

Think some more about the 95% confidence interval we just created for the proportion of French adults who were sympathetic to bossnapping.

- 1 If we wanted to be 98% confident, would our confidence interval need to be wider or narrower?
- 2 Our margin of error was about  $\pm 3\%$ . If we wanted to reduce it to  $\pm 2\%$  without increasing the sample size, would our level of confidence be higher or lower?
- 3 If the organization had polled more people, would the interval's margin of error have likely been larger or smaller?

<sup>5</sup>If you are following along on your calculator and not rounding off (as we have done for this example), you'll get  $SE = 0.0151918$  and a  $ME$  of  $0.029776$ .

## 11.4 Choosing the Sample Size

Every confidence interval must balance precision—the width of the interval—against confidence. Although it is good to be precise and comforting to be confident, there is a trade-off between the two. A confidence interval that says that the percentage is between 10% and 90% wouldn't be of much use, although you could be quite confident that it covered the true proportion. An interval from 43% to 44% is reassuringly precise, but not if it carries a confidence level of 35%. It's a rare study that reports confidence levels lower than 80%. Levels of 95% or 99% are more common.

The time to decide whether the 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.

Consider a company planning to offer a new service to their customers. Product managers want to estimate the proportion of customers who are likely to purchase this new service to within 3% with 95% confidence. How large a sample do they need?

Let's look at the margin of error:

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

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

They want to find  $n$ , the sample size. To find  $n$ , they need a value for  $\hat{p}$ . They don't know  $\hat{p}$  because they don't have a sample yet, but they can probably guess a value. The worst case—the value that makes the SD (and therefore  $n$ ) largest—is 0.50, so if they use that value for  $\hat{p}$ , they'll certainly be safe.

The company's equation, then, is:

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

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

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

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

Then square the result to find  $n$ :

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

That method will probably give a value with a fraction. To be safe, always round up. The company will need at least 1068 respondents to keep the margin of error as small as 3% with a confidence level of 95%.

Unfortunately, bigger samples cost more money and require 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, you must *quadruple* the sample size.

Generally a margin of error of 5% or less is acceptable, but different circumstances call for different standards. The size of the margin of error may be a marketing decision or one determined by the amount of financial risk you (or the company) are willing to accept. Drawing a large sample to get a smaller ME,

### What $\hat{p}$ Should We Use?

Often you'll have an estimate of the population proportion based on experience or perhaps on 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  $\hat{p} = 0.5$ . That will determine the largest sample necessary regardless of the true proportion. It's the *worst case* scenario.

**Why 1000?**

Public opinion polls often use a sample size of 1000, which gives an ME of about 3% (at 95% confidence) when  $p$  is near 0.5. But businesses and nonprofit organizations often use much larger samples to estimate the response to a direct mail campaign. Why? Because the proportion of people who respond to these mailings is very low, often 5% or even less. An ME of 3% may not be precise enough if the response rate is that low. Instead, an ME like 0.1% would be more useful, and that requires a very large sample size.



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. A news event or new product announcement can change opinions 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. Also 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 complete or nearly complete response by a modest-size sample can yield useful results.

Surveys are not the only place where proportions pop up. Credit card banks sample huge mailing lists to estimate what proportion of people will accept a credit card offer. Even pilot studies may be mailed to 50,000 customers or more. Most of these customers don't respond. But in this case, that doesn't make the sample smaller. In fact, they did respond in a way—they just said "No thanks." To the bank, the response rate<sup>6</sup> is  $\hat{p}$ . With a typical success rate below 1%, 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 should take care when estimating the size needed. For our election poll example, 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 value of  $p$  that they expect to find from their experience.

**How much of a difference can it make?**

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. To be within a tenth of a percentage point, or 0.001 of the true acquisition rate with 95% confidence, how big does the test mailing have to be? Similar mailings in the past lead them to expect that about 0.5% of the people receiving the offer will accept it. Using those values, they find:

$$ME = 0.001 = z^* \sqrt{\frac{pq}{n}} = 1.96 \sqrt{\frac{(0.005)(0.995)}{n}}$$

$$(0.001)^2 = 1.96^2 \frac{(0.005)(0.995)}{n} \Rightarrow n = \frac{1.96^2(0.005)(0.995)}{(0.001)^2}$$

$$= 19,111.96 \text{ or } 19,112$$

That's a perfectly reasonable size for a trial mailing. But if they had used 0.50 for their estimate 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 result!

<sup>6</sup>Be careful. In marketing studies like this *every* mailing yields a response—"yes" or "no"—and response rate means the success rate, the proportion of customers who accept the offer. That's a different use of the term response rate from the one used in survey response.



## For Example

## Sample size calculations for a confidence interval for a proportion

The President of the Chamber of Commerce in the previous examples is worried that the 99% confidence interval is too wide. Recall that it was  $(0.503, 0.617)$ , which has a width of 0.114.

**Question:** How large a sample would she need to take to have a 99% interval half as wide? One quarter as wide? What if she wanted a 95% confidence interval that was plus or minus 3 percentage points? How large a sample would she need?

**Answer:** Because the formula for the confidence interval is dependent on the inverse of the square root of the sample size:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}},$$

a sample size four times as large will produce a confidence interval *half* as wide. The original 99% confidence interval had a sample size of 516. If she wants it half as wide, she will need about  $4 \times 516 = 2064$  respondents. To get it a quarter as wide she'd need  $4^2 \times 516 = 8192$  respondents!

If she wants a 99% confidence interval that's plus or minus 3 percentage points, she must calculate

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \hat{p} \pm 0.03$$

so

$$2.576 \sqrt{\frac{(0.5)(0.5)}{n}} = 0.03$$

which means that

$$n \approx \left(\frac{2.576}{0.03}\right)^2 (0.5)(0.5) = 1843.27$$

Rounding up, she'd need 1844 respondents. We used 0.5 because we didn't have any information about the election before taking the survey. Using  $p = 0.56$  instead would give  $n = 1817$ .

## \*11.5 A Confidence Interval for Small Samples

When the Success/Failure condition fails, all is not lost. A simple adjustment to the calculation lets us make a confidence interval anyway. All we do is add four *synthetic* observations, two to the successes and two to the failures. So instead of the proportion

$\hat{p} = \frac{y}{n}$ , we use the adjusted proportion  $\tilde{p} = \frac{y+2}{n+4}$ , and for convenience, we write

$\tilde{n} = n + 4$ . We modify the interval by using these adjusted values for both the center of the interval *and* the margin of error. Now the adjusted interval is:

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

This adjusted form gives better performance overall<sup>7</sup> and works much better for proportions near 0 or 1. It has the additional advantage that we don't need to check the Success/Failure condition that  $n\hat{p}$  and  $n\hat{q}$  are greater than 10.

Suppose a student in an advertising class is studying the impact of ads placed during the Super Bowl, and wants to know what proportion of students

<sup>7</sup>By "better performance" we mean that the 95% confidence interval's actual chance of covering the true population proportion is closer to 95%. Simulation studies have shown that our original, simpler confidence interval covers the true population proportion less than 95% of the time when the sample size is small or the proportion is very close to 0 or 1. The original idea was E. B. Wilson's, but the simplified approach we suggest here appeared in A. Agresti and B. A. Coull, "Approximate Is Better Than 'Exact' for Interval Estimation of Binomial Proportions," *The American Statistician*, 52 (1998): 119–126.

on campus watched it. She takes a random sample of 25 students and find that all 25 watched the Super Bowl for a  $\hat{p}$  of 100%. A 95% confidence interval is  $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.0 \pm 1.96\sqrt{\frac{1.0(0.0)}{25}} = (1.0, 1.0)$ . Does she really believe that *every* one of the 30,000 students on her campus watched the Super Bowl? Probably not. And she realizes that the Success/Failure condition is severely violated because there are *no* failures.

Using the pseudo observation method described above, she adds two successes and two failures to the sample to get 27/29 successes, for  $\tilde{p} = \frac{27}{29} = 0.931$ . The standard error is no longer 0, but  $SE(\tilde{p}) = \sqrt{\frac{\tilde{p}\tilde{q}}{\tilde{n}}} = \sqrt{\frac{(0.931)(0.069)}{29}} = 0.047$ .

Now, a 95% confidence interval is  $0.931 \pm 1.96(0.047) = (0.839, 1.023)$ . In other words, she's 95% confident that between 83.9% and 102.3% of all students on campus watched the Super Bowl. Because any number greater than 100% makes no sense, she will report simply that with 95% confidence the proportion is at least 83.9%.

## What Can Go Wrong?

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

- **Be sure to use the right language to describe your confidence intervals.** Technically, you should say “I am 95% confident that the interval from 32.2% to 35.8% captures the true proportion of U.S. adults who thought the economy was improving in March 2010.” That formal phrasing emphasizes that *your confidence (and your uncertainty) is about the interval, not the true proportion*. But you may choose a more casual phrasing like “I am 95% confident that between 32.2% and 35.8% of U.S. adults thought the economy was improving in March 2010.” 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. It's the focus of both our confidence and our doubt.
- **Don't suggest that the parameter varies.** A statement like “there is a 95% chance that the true proportion is between 32.2% and 35.8%” sounds as though you think the population proportion wanders around and sometimes happens to fall between 32.2% and 35.8%. 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 thought the economy was improving in March 2010 will be between 32.2% and 35.8%” is just wrong. The interval isn't about sample proportions but about the population proportion. There is nothing special about the sample we happen to have; it doesn't establish a standard for other samples.
- **Don't be certain about the parameter.** Saying “between 32.2% and 35.8% of U.S. adults thought the economy was improving in March 2010” asserts that the population proportion cannot be outside that interval. Of course, you 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 32.2% and 35.8%." Of course, you are—in fact, we calculated that our sample proportion was 34.0%. 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 32.2% and 35.8% of all U.S. adults think the economy is improving." Gallup sampled adults during March 2010, and public opinion shifts over time.
- **Do take responsibility.** Confidence intervals are about *uncertainty*. *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 32.2% and 35.8% of U.S. adults thought the economy was improving in March 2010."

Confidence intervals and margins of error depend crucially on the assumption and conditions. When they're not true the results may be invalid. For your own surveys, follow the survey designs from Chapter 3. For surveys you read about, so be sure to:

- **Watch out for biased sampling.** 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 a sample are mutually independent is one that you usually cannot check. It always pays to think about it, though.
- **Be careful of sample size.** The validity of the confidence interval for proportions may be affected by sample size. Avoid using the confidence interval on "small" samples.

## Ethics in Action

One of Tim Solsby's major responsibilities at MassEast Federal Credit Union is managing online services and website content. In an effort to better serve MassEast members, Tim routinely visits the sites of other financial institutions to get ideas on how he can improve MassEast's online presence. One of the features that caught his attention was a "teen network" that focused on educating teenagers about personal finances. He thought that this was a novel idea and one that could help build a stronger online community among MassEast's members. The executive board of MassEast was meeting next month to consider proposals for improving credit union services, and Tim was eager to present his idea for adding an online teen network. To strengthen his proposal, he decided to poll current credit union members. On the MassEast Federal Credit Union website, he posted an online survey. Among the questions he asked are "Do you have teenage children in your household?" and "Would you encourage your teenage children to

learn more about managing personal finances?" Based on 850 responses, Tim constructed a 95% confidence interval and was able to estimate (with 95% confidence) that between 69% and 75% of MassEast members had teenage children at home and that between 62% and 68% would encourage their teenagers to learn more about managing personal finances. Tim believed these results would help convince the executive board that MassEast should add this feature to its website.

**ETHICAL ISSUE** *The sampling method introduces bias because it is a voluntary response sample and not a random sample. Customers who do have teenagers are more likely to respond than those that do not (related to Item A, ASA Ethical Guidelines).*

**ETHICAL SOLUTION** *Tim should revise his sampling methods. He might draw a simple random sample of credit union customers and try and contact them by mail or telephone. Whatever method he uses, Tim needs to disclose the sampling procedure to the Board and discuss possible sources of bias.*

## What Have We Learned?

### Learning Objectives

- Construct a confidence interval for a proportion,  $p$ , as the statistic,  $\hat{p}$ , plus and minus a **margin of error**.
  - The margin of error consists of a **critical value** based on the sampling model times a **standard error** based on the sample.
  - The critical value is found from the Normal model.
  - The standard error of a sample proportion is calculated as  $\sqrt{\frac{\hat{p}\hat{q}}{n}}$ .
- Interpret a confidence interval correctly.
  - You can claim to have the specified level of confidence that the interval you have computed actually covers the true value.
- Understand the importance of the sample size,  $n$ , in improving both the certainty (confidence level) and precision (margin of error).
  - For the same sample size and proportion, more certainty requires less precision and more precision requires less certainty.
- Know and check the assumptions and conditions for finding and interpreting confidence intervals.
  - Independence Assumption or Randomization Condition
  - 10% Condition
  - Success/Failure Condition
- Be able to invert the calculation of the margin of error to find the sample size required, given a proportion, a confidence level, and a desired margin of error.

### Terms

Confidence interval	<p>An interval of values usually of the form</p> $\text{estimate} \pm \text{margin of error}$ <p>found from data in such a way that a percentage of all random samples can be expected to yield intervals that capture the true parameter value.</p>
Critical value	<p>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 <math>z^*</math>, is usually found from a table or with technology.</p>
Margin of error (ME)	<p>In a confidence interval, the extent of the interval on either side of the observed statistic value. 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.</p>
One-proportion $z$ -interval	<p>A confidence interval for the true value of a proportion. The confidence interval is</p> $\hat{p} \pm z^*SE(\hat{p})$ <p>where <math>z^*</math> is a critical value from the Standard Normal model corresponding to the specified confidence level.</p>

## Technology Help: Confidence Intervals for Proportions

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.

### EXCEL

$\lambda$ LSTAT  $\triangleright$

Inference methods for proportions are not part of the standard Excel tool set, but you can compute a confidence interval using Excel's equations:

For example, suppose you have 100 observations in cells A1:A100 and each cell is "yes" or "no."

- In cell B2, enter `=countif(A1:A100,"yes")/100` to compute the proportion of "yes" responses. (The 100 is because you have 100 observations. Replace it with the number of observations you actually have.)
- In cell B3, enter `=sqrt(B2*(1-B2)/100)` to compute the standard error.
- In cell B4, enter `=normsinv(.975)` for a 95% confidence interval.
- In cell B5, enter `=B2-B4*B3` as the lower end of the CI.
- In cell B6, enter `=B2+B4*B3` as the upper end of the CI.

### Comments

For summarized data, compute the proportion in cell B2 according to whether your summaries are counts, percentages, or already

proportions, and continue with the example, using total count in place of the "100" in the second step.

### JMP

For a **categorical** variable that holds category labels, the **Distribution** platform includes tests and intervals for proportions.

For summarized data,

- Put the category names in one variable and the frequencies in an adjacent variable.
- Designate the frequency column to have the role of frequency.
- Then use the Distribution platform.

### Comments

JMP uses slightly different methods for proportion inferences than those discussed in this text. Your answers are likely to be slightly different, especially for small samples.

### MINITAB

Choose **Basic Statistics** from the **Stat** menu.

- Choose **1Proportion** from the Basic Statistics submenu.
- If the data are category names in a variable, assign the variable from the variable list box to the **Samples in columns** box. If you have summarized data, click the **Summarized Data** button and fill in the number of trials and the number of successes.
- Click the **Options** button and specify the remaining details.
- If you have a large sample, check **Use test and interval based on normal distribution**. Click the **OK** button.

### Comments

When working from a variable that names categories, MINITAB treats the last category as the "success" category. You can specify how the categories should be ordered.

### SPSS

SPSS does not find confidence intervals for proportions.

## Brief CASE

### Investment

During the period from June 27–29, 2003, the Gallup Organization asked stock market investors questions about the amount and type of their investments. The questions asked the investors were:

1. Is the total amount of your investments right now \$10,000 or more, or is it less than \$10,000?
2. If you had \$1000 to invest, would you be more likely to invest it in stocks or bonds?

In response to the first question, 65% of the 692 investors reported that they currently have at least \$10,000 invested in the stock market. In response to the second question, 48% of the 692 investors reported that they would be more likely to invest in stocks (over bonds). Compute the standard error for each sample proportion. Compute and describe the 95% confidence intervals in the context of the question. What would the size of the sample need to be for the margin of error to be 3%?

Find a recent survey about investment practices or opinions and write up a short report on your findings.

### Forecasting Demand

Utilities must forecast the demand for energy use far into the future because it takes decades to plan and build new power plants. Ron Bears, who worked for a northeast utility company, had the job of predicting the proportion of homes that would choose to use electricity to heat their homes. Although he was prepared to report a confidence interval for the true proportion, after seeing his preliminary report, his management demanded a single number as his prediction.

Help Ron explain to his management why a confidence interval for the desired proportion would be more useful for planning purposes. Explain how the precision of the interval and the confidence we can have in it are related to each other. Discuss the business consequences of an interval that is too narrow and the consequences of an interval with too low a confidence level.

## Exercises

### SECTION 11.1

1. For each situation below identify the population and the sample and identify  $p$  and  $\hat{p}$  if appropriate and what the value of  $\hat{p}$  is. Would you trust a confidence interval for the true proportion based on these data? Explain briefly why or why not.

a) As concertgoers enter a stadium, a security guard randomly inspects their backpacks for alcoholic beverages. Of the 130 backpacks checked so far, 17 contained alcoholic beverages of some kind. The guards want to estimate the percentage of all backpacks of concertgoers at this concert that contain alcoholic beverages.

b) The website of the English newspaper *The Guardian* asked visitors to the site to say whether they approved of recent “bossnapping” actions by British workers who were outraged over being fired. Of those who responded, 49.2% said “Yes. Desperate times, desperate measures.”

c) An airline wants to know the weight of carry-on baggage that customers take on their international routes, so they take a random sample of 50 bags and find that the average weight is 17.3 pounds.

2. For each situation below identify the population and the sample and explain what  $p$  and  $\hat{p}$  represent and what the

value of  $\hat{p}$  is. Would you trust a confidence interval for the true proportion based on these data? Explain briefly why or why not.

- A marketing analyst conducts a large survey of her customers to find out how much money they plan to spend at the company website in the next 6 months. The average amount reported from the 534 respondents is \$145.34.
- A campus survey on a large campus (40,000 students) is trying to find out whether students approve of a new parking policy allowing students to park in previously inaccessible parking lots, but for a small fee. Surveys are sent out by mail and e-mail. Of the 243 surveys returned, 134 are in favor of the change.
- The human resources department of a large Fortune 100 company wants to find out how many employees would take advantage of an on-site day care facility. They send out an e-mail to 500 employees and receive responses from 450 of them. Of those responding, 75 say that they would take advantage of such a facility.

**3.** A survey of 200 students is selected randomly on a large university campus. They are asked if they use a laptop in class to take notes. Suppose that based on the survey, 70 of the 200 students responded “yes.”

- What is the value of the sample proportion  $\hat{p}$ ?
- What is the standard error of the sample proportion?
- Construct an approximate 95% confidence interval for the true proportion  $p$  by taking  $\pm 2$  SEs from the sample proportion.

**4.** From a survey of 250 coworkers you find that 155 would like the company to provide on-site day care.

- What is the value of the sample proportion  $\hat{p}$ ?
- What is the standard error of the sample proportion?
- Construct an approximate 95% confidence interval for the true proportion  $p$  by taking  $\pm 2$  SEs from the sample proportion.

### SECTION 11.2

**5.** From a survey of coworkers you find that 48% of 200 have already received this year’s flu vaccine. An approximate 95% confidence interval is (0.409, 0.551). Which of the following are true? If not, explain briefly.

- 95% of the coworkers fall in the interval (0.409, 0.551).
- We are 95% confident that the proportion of coworkers who have received this year’s flu vaccine is between 40.9% and 55.1%.
- There is a 95% chance that a random selected coworker has received the vaccine.
- There is a 48% chance that a random selected coworker has received the vaccine.
- We are 95% confident that between 40.9% and 55.1% of the samples will have a proportion near 48%.

**6.** As in Exercise 5, from a survey of coworkers you find that 48% of 200 have already received this year’s flu

vaccine. An approximate 95% confidence interval is (0.409, 0.551).

- How would the confidence interval change if the sample size had been 800 instead of 200?
- How would the confidence interval change if the confidence level had been 90% instead of 95%?
- How would the confidence interval change if the confidence level had been 99% instead of 95%?

### SECTION 11.3

**7.** Consider each situation described. Identify the population and the sample, explain what  $p$  and  $\hat{p}$  represent, and tell whether the methods of this chapter can be used to create a confidence interval.

- A consumer group hoping to assess customer experiences with auto dealers surveys 167 people who recently bought new cars; 3% of them expressed dissatisfaction with the salesperson.
- A cell phone service provider wants to know what percent of U.S. college students have cell phones. A total of 2883 students were asked as they entered a football stadium, and 2243 indicated they had phones with them.

**8.** Consider each situation described. Identify the population and the sample, explain what  $p$  and  $\hat{p}$  represent, and tell whether the methods of this chapter can be used to create a confidence interval.

- A total of 240 potato plants in a field in Maine are randomly checked, and only 7 show signs of blight. How severe is the blight problem for the U.S. potato industry?
- Concerned about workers’ compensation costs, a small company decided to investigate on-the-job injuries. The company reported that 12 of their 309 employees suffered an injury on the job last year. What can the company expect in future years?

### SECTION 11.4

**9.** Suppose you want to estimate the proportion of traditional college students on your campus who own their own car. You have no preconceived idea of what that proportion might be.

- What sample size is needed if you wish to be 95% confident that your estimate is within 0.02 of the true proportion?
- What sample size is needed if you wish to be 99% confident that your estimate is within 0.02 of the true proportion?
- What sample size is needed if you wish to be 95% confident that your estimate is within 0.05 of the true proportion?

**10.** As in Exercise 9, you want to estimate the proportion of traditional college students on your campus who own their own car. However, from some research on other college campuses, you believe the proportion will be near 20%.

- a) What sample size is needed if you wish to be 95% confident that your estimate is within 0.02 of the true proportion?  
 b) What sample size is needed if you wish to be 99% confident that your estimate is within 0.02 of the true proportion?  
 c) What sample size is needed if you wish to be 95% confident that your estimate is within 0.05 of the true proportion?

**11.** It's believed that as many as 25% of adults over age 50 never graduated from high school. We wish to see if this percentage is the same among the 25 to 30 age group.

- a) How many of this younger age group must we survey in order to estimate the proportion of nongrads to within 6% with 90% confidence?  
 b) Suppose we want to cut the margin of error to 4%. What's the necessary sample size?  
 c) What sample size would produce a margin of error of 3%?

**12.** In preparing a report on the economy, we need to estimate the percentage of businesses that plan to hire additional employees in the next 60 days.

- a) How many randomly selected employers must we contact in order to create an estimate in which we are 98% confident with a margin of error of 5%?  
 b) Suppose we want to reduce the margin of error to 3%. What sample size will suffice?  
 c) Why might it not be worth the effort to try to get an interval with a margin of error of 1%?

## CHAPTER EXERCISES

**13. Margin of error.** A corporate executive reports the results of an employee satisfaction survey, stating that 52% of employees say they are either "satisfied" or "extremely satisfied" with their jobs, and then says "the margin of error is plus or minus 4%." Explain carefully what that means.

**14. Margin of error.** A market researcher estimates the percentage of adults between the ages of 21 and 39 who will see their television ad is 15%, adding that he believes his estimate has a margin of error of about 3%. Explain what the margin of error means.

**15. Conditions.** Consider each situation described below. Identify the population and the sample, explain what  $p$  and  $\hat{p}$  represent, and tell whether the methods of this chapter can be used to create a confidence interval.

- a) Police set up an auto checkpoint at which drivers are stopped and their cars inspected for safety problems. They find that 14 of the 134 cars stopped have at least one safety violation. They want to estimate the proportion of all cars in this area that may be unsafe.  
 b) A CNN show asks viewers to register their opinions on corporate corruption by logging onto a website. Of the 602 people who voted, 488 thought corporate corruption was "worse" this year than last year. The show wants to estimate the level of support among the general public.

**16. More conditions.** Consider each situation described below. Identify the population and the sample, explain what  $p$  and  $\hat{p}$  represent, and tell whether the methods of this chapter can be used to create a confidence interval.

- a) A large company with 10,000 employees at their main research site is considering moving its day care center off-site to save money. Human resources gathers employees' opinions by sending a questionnaire home with all employees; 380 surveys are returned, with 228 employees in favor of the change.  
 b) A company sold 1632 MP3 players last month, and within a week, 1388 of the customers had registered their products online at the company website. The company wants to estimate the percentage of all their customers who enroll their products.

**17. Catalog sales.** A catalog sales company promises to deliver orders placed on the Internet within 3 days. Follow-up calls to a few randomly selected customers show that a 95% confidence interval for the proportion of all orders that arrive on time is  $88\% \pm 6\%$ . What does this mean? Are the conclusions in parts a–e correct? Explain.

- a) Between 82% and 94% of all orders arrive on time.  
 b) 95% of all random samples of customers will show that 88% of orders arrive on time.  
 c) 95% of all random samples of customers will show that 82% to 94% of orders arrive on time.  
 d) The company is 95% sure that between 82% and 94% of the orders placed by the customers in this sample arrived on time.  
 e) On 95% of the days, between 82% and 94% of the orders will arrive on time.

**18. Belgian euro.** Recently, two students made worldwide headlines by spinning a Belgian euro 250 times and getting 140 heads—that's 56%. That makes the 90% confidence interval (51%, 61%). What does this mean? Are the conclusions in parts a–e correct? Explain your answers.

- a) Between 51% and 61% of all euros are unfair.  
 b) We are 90% sure that in this experiment this euro landed heads between 51% and 61% of the spins.  
 c) We are 90% sure that spun euros will land heads between 51% and 61% of the time.  
 d) If you spin a euro many times, you can be 90% sure of getting between 51% and 61% heads.  
 e) 90% of all spun euros will land heads between 51% and 61% of the time.

**19. Confidence intervals.** Several factors are involved in the creation of a confidence interval. Among them are the sample size, the level of confidence, and the margin of error. Which statements are true?

- a) For a given sample size, higher confidence means a smaller margin of error.  
 b) For a specified confidence level, larger samples provide smaller margins of error.



- c) For a fixed margin of error, larger samples provide greater confidence.  
 d) For a given confidence level, halving the margin of error requires a sample twice as large.

**20. Confidence intervals, again.** Several factors are involved in the creation of a confidence interval. Among them are the sample size, the level of confidence, and the margin of error. Which statements are true?

- a) For a given sample size, reducing the margin of error will mean lower confidence.  
 b) For a certain confidence level, you can get a smaller margin of error by selecting a bigger sample.  
 c) For a fixed margin of error, smaller samples will mean lower confidence.  
 d) For a given confidence level, a sample 9 times as large will make a margin of error one third as big.

**21. Cars.** A student is considering publishing a new magazine aimed directly at owners of Japanese automobiles. He wanted to estimate the fraction of cars in the United States that are made in Japan. The computer output summarizes the results of a random sample of 50 autos. Explain carefully what it tells you.

z-interval for proportion  
 With 90.00% confidence  
 $0.29938661 < p(\text{japan}) < 0.46984416$

**22. Quality control.** For quality control purposes, 900 ceramic tiles were inspected to determine the proportion of defective (e.g., cracked, uneven finish, etc.) tiles. Assuming that these tiles are representative of all tiles manufactured by an Italian tile company, what can you conclude based on the computer output?

z-interval for proportion  
 With 95.00% confidence  
 $0.025 < p(\text{defective}) < 0.035$

**23. E-mail.** A small company involved in e-commerce is interested in statistics concerning the use of e-mail. A poll found that 38% of a random sample of 1012 adults, who use a computer at their home, work, or school, said that they do not send or receive e-mail.

- a) Find the margin of error for this poll if we want 90% confidence in our estimate of the percent of American adults who do not use e-mail.  
 b) Explain what that margin of error means.  
 c) If we want to be 99% confident, will the margin of error be larger or smaller? Explain.  
 d) Find that margin of error.  
 e) In general, if all other aspects of the situation remain the same, will smaller margins of error involve greater or less confidence in the interval?

**24. Biotechnology.** A biotechnology firm in Boston is planning its investment strategy for future products and

research labs. A poll found that only 8% of a random sample of 1012 U.S. adults approved of attempts to clone a human.

- a) Find the margin of error for this poll if we want 95% confidence in our estimate of the percent of American adults who approve of cloning humans.  
 b) Explain what that margin of error means.  
 c) If we only need to be 90% confident, will the margin of error be larger or smaller? Explain.  
 d) Find that margin of error.  
 e) In general, if all other aspects of the situation remain the same, would smaller samples produce smaller or larger margins of error?

**25. Teenage drivers.** An insurance company checks police records on 582 accidents selected at random and notes that teenagers were at the wheel in 91 of them.

- a) Create a 95% confidence interval for the percentage of all auto accidents that involve teenage drivers.  
 b) Explain what your interval means.  
 c) Explain what “95% confidence” means.  
 d) A politician urging tighter restrictions on drivers’ licenses issued to teens says, “In one of every five auto accidents, a teenager is behind the wheel.” Does your confidence interval support or contradict this statement? Explain.

**26. Advertisers.** Direct mail advertisers send solicitations (“junk mail”) to thousands of potential customers in the hope that some will buy the company’s product. The response rate is usually quite low. Suppose a company wants to test the response to a new flyer and sends it to 1000 people randomly selected from their mailing list of over 200,000 people. They get orders from 123 of the recipients.

- a) Create a 90% confidence interval for the percentage of people the company contacts who may buy something.  
 b) Explain what this interval means.  
 c) Explain what “90% confidence” means.  
 d) The company must decide whether to now do a mass mailing. The mailing won’t be cost-effective unless it produces at least a 5% return. What does your confidence interval suggest? Explain.

**27. Retailers.** Some food retailers propose subjecting food to a low level of radiation in order to improve safety, but sale of such “irradiated” food is opposed by many people. Suppose a grocer wants to find out what his customers think. He has cashiers distribute surveys at checkout and ask customers to fill them out and drop them in a box near the front door. He gets responses from 122 customers, of whom 78 oppose the radiation treatments. What can the grocer conclude about the opinions of all his customers?

**28. Local news.** The mayor of a small city has suggested that the state locate a new prison there, arguing that the construction project and resulting jobs will be good for the local economy. A total of 183 residents show up for a

public hearing on the proposal, and a show of hands finds 31 in favor of the prison project. What can the city council conclude about public support for the mayor's initiative?

**29. Internet music.** In a survey on downloading music, the Gallup Poll asked 703 Internet users if they “ever downloaded music from an Internet site that was not authorized by a record company, or not,” and 18% responded “yes.” Construct a 95% confidence interval for the true proportion of Internet users who have downloaded music from an Internet site that was not authorized.

**30. Economy worries.** In 2008, a Gallup Poll asked 2335 U.S. adults, aged 18 or over, how they rated economic conditions. In a poll conducted from January 27–February 1, 2008, only 24% rated the economy as Excellent/Good. Construct a 95% confidence interval for the true proportion of Americans who rated the U.S. economy as Excellent/Good.

**31. International business.** In Canada, the vast majority (90%) of companies in the chemical industry are ISO 14001 certified. The ISO 14001 is an international standard for environmental management systems. An environmental group wished to estimate the percentage of U.S. chemical companies that are ISO 14001 certified. Of the 550 chemical companies sampled, 385 are certified.

a) What proportion of the sample reported being certified?  
b) Create a 95% confidence interval for the proportion of U.S. chemical companies with ISO 14001 certification. (Be sure to check conditions.) Compare to the Canadian proportion.

**32. Worldwide survey.** In Chapter 4, Exercise 27, we learned that GfK Roper surveyed people worldwide asking them “how important is acquiring wealth to you.” Of 1535 respondents in India, 1168 said that it was of more than average importance. In the United States of 1317 respondents, 596 said it was of more than average importance.

a) What proportion thought acquiring wealth was of more than average importance in each country's sample?  
b) Create a 95% confidence interval for the proportion who thought it was of more than average importance in India. (Be sure to test conditions.) Compare that to a confidence interval for the U.S. population.

**33. Business ethics.** In a survey on corporate ethics, a poll split a sample at random, asking 538 faculty and corporate recruiters the question: “Generally speaking, do you believe that MBAs are more or less aware of ethical issues in business today than five years ago?” The other half were asked: “Generally speaking, do you believe that MBAs are less or more aware of ethical issues in business today than five years ago?” These may seem like the same questions, but sometimes the order of the choices matters. In response to the first question, 53% thought MBA graduates were more aware of

ethical issues, but when the question was phrased differently, this proportion dropped to 44%.

a) What kind of bias may be present here?  
b) Each group consisted of 538 respondents. If we combine them, considering the overall group to be one larger random sample, what is a 95% confidence interval for the proportion of the faculty and corporate recruiters that believe MBAs are more aware of ethical issues today?  
c) How does the margin of error based on this pooled sample compare with the margins of error from the separate groups? Why?

**34. Media survey.** In 2007, a Gallup Poll conducted face-to-face interviews with 1006 adults in Saudi Arabia, aged 15 and older, asking them questions about how they get information. Among them was the question: “Is international television very important in keeping you well-informed about events in your country?” Gallup reported that 82% answered “yes” and noted that at 95% confidence there was a 3% margin of error and that “in addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of public opinion polls.”

a) What kinds of bias might they be referring to?  
b) Do you agree with their margin of error? Explain.

**35. Gambling.** A city ballot includes a local initiative that would legalize gambling. The issue is hotly contested, and two groups decide to conduct polls to predict the outcome. The local newspaper finds that 53% of 1200 randomly selected voters plan to vote “yes,” while a college Statistics class finds 54% of 450 randomly selected voters in support. Both groups will create 95% confidence intervals.

a) Without finding the confidence intervals, explain which one will have the larger margin of error.  
b) Find both confidence intervals.  
c) Which group concludes that the outcome is too close to call? Why?

**36. Casinos.** Governor Deval Patrick of Massachusetts proposed legalizing casinos in Massachusetts although they are not currently legal, and he included the revenue from them in his latest state budget. The website [www.boston.com](http://www.boston.com) conducted an Internet poll on the question: “Do you agree with the casino plan the governor is expected to unveil?” As of the end of 2007, there were 8663 votes cast, of which 63.5% of respondents said: “No. Raising revenues by allowing gambling is shortsighted.”

a) Find a 95% confidence interval for the proportion of voters in Massachusetts who would respond this way.  
b) Are the assumptions and conditions satisfied? Explain.

**37. Pharmaceutical company.** A pharmaceutical company is considering investing in a “new and improved” vitamin D supplement for children. Vitamin D, whether ingested as a dietary supplement or produced naturally when sunlight

falls upon the skin, is essential for strong, healthy bones. The bone disease rickets was largely eliminated in England during the 1950s, but now there is concern that a generation of children more likely to watch TV or play computer games than spend time outdoors is at increased risk. A recent study of 2700 children randomly selected from all parts of England found 20% of them deficient in vitamin D.

- Find a 98% confidence interval for the proportion of children in England who are deficient in vitamin D.
- Explain carefully what your interval means.
- Explain what “98% confidence” means.
- Does the study show that computer games are a likely cause of rickets? Explain.

**38. Wireless access.** In Chapter 4, Exercise 36, we saw that the Pew Internet and American Life Project polled 798 Internet users in December 2006, asking whether they have logged on to the Internet using a wireless device or not and 243 responded “Yes.”

- Find a 98% confidence interval for the proportion of all U.S. Internet users who have logged in using a wireless device.
- Explain carefully what your interval means.
- Explain what “98% confidence” means.

**\*39. Funding.** In 2005, a survey developed by Babson College and the Association of Women’s Business Centers (WBCs) was distributed to WBCs in the United States. Of a representative sample of 20 WBCs, 40% reported that they had received funding from the national Small Business Association (SBA).

- Check the assumptions and conditions for inference on proportions.
- If it’s appropriate, find a 90% confidence interval for the proportion of WBCs that receive SBA funding. If it’s not appropriate, explain and/or recommend an alternative action.

**\*40. Real estate survey.** A real estate agent looks over the 15 listings she has in a particular zip code in California and finds that 80% of them have swimming pools.

- Check the assumptions and conditions for inference on proportions.
- If it’s appropriate, find a 90% confidence interval for the proportion of houses in this zip code that have swimming pools. If it’s not appropriate, explain and/or recommend an alternative action.

**\*41. Benefits survey.** A paralegal at the Vermont State Attorney General’s office wants to know how many companies in Vermont provide health insurance benefits to all employees. She chooses 12 companies at random and finds that all 12 offer benefits.

- Check the assumptions and conditions for inference on proportions.
- Find a 95% confidence interval for the true proportion of companies that provide health insurance benefits to all their employees.

**\*42. Awareness survey.** A telemarketer at a credit card company is instructed to ask the next 18 customers that call into the 800 number whether they are aware of the new Platinum card that the company is offering. Of the 18, 17 said they were aware of the program.

- Check the assumptions and conditions for inference on proportions.
- Find a 95% confidence interval for the true proportion of customers who are aware of the new card.

**43. IRS.** In a random survey of 226 self-employed individuals, 20 reported having had their tax returns audited by the IRS in the past year. Estimate the proportion of self-employed individuals nationwide who’ve been audited by the IRS in the past year.

- Check the assumptions and conditions (to the extent you can) for constructing a confidence interval.
- Construct a 95% confidence interval.
- Interpret your interval.
- Explain what “95% confidence” means in this context.

**44. ACT, Inc.** In 2004, ACT, Inc. reported that 74% of 1644 randomly selected college freshmen returned to college the next year. Estimate the national freshman-to-sophomore retention rate.

- Check that the assumptions and conditions are met for inference on proportions.
- Construct a 98% confidence interval.
- Interpret your interval.
- Explain what “98% confidence” means in this context.

**45. Internet music, again.** A Gallup Poll (Exercise 29) asked Americans if the fact that they can make copies of songs on the Internet for free made them more likely—or less likely—to buy a performer’s CD. Only 13% responded that it made them “less likely.” The poll was based on a random sample of 703 Internet users.

- Check that the assumptions and conditions are met for inference on proportions.
- Find the 95% confidence interval for the true proportion of all U.S. Internet users who are “less likely” to buy CDs.

**46. ACT, Inc., again.** The ACT, Inc. study described in Exercise 44 was actually stratified by type of college—public or private. The retention rates were 71.9% among 505 students enrolled in public colleges and 74.9% among 1139 students enrolled in private colleges.

- Will the 95% confidence interval for the true national retention rate in private colleges be wider or narrower than the 95% confidence interval for the retention rate in public colleges? Explain.
- Find the 95% confidence interval for the public college retention rate.
- Should a public college whose retention rate is 75% proclaim that they do a better job than other public colleges of keeping freshmen in school? Explain.

**47. Politics.** A poll of 1005 U.S. adults split the sample into four age groups: ages 18–29, 30–49, 50–64, and 65+. In the youngest age group, 62% said that they thought the U.S. was ready for a woman president, as opposed to 35% who said “no, the country was not ready” (3% were undecided). The sample included 250 18- to 29-year-olds.

a) Do you expect the 95% confidence interval for the true proportion of all 18- to 29-year-olds who think the U.S. is ready for a woman president to be wider or narrower than the 95% confidence interval for the true proportion of all U.S. adults? Explain.

b) Find the 95% confidence interval for the true proportion of all 18- to 29-year-olds who believe the U.S. is ready for a woman president.

**48. Wireless access, again.** The survey in Exercise 38 asking about wireless Internet access also classified the 798 respondents by income.

	A	B	C	D
		Wireless Users	Other Internet Users	Total
1				
2	Under \$30K	34	128	162
3	\$30K--\$50K	31	133	164
4	\$50K--\$75K	44	72	116
5	Over \$75K	83	111	194
6	Don't know/refused	51	111	162
7	Total	243	555	798

a) Do you expect the 95% confidence interval for the true proportion of all those making more than \$75K who are wireless users to be wider or narrower than the 95% confidence interval for the true proportion among those who make between \$50K and \$75K? Explain briefly.

b) Find the 95% confidence interval for the true proportion of those making more than \$75K who are wireless users.

**49. More Internet music.** A random sample of 168 students was asked how many songs were in their digital music library and what fraction of them was legally purchased. Overall, they reported having a total of 117,079 songs, of which 23.1% were legal. The music industry would like a good estimate of the proportion of songs in students' digital music libraries that are legal.

a) Think carefully. What is the parameter being estimated? What is the population? What is the sample size?

b) Check the conditions for making a confidence interval.

c) Construct a 95% confidence interval for the fraction of legal digital music.

d) Explain what this interval means. Do you believe that you can be this confident about your result? Why or why not?

**50. Trade agreement.** Results from a January 2008 telephone survey conducted by Gallup showed that 57% of urban Colombian adults support a free trade agreement (FTA) with the United States. Gallup used a sample of 1000 urban Colombians aged 15 and older.

a) What is the parameter being estimated? What is the population? What is the sample size?

b) Check the conditions for making a confidence interval.

c) Construct a 95% confidence interval for the fraction of Colombians in agreement with the FTA.

d) Explain what this interval means. Do you believe that you can be this confident about your result? Why or why not?

**51. CDs.** A company manufacturing CDs is working on a new technology. A random sample of 703 Internet users were asked: “As you may know, some CDs are being manufactured so that you can only make one copy of the CD after you purchase it. Would you buy a CD with this technology, or would you refuse to buy it even if it was one you would normally buy?” Of these users, 64% responded that they would buy the CD.

a) Create a 90% confidence interval for this percentage.

b) If the company wants to cut the margin of error in half, how many users must they survey?

**52. Internet music, last time.** The research group that conducted the survey in Exercise 49 wants to provide the music industry with definitive information, but they believe that they could use a smaller sample next time. If the group is willing to have twice as big a margin of error, how many songs must be included?

**53. Graduation.** As in Exercise 11, we hope to estimate the percentage of adults aged 25 to 30 who never graduated from high school. What sample size would allow us to increase our confidence level to 95% while reducing the margin of error to only 2%?

**54. Better hiring info.** Editors of the business report in Exercise 12 are willing to accept a margin of error of 4% but want 99% confidence. How many randomly selected employers will they need to contact?

**55. Pilot study.** A state's environmental agency worries that a large percentage of cars may be violating clean air emissions standards. The agency hopes to check a sample of vehicles in order to estimate that percentage with a margin of error of 3% and 90% confidence. To gauge the size of the problem, the agency first picks 60 cars and finds 9 with faulty emissions systems. How many should be sampled for a full investigation?

**56. Another pilot study.** During routine conversations, the CEO of a new start-up reports that 22% of adults between the ages of 21 and 39 will purchase her new product. Hearing this, some investors decide to conduct a large-scale

study, hoping to estimate the proportion to within 4% with 98% confidence. How many randomly selected adults between the ages of 21 and 39 must they survey?

- 57. Approval rating.** A newspaper reports that the governor's approval rating stands at 65%. The article adds that the poll is based on a random sample of 972 adults and has a margin of error of 2.5%. What level of confidence did the pollsters use?
- 58. Amendment.** The Board of Directors of a publicly traded company says that a proposed amendment to their bylaws is likely to win approval in the upcoming election because a poll of 1505 stock owners indicated that 52% would vote in favor. The Board goes on to say that the margin of error for this poll was 3%.
- Explain why the poll is actually inconclusive.
  - What confidence level did the pollsters use?
- T 59. Customer spending.** The data set provided contains last month's credit card purchases of 500 customers randomly chosen from a segment of a major credit card issuer. The marketing department is considering a special offer for customers who spend more than \$1000 per month on their card. From these data construct a 95% confidence interval for the proportion of customers in this segment who will qualify.
- T 60. Advertising.** A philanthropic organization knows that its donors have an average age near 60 and is considering taking out an ad in the *American Association of Retired People (AARP)* magazine. An analyst wonders what proportion of their donors are actually 50 years old or older. He takes a

random sample of the records of 500 donors. From the data provided, construct a 95% confidence interval for the proportion of donors who are 50 years old or older.

**61. Health insurance.** Based on a 2007 survey of U.S. households (see [www.census.gov](http://www.census.gov)), 87% (out of 3060) of males in Massachusetts (MA) have health insurance.

- Examine the conditions for constructing a confidence interval for the proportion males in MA who had health insurance.
- Find the 95% confidence interval for the percent of males who have health insurance.
- Interpret your confidence interval.

**62. Health insurance, part 2.** Using the same survey and data as in Exercise 61, we find that 84% of those respondents in Massachusetts who identified themselves as Black/African-Americans (out of 440) had health insurance.

- Examine the conditions for constructing a confidence interval for the proportion of Black/African-Americans in MA who had health insurance.
- Find the 95% confidence interval.
- Interpret your confidence interval.

### Just Checking Answers

- Wider
- Lower
- Smaller