

FOUNDATIONS OF STATISTICAL INFERENCE

You're falling asleep and you hear a sound. Should you get up or go to sleep? Chances are, you've had an experience like this. Whether you get up or go to sleep depends on what kind of sound you heard, right? You can probably sense the difference between the noises you expect to hear at night (floorboard creaking, wind rustling trees outside, a family member fixing a late-night snack) and sounds that demand your attention (breaking glass, a crying baby, or someone calling your name). Without too much conscious thought on your part, you make inferences about information all the time. You tune out background noise but pay attention to people speaking to you and sounds that signal danger or distress.

Tuning out background noise may come naturally to you, but some inferences about sounds require concerted effort and specialized training. Picture a doctor listening to your heartbeat with a stethoscope. Does your heartbeat sound like a healthy person's or does it sound unusual? If you don't know what normal, healthy heartbeats sound like, it's impossible to identify an irregular heartbeat, even if you listen to one with a stethoscope. The doctor has been trained to distinguish normal, expected heartbeats from unusual sounds that may signal health problems.

When we conduct research and analyze data, we face a similar challenge of distinguishing mere random noise from meaningful results. Suppose that in one of our analyses of the American National Election Study we find that men give the Democratic Party an average feeling thermometer rating of 49, compared with a mean rating of 54 among women. Is this 5-point difference "big enough" to support the conclusion that females have higher regard for the Democratic Party than do males? Or should we instead decide that the difference is "too small" to warrant that conclusion? Suppose we are investigating the electoral mobilization of military veterans. One of our cross-tabulation analyses shows that 86 percent of veterans reported voting in the presidential election, compared with 77 percent of nonveterans. Does a 9-percentage-point difference allow us

Learning Objectives

In this chapter you will learn:

- Why random sampling is of cardinal importance in political research
- Why samples that seem small can yield accurate information about much larger groups
- How to figure out the margin of error for the information in a sample
- How to use the normal curve to make inferences about the information in a sample

Get the edge on your studies.
edge.sagepub.com/pollock

- Take a quiz to find out what you've learned.
- Review key terms with eFlashcards.
- Watch videos that enhance chapter content.

 **SAGE** edge™
for CQ Press

to say that veterans are more likely to vote than are nonveterans, or is the difference too fragile to support this conclusion?

As humans, we are prone to being fooled by randomness. We try to make sense of things by perceiving patterns and connections where none exist. Inferential statistics help us put analysis in perspective and properly interpret empirical results. **Inferential statistics** refers to a set of procedures for deciding how closely a relationship we observe in a sample corresponds to the unobserved relationship in the population from which the sample was drawn. Inferential statistics can help us decide whether the 5-point feeling thermometer difference between men and women represents a real gender difference in the population or whether the difference occurred by happenstance when the sample was taken. Inferential statistics will tell us how often a random sample will produce a 9-percentage-point difference in voting between veterans and nonveterans if, in fact, no difference exists in the population.

In this chapter we cover the essential foundations of inferential statistics. The techniques described in this chapter allow researchers to quantify how much noise can be expected when a statistic is estimated from a sample. A solid foundation in inferential statistics can help us avoid two potential pitfalls: paying too much attention to random noise and failing to perceive a reliable signal.

POPULATION PARAMETERS AND SAMPLE STATISTICS

In this section, we cover some specialized terminology that is essential to understanding inferential statistics. We've encountered some of these terms before. For example, we discussed different strategies for creating a sample of observations in Chapter 4.

Anyone who is interested in politics, society, or the economy wants to understand the attitudes, beliefs, or behavior of very large groups. These large aggregations of units are populations. A **population** may be defined generically as the universe of cases the researcher wants to describe. If we were studying the financial activity of political action committees (PACs) in the most recent congressional election, for example, our population would include all PAC contributions in the most recent election. Political scientists analyzing vote choice in the most recent congressional elections, by contrast, would define their population as all voting-eligible adults.

A characteristic of a population—the dollar amount of the average PAC contribution or the percentage of voting-age adults who voted—is called a **population parameter**. Figuring out a population's characteristics, its parameters, is a main goal of the social science investigator. In some situations, the researcher's goal is to describe the population, but the goal could also be to estimate causal effects—like the gender difference in party evaluation or the effect of being a military veteran on the likelihood of voting.

Researchers who enjoy complete access to their populations of interest—they can observe and measure every PAC, eligible voter, every member of Congress, Supreme Court decision, or whatever—are working with a **census**. A census allows the researcher to obtain measurements for all members of a population. Thus, the researcher does not need to infer or estimate any population parameters when describing the cases.¹

More often, however, researchers are unable to examine a population directly and must rely, instead, on a sample. A **sample** is a set of cases or observations drawn from a population. Samples, like death and taxes, are fixtures of life in social research.

Because population characteristics are frequently hidden from direct view, we turn to samples, which yield observable sample statistics. A **sample statistic** is an estimate of a population parameter, based on a sample drawn from the population.

Public opinion polls, for example, never survey every person in the population of interest (for example, all voting-eligible adults). Instead, a pollster surveys a random sample of likely voters, analyzes the survey results, and then infers or estimates a population characteristic from this sample statistic. A **random sample** is one where each and every member of the population has an equal chance of being studied. As you know from Chapter 4, researchers don't always study randomly selected observations, but random sampling is essential for the researcher to make inferences about the population based on sample data. Sometimes random samples, which typically have 1,000 to 1,500 observations, seem too small to faithfully represent their population parameters. Just how accurately does a sample statistic estimate a population parameter? The answer to this question lies at the heart of inferential statistics.

In common practice, the true value of a population parameter is unknown. We analyze sample statistics to estimate population parameters. When we analyze a random sample of the population instead of analyzing the entire population, the statistics we estimate will vary from the population parameters. In fact, in drawing a random sample, we are consciously introducing **random sampling error**. Random sampling error is defined as the extent to which a sample statistic differs, *by chance*, from a population parameter. Don't let the word "error" mislead you. Random sampling error isn't a mistake or oversight by the researcher; it's inherent in estimating a parameter from a sample. (Recall from Chapter 1 the difference between a random error and a systematic error.) If we are working with a random sample, the population parameter will be equal to the statistic we obtain from the sample, plus any random error that was introduced by taking the sample:

$$\text{Population parameter} = \text{Sample statistic} \pm \text{Random sampling error}$$

Consider student researchers who want a sample statistic that provides an unbiased estimate of a characteristic of all students at the university. The true value of the characteristic in the full student body is the population parameter. They eliminate selection bias by taking a random sample. But they know that random sampling error affects their estimate of the population parameter. Assume that the student researchers use a feeling thermometer scale to measure the sample's attitudes toward the Democratic Party. Having collected this information on each member of the sample, they calculate the mean rating of the Democratic Party. Because they are working with a random sample, the student pollsters know that the sample's mean Democratic rating is the same as the population's mean Democratic rating, plus (or minus) the random error introduced by taking the sample. Similarly, if the student researchers ask whether those in the sample voted in the last presidential election, the proportion of voters in the sample will equal the proportion of voters in the full student body plus random sampling error.

Some estimates of a population parameter are closer to the true value than other estimates are. The heart of inferential statistics is understanding random sampling error. The **standard error** of a statistic tells us how much we can expect a sample statistic to vary from the population parameter. What makes random sampling error a "better" kind of error than systematic error is that we have the statistical tools for figuring out how much a sample statistic is affected by random sampling error. We know how random sampling error affects a sample statistic, and we fully understand how to estimate its magnitude. A statistic's standard error tells us precisely how much noise to expect when we analyze a sample.

In addition to calculating the standard error of a statistic, which is a single number, we can also plot a statistic's **sampling distribution**, which shows how random sampling error shapes the expected values of sampling statistics. A sampling distribution is an abstract, statistical concept. It shows the expected distribution of sample statistics, like sample proportions or sample means. We don't observe a sampling distribution directly like we do sample data, but we can derive it from sample statistics. With a firm understanding of the random sampling error you should expect when analyzing data, you'll be able to confidently tell the difference between random error and meaningful statistics.

THE CENTRAL LIMIT THEOREM AND THE NORMAL DISTRIBUTION

When we estimate a population parameter with a sample statistic, we typically don't know how much our estimate deviates from the true population parameter. After all, if we knew the true population parameter, we wouldn't need to estimate it. How, then, do we know how much error is standard? To get a better sense of the effect of random sampling error on statistical estimates, we can conduct a thought experiment and simulate the outcome of this thought experiment using a computer.²

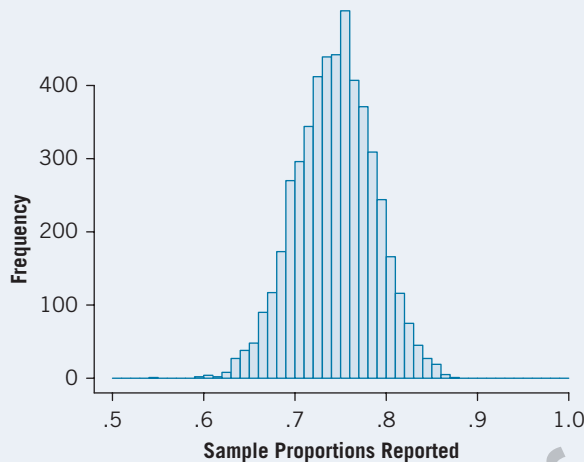
Imagine we knew the truth about a population but had researchers try over and over to estimate population parameters by conducting random samples from the population. We already know how much sample means and sample proportions would tend to deviate from the population parameter, but let's see how we got those numbers and what else we can learn about random sampling error.

Assume we knew the true proportion of students who voted in the last presidential election. The true population proportion equals .75. If researchers tried to estimate the proportion of student voters by taking random samples of 100 students over and over, what sample proportions would they report? We can use a computer to simulate 100 randomly sampled observations from a population with .75 Y (voters) and .25 N (nonvoters). Let's simulate a few samples and calculate the proportion of voters in each.

Sample 1	Sample 2	Sample 3
Y Y Y Y Y N Y Y N	Y Y Y Y Y Y Y Y Y N	Y Y Y Y Y N N N N Y
Y Y Y Y Y Y Y Y N	Y Y Y Y N N N N N Y	Y Y N Y Y Y Y N N
N N Y Y Y Y Y N Y Y	Y Y N Y N Y Y Y Y Y	Y Y Y Y Y Y Y N Y Y
Y N Y Y N Y Y N Y Y	Y Y Y N Y Y Y N N Y	Y N Y Y Y Y Y Y Y N
Y Y Y Y Y N Y N Y Y	N Y Y Y Y N Y N Y Y	N N Y N Y Y Y N Y Y
N Y N Y N N Y Y Y Y	Y Y Y Y N Y N Y N Y	Y N Y N Y Y Y Y Y Y
Y Y Y N Y Y Y Y Y Y	Y Y Y Y N Y Y Y Y Y	Y Y Y Y Y Y Y Y Y N
Y N Y Y N Y Y Y N Y	N Y Y N Y Y Y Y Y Y	N Y Y Y Y Y N N Y Y
N N Y N N Y Y Y N Y	Y Y N N Y N Y Y Y N	Y Y N Y Y Y Y Y Y Y
Y Y Y Y N Y N N Y Y	Y Y Y Y Y Y Y Y Y Y	Y Y Y Y Y Y Y Y N Y
"Yes" Proportion = .73	"Yes" Proportion = .76	"Yes" Proportion = .79

None of these first three simulated samples exactly estimate the true population parameter (.75) but that's not too surprising. Random sampling error adds noise to these estimates of the population parameter.

Figure 6-1 Distribution of Proportions from 5,000 Random Samples



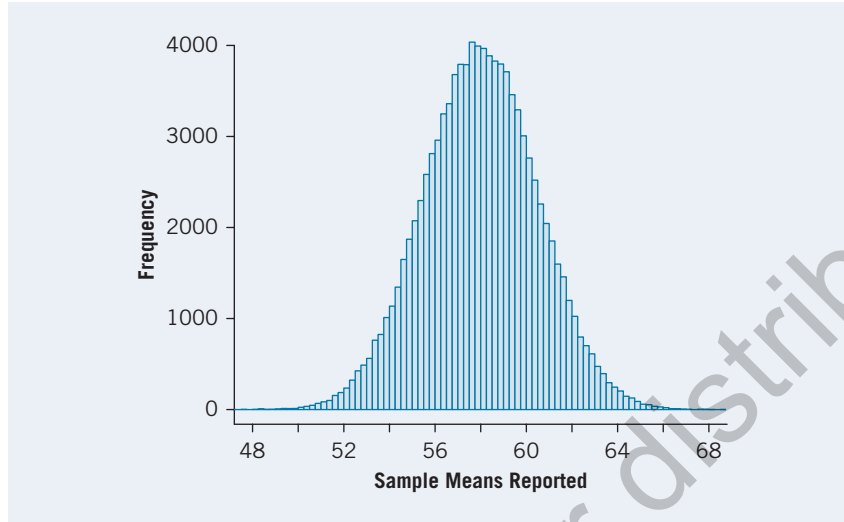
Keep the thought experiment going and imagine we had researchers randomly sample 100 students repeatedly and keep reporting the proportion of voters so we could plot the sample proportions. Since we're simulating this research with a computer (and don't need to pay our imaginary researchers), let's generate 5,000 sample proportions and create a histogram to view the distribution of sample statistics.

This figure shows a sampling distribution of the proportion of students who vote when $p = .75$ and $n = 100$. You can see that the sample proportion we're most likely to glean from a random sample is the true population proportion, $.75$, but the sample proportions vary because of random sampling error. The shape of this sampling distribution probably looks familiar to you—it's often called a bell curve—but think for a moment about how we produced this shape. We saved the results of many random samples of our $.75$ voting population and found that the sampling distribution had this shape. We didn't tell the computer to create a sampling distribution with this shape, it just works out that way.

Let's continue our thought experiment but change it up slightly to analyze sample means. Assume we know the distribution of feeling thermometer scores in the population. Imagine drawing an extremely large number of random samples of $n = 100$ from a population with a mean of 58 and a standard deviation equal to 24.8.³ Researchers draw a sample, report the mean thermometer rating, and return the cases to the population. Researchers draw another sample, report the sample mean, and return the cases to the population. They draw another sample, and another, and another—until hundreds of thousands of means from hundreds of thousands of samples of $n = 100$ are calculated and recorded. What would the distribution of all those sample means look like? Consider Figure 6-2, which shows the distribution of the means of 100,000 random samples drawn from the population.

The sampling distribution of these sample means should look familiar. Study Figure 6-2 for a moment. Notice some important aspects of this sampling distribution. It is centered on the true population mean of 58. In fact, if we were to calculate the overall mean of the 100,000 sample means represented in Figure 6-2, we would arrive at the population mean (within $.001$). The bulk of the sample means reside in

Figure 6-2 Distribution of Means from 100,000 Random Samples



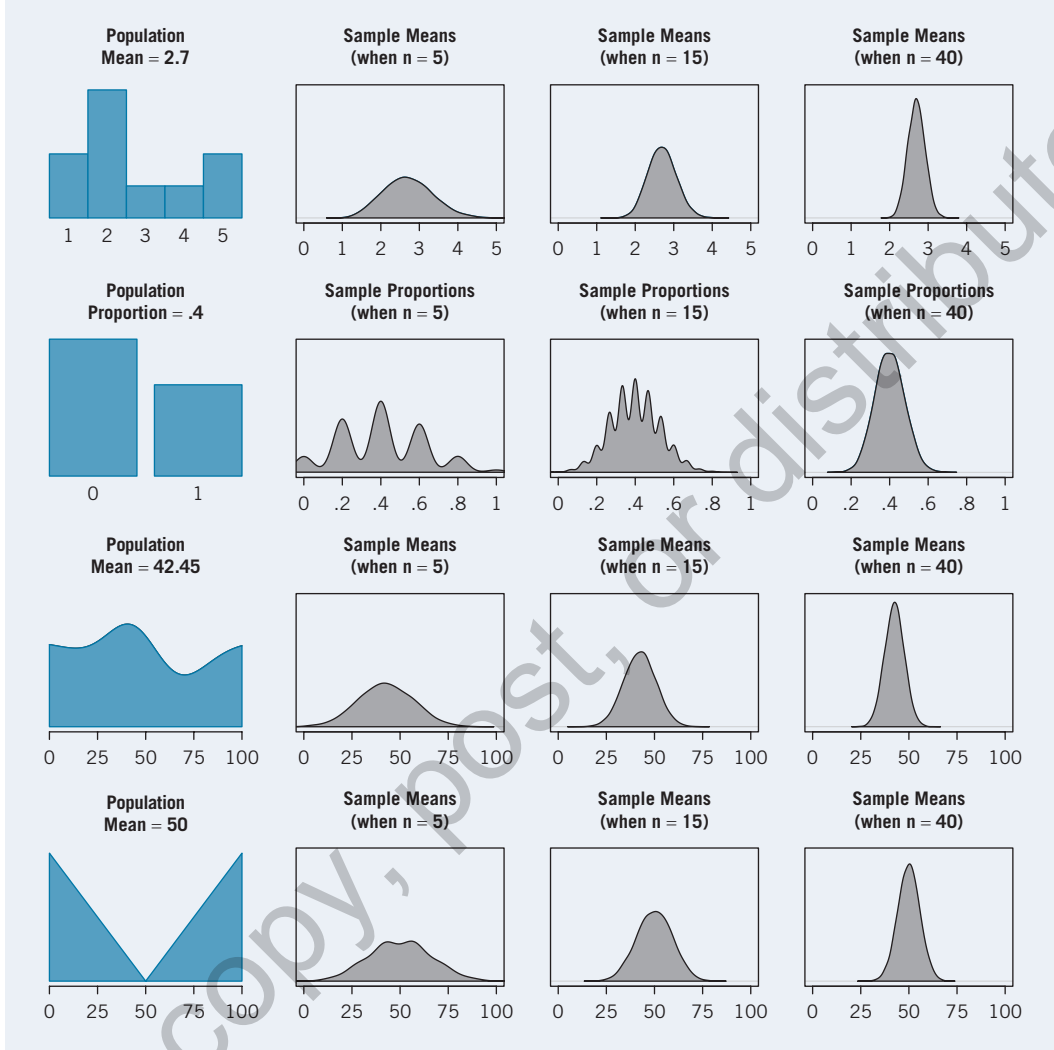
Note: The figure shows means from 100,000 samples of $n = 100$. Population parameters: $\mu = 58$ and $\sigma = 24.8$.

the thick part of the distribution around 58, between about 56 and 60. Even so, a few samples serve as very poor representations of the population mean ($\mu = 58$), returning values of less than about 53 or greater than 63.⁴ Some sample means are higher than the true population mean and others are lower. Despite some random “notchiness” here and there, the distribution has a symmetrical bell-like shape. This bell-like symmetry is the signature of the normal distribution. Again, we didn’t program the computer to create this shape, it just emerges when you repeatedly conduct random samples. Random sampling error produces a bell-shaped curve.

Figures 6-1 and 6-2 illustrate the central limit theorem. The **central limit theorem** is an established statistical rule that tells us that if we were to take an infinite number of samples of size n from a population of N members, the sample means will follow a normal distribution. This distribution of sample means, furthermore, would have a mean equal to the true population mean and have random sampling error equal to σ , the population standard deviation, divided by the square root of n . Thus, most random samples of $n = 100$ that are drawn from a population with a mean of 58 and a standard deviation of 24.8 will yield means that are equal to 58, give or take 2.5 or so. In fact, the normal distribution allows us to make precise inferences about the percentage of sample means that will fall within any given number of standard errors of the true population mean.

Here is what’s truly incredible about the central limit theorem: The expected sampling distribution will be a normal distribution regardless of the distribution of sample values. The distribution of variables in the sample can be split between two categories, right skewed, left skewed, multimodal, or uniform. The central limit theorem applies to proportions, as seen in Figure 6-1 and the second row of Figure 6-3, because the number of “yes” responses or 1s can be considered a sum of random numbers. As long as the sample size is sufficiently large, the sampling distribution follows a well-defined normal distribution. There should be at least 30 observations and, when the sample statistic is a proportion, at least 10 cases with each value of the

Figure 6-3 Illustration of the Central Limit Theorem



variable. When we visualize repeated sampling, as in Figures 6-1 and 6-2, we'll see some minor deviations and asymmetries but the more we repeatedly sample, the more our simulated sampling distributions take the ideal form.

You can see the central limit theorem in action in Figure 6-3. Each row of the figure shows a different distribution of variable values in a population; the first row, far-left panel, shows a hypothetical discrete variable with values 1 to 5 and a mean of 2.7. The second row, far-left panel, shows a .4 population proportion. The third and fourth rows show distributions of interval-level variables in the population with means of 42.25 and 50, respectively. If we estimate a mean or proportion using a very small sample, such as five, ten, or fifteen observations, the sampling distribution of our sample statistic does not follow a normal distribution. This is evident in the second and third columns of Figure 6-3. But in the fourth column, we can see that sampling distributions become normally distributed with a relatively small sample,

such as 40 observations, regardless of how the underlying variable is distributed in the population. Notice how the sampling distributions consolidate around the population means as the sample sizes increase from 5 to 15 and then to 40. The sampling distributions in the far-right column are all normal distributions, but you can see that their widths vary.

To this point, we've used the casual term "bell curve" to describe the sampling distributions of means and proportions to help to see the big picture without getting too caught up in technical details. We don't observe sampling distributions directly, but we can use computers to carry out the thought experiment of repeatedly sampling some population and see what sampling distributions would look like. We generally don't know what the probability distribution that generates the observed data looks like, if there even is one, so the central limit theorem is very useful. The central limit theorem tells us that sample means and proportions will be normally distributed around population means and proportions.

Let's take a closer look at what we mean when we say a statistic would have a normal distribution in repeated samples. The **normal distribution** is a probability distribution with a very specific shape. Normal distribution is defined by mean and standard deviation. Just like only some round shapes are perfect circles, many distributions can be casually described as bell shaped, but very few of them will conform to a normal distribution. When we say something follows a normal distribution, we are saying something very specific. A normally distributed variable follows the **empirical rule**. We know exactly how often a normally distributed variable falls within one, two, and three standard errors of its mean: 68, 95, and 99 percent of the time, respectively. As we shall see, the empirical rule is an important and convenient rule of thumb in inferential statistics.

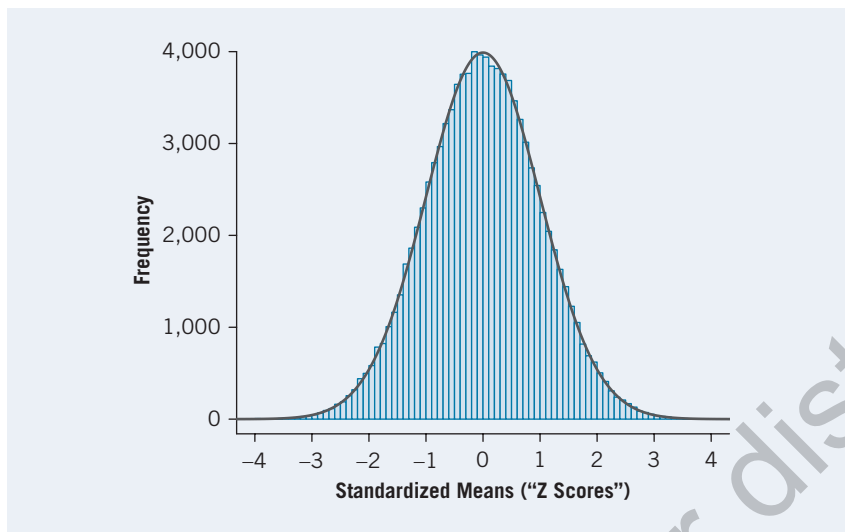
Examine Figure 6-4, which again displays the distribution of the 100,000 sample means. Figure 6-4 differs from Figure 6-2 in two respects. First, a line representing the normal curve has been drawn around the distribution, summarizing its shape. You can see how closely the normal curve fits the simulated sampling distribution. Second, the horizontal axis now shows standardized means, the *Z* scores of the reported means, and Figure 6-2's units of measurement (points on the Democratic feeling thermometer) are replaced in Figure 6-4 with a scale of single-digit numbers. The mean, which appeared as the raw value "58" in Figure 6-2, is labeled "0" in Figure 6-4. Values below the mean are negative, and values above the mean are positive.

The horizontal axis in Figure 6-4 is a standardized transformation of the axis in Figure 6-2. **Standardization** occurs when the numbers in a distribution are converted into standard units of deviation from the mean of the distribution. A value that is expressed in original form—points on a thermometer scale, years of age, dollars of income—is called a raw score or unstandardized score. A standardized value is called a *Z* value or ***Z* score**. To transform an unstandardized score into a *Z* score, you would divide the score's deviation from the expected value (i.e., the population mean) by the standard error of the sample statistic:

$$Z \text{ score} = \frac{\text{Deviation from expected value}}{\text{Standard error}}$$

The standard error, as its name suggests, is also a standard unit of deviation from the mean. Consider three sample means: 54, 58, and 60. Given a standard error equal to 2.5, a sample mean of 54 has a *Z* score equal to $(54 - 58)/2.5 = -1.6$, or 1.6 standard errors below the true population mean of 58. So, a sample mean of 54

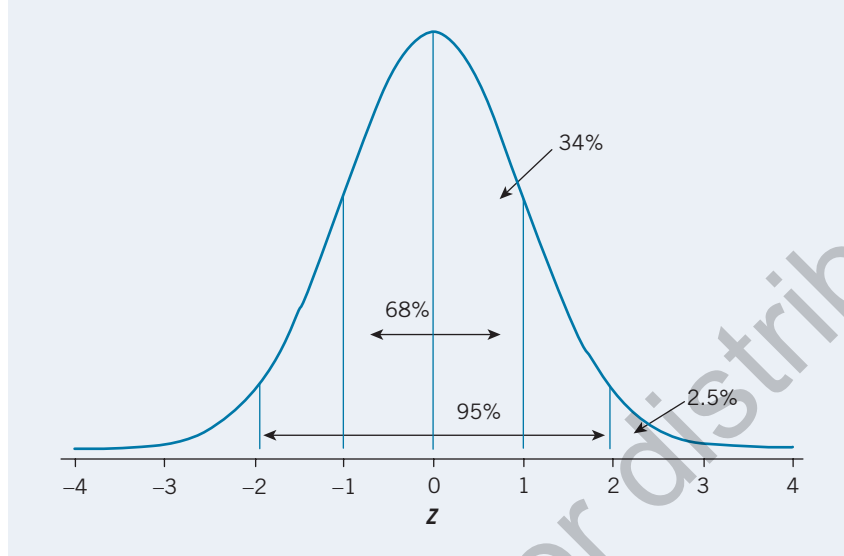
Figure 6-4 Raw Values Converted to Z Scores



in Figure 6-2 appears as a Z score of -1.6 in Figure 6-4. A sample mean of 58 in Figure 6-2 hits the population mean right on the money, so in Figure 6-4 it has a Z score equal to 0: $(58 - 58)/2.5 = 0$. The sample mean of 60 converts into a Z score of $(60 - 58)/2.5 = 2.8$, very close to a Z score of 1 in Figure 6-4.

When we convert sample statistics to Z scores, we standardize the expected sampling distribution. Why go through the ritualistic procedure of converting an untransformed score into a Z score? Because Z scores are the key to the inferential leverage of the normal distribution. Figure 6-5 again presents the bell-shaped curve, this time displaying percentages inside the curve. The arrow stretching between $Z = -1$ and $Z = +1$ bears the label “68%.” What does this mean? It means this: If a variable is normally distributed, then 68 percent of the cases in the distribution will have Z scores between -1 (one standard unit below the mean) and $+1$ (one standard unit above the mean). Moreover, since the curve is perfectly symmetrical, half of that 68 percent—34 percent of the cases—will fall between the mean (Z equal to 0) and a Z score of $+1$, and the other 34 percent will fall between the mean and a Z score of -1 . So the range between $Z = -1$ and $Z = +1$ is the fattest and tallest part of the curve, containing over two-thirds of the cases. Notice the arrow labeled “95%,” the one stretching between $Z = -1.96$ and $Z = +1.96$. These numbers tell us that, in a normal distribution, 95 percent of the cases will have Z scores in the long interval between 1.96 standard units below the mean and 1.96 standard units above the mean. This long interval, in other words, will contain just about all the cases. But, of course, 5 percent of the cases—those with Z scores of less than -1.96 or greater than $+1.96$ —will lie outside this interval, in the sparsely populated tails of the distribution. Again, since the curve is symmetrical, half of this 5 percent, or 2.5 percent, will fall in the region below $Z = -1.96$, and the other 2.5 percent will fall in the region above $Z = +1.96$. As you can see, the density (height) of the standard normal distribution is very close to 0 for Z scores less than -3 and scores greater than 3. Technically, the left and right tails approach, yet never reach 0, but the density outside the range of Figure 6-5 is microscopic.

Figure 6-5 Areas under the Standard Normal Curve



The **standard normal distribution** has a mean equal to 0 and a standard deviation of 1. There are many normal distributions, but there is only one standard normal distribution. It's a beauty to behold, for sure, and our appreciation only grows when we think about how we came upon it and what it represents. When we conducted our thought experiment of repeatedly sampling some statistic, we didn't tell the computer we wanted the statistic to be normally distributed; it just happens naturally, and it happens no matter how the values are distributed in the population (see Figure 6-3). The normal distribution represents noise and unpredictable variation in measurement; its well-defined shape shows us, quite remarkably, that there is order in randomness. It's quite a sight to see. Table 6-1 points out some of the standard normal distribution's critical features.

To make sure you understand what the critical values listed in Table 6-1 tell us, put one finger on $Z = -1.64$ on the x -axis of Figure 6-5 and another finger on $Z = +1.64$. Ninety percent of the area of the standard normal distribution lies between your fingers and 10 percent of the probability distribution lies outside your fingers in the right and left tails of the Figure 6-5. Move your fingers to $Z = -1.96$ and $Z = +1.96$. Now, 95 percent of the area is between your fingers and 5 percent is more than that, in the right and left tails. (The empirical rule that 95 percent of a normal distribution lies within 2 standard deviations of the mean marked on Figure 6-5 is a rule of thumb; 1.96 is close to 2.) The ± 1.96 interval is enormously important, as you'll soon discover. If you slide your fingers out to $Z = -2.58$ and $Z = +2.58$, you've now got 99 percent of the standard normal distribution's area between your fingers; only 1 percent of the distribution is more extreme than ± 2.58 . Stretch your fingers out to $Z = -3.29$ and $Z = +3.29$ and the extreme tails are now microscopic; only one-tenth of 1 percent (.1 percent) of the distribution is outside your fingers. For further detail, see Box 7-1 in Chapter 7.

Sometimes, researchers will refer to the **cumulative density** of a probability distribution. The cumulative density of a standard normal distribution equals the proportion (or percentage) of the area under the curve less than a given Z score.

Table 6-1 Critical Z Scores for Making Inferences

Value of Z	Percentage of Standard Normal Beyond
± 1.64	10%
± 1.96	5%
± 2.58	1%
± 3.29	.1%

For example, the cumulative density at -2 is about .05; 5 percent of the area of the standard normal distribution lies in the left tail from negative infinity to -2 . The cumulative density at 0 is .50 because .50 (or 50 percent) of the area is less than 0. What's the cumulative density at 1? It's .84 (the 68 percent of the area between -1 and 1 plus the 16 percent between minus infinity and -1). Sometimes researchers and statistical software will report cumulative densities, so it's essential to get familiar with the standard normal distribution.

QUANTIFYING STANDARD ERRORS

In this section, we closely examine the expected sampling distributions of two of the most common statistics we estimate from samples: means and proportions. When we calculate the mean of an interval-level variable in a sample to estimate a population mean, how much random sampling error should we expect? When we find the proportion of observations with a certain value of a nominal- or ordinal-level variable in a sample, how well does this statistic reflect the population parameter?

Three factors determine how well a sample statistic reflects a population parameter: sample selection method, variation in population, and sample size. We'll discuss these three factors generally to help you see the big picture and then we'll apply the second and third factors to means and proportions.

The first factor that determines how well a statistic estimates a parameter, selection method, was discussed in Chapter 4 when we discussed selecting observations to test hypotheses. When a researcher analyzes a random sample of the population, sample statistics will be unbiased estimates of population parameters. If observations are not selected randomly, sample statistics will systematically vary from population parameters no matter how large the sample is. Researchers have some statistical methods that correct for nonrandom sampling and mitigate the potential for bias, such as weighting observations, but these remedial measures require fairly extensive information about the sample observations. An online poll conducted on Fox News.com or CNN.com won't generate an accurate estimate of the general public's political views no matter how many of those website visitors respond to the poll.

Variation in the population characteristic has a direct relationship with random sampling error: As variation goes up, random sampling error goes up. Recall from Chapter 2 that variables have more or less dispersion. When a variable is measured at the interval level, we can calculate its standard deviation, which tells us how much its observed values deviate from its mean value. When a variable is measured at the nominal or ordinal level, we don't calculate its dispersion but we say dispersion is maximized when observations are evenly divided among possible

values. For example, if the possible responses to a survey question are “yes” and “no,” the greatest possible variation is an evenly split response: 50 percent yes and 50 percent no. The variance component of standard error reflects how homogeneous the sample is with respect to the characteristic being measured. If people share the same opinion, you don’t need to sample many people to obtain an accurate estimate of the population’s opinion. If *everyone* gave the same response to a yes or no survey question, you’d only need to survey one person to estimate the population parameter.

Finally, sample size has an inverse relationship with random sampling error. As the sample size goes up, random sampling error goes down. With a larger sample, we can be more confident in our estimates of population parameters. As we’ll see, however, the relationship between sample size and random sampling error is curvilinear; the marginal benefit of increasing sample size diminishes, so it takes larger and larger sample sizes to reduce random sampling error.

Adopting a conventional notation in which sample size is denoted by a lowercase n , we would have to say that a sample of $n = 400$ is preferable to a sample of $n = 100$, since the larger sample would provide a more accurate picture of what we are after.⁵ However, the inverse relationship between sample size and sampling error is nonlinear. Even though the larger sample is four times the size of the smaller one, going from $n = 100$ to $n = 400$ delivers only a twofold reduction in random sampling error. In ordinary language, if you wish to cut random error in half, you must quadruple the sample size. In mathematical language, the sample size component of random sampling error is equal to the square root of the sample size, \sqrt{n} .

Let’s pause and review the statistical components discussed thus far.

Sample size component: As the sample size goes up, random sampling error declines as a function of the square root of the sample size.

Variation component: As variation goes up, random sampling error increases in direct relation to the population’s standard deviation.

These factors—the variation component and the sample size component—are not separate and independent. Rather, they work together, in a partnership of sorts, in determining the size of random sampling error. This partnership can be defined by using ideas and terminology that we have already discussed:

$$\text{Random sampling error} = \frac{\text{Variation component}}{\text{Sample size component}}$$

Before exploring the exact properties of this general formula for random sampling error, consider its intuitive appeal. Notice that “Variation component” is the numerator. This reflects its direct relationship with random sampling error. “Sample size component” is the denominator, depicting its inverse relationship with random sampling error. Both components, the variation component and the sample size component, have known properties that give the researcher a good idea of just how much random sampling error is contained in a sample statistic.

Standard Error of a Sample Mean

Let’s apply these general principles of random sampling error to sample means. The variation component of the random sampling error of a mean is defined statistically by a measure you may have encountered before: standard deviation (symbolized as σ).

As discussed in Chapter 2 (see “Interval Variables” on p. 46), a variable’s standard deviation tells you how much the variable’s values vary from its mean value. When you calculate a sample mean to estimate the population mean, how much random sampling error can you expect? The standard error of a sample mean is directly related to the variable’s variation (signified by σ , the population standard deviation) and inversely related to the square root of the sample size (n):

$$\text{Standard error of sample mean} = \frac{\sigma}{\sqrt{n}}$$

In this chapter we have been using the generic term *random sampling error* to describe the error introduced when a random sample is drawn. The size of this error, as we have just seen, is determined by dividing the variation component by the sample size component. However, when researchers are describing the random sampling error associated with a sample statistic, they do not ordinarily use the term *random sampling error*. Rather, they refer to the standard error of the mean. Computer analysis programs routinely calculate standard errors for mean estimates, and political researchers always report the standard errors for the sample estimates they publish in quantitative research articles. Let’s be clear. The terms *standard error of a sample mean* and *random sampling error of a sample mean* are synonymous. Both terms refer to the bedrock foundation of inferential statistics. But because you will often encounter the term *standard error*, this book will use the term, too.

How can we use the standard error of mean formula to figure out how closely a sample mean matches the population mean? Assume we record values of an interval-level feeling thermometer measurement from a random sample of $n = 100$ and find that the sample mean equals 58 and the sample standard deviation equals 24.8. Our best estimate of the population mean is 58, but we can expect our sample mean to vary from the population parameter due to random sampling error. According to the formula for the standard error of a sample mean, the magnitude of the standard error of this sample mean is equal to:

$$\text{Standard error of mean} = \frac{24.8}{\sqrt{100}} = \frac{24.8}{10} = 2.48$$

Based on the random sample, we’d estimate the mean value of the feeling thermometer in the population as 58 (this is sometimes called the point estimate to help visualize the result) but we could be pretty sure the random “noise” of this estimate is plus or minus 2.48 degrees.

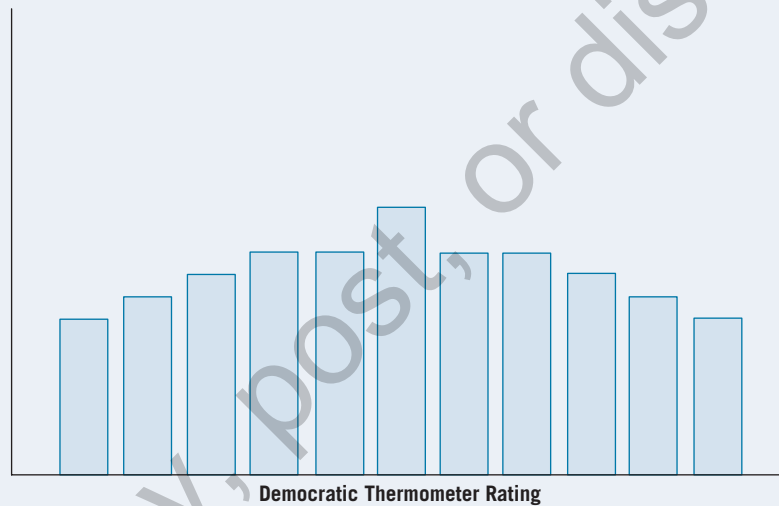
Take another look at Figure 6-2 that shows a simulated distribution of sample means. How much deviation from the population mean of 58 looks typical? It might help to imagine folding the left side of the distribution onto the right side. Can you see how much noise there is in this estimate? The standard error of the sample means equals the standard deviation of the feeling thermometer divided by the square root of the sample size, σ/\sqrt{n} , which equals 2.48 in this example ($\sigma = 24.8$ and $n = 100$). We could have approximated the answer from our simulated sampling distribution, but the standard error of mean formula gives us an exact answer.

Consider how the two components of the random sampling error formula, the variance component and the sample size component, work together. Figure 6-6 depicts two possible ways that student ratings of the Democrats might be distributed within the student population. First, suppose that in the general population of students there is a great deal of variation in ratings of the Democratic Party, as in

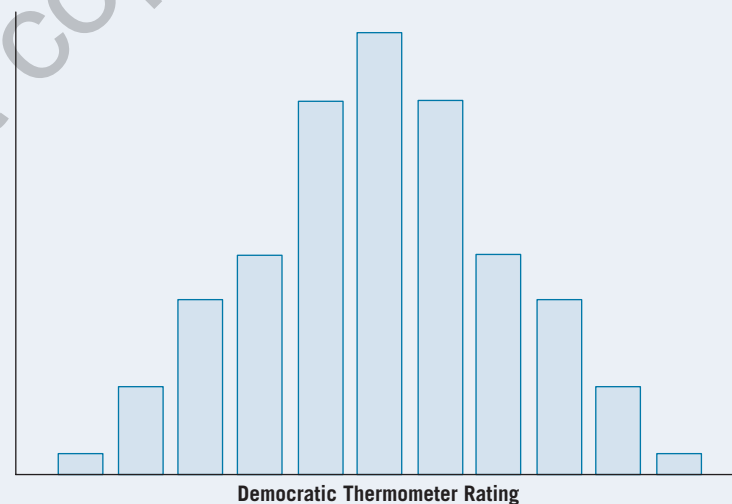
panel A of Figure 6-6. There are appreciable numbers of students in every range of the rating scale, from lower to higher, with only a slight amount of clustering around the center of the distribution. Large numbers of students dislike the Democrats and give them ratings between 0 and 40. Many students like the Democrats and give them ratings between 60 and 100. In this scenario, the population parameter the student researchers wish to estimate, Democratic Party thermometer ratings, would have a large variation component. Since variation in the population characteristic is high, the variation component of random sampling error is high. A random sample taken from the population would produce a sample mean that may or may not be

Figure 6-6 High Variation and Low Variation in a Population

A. High Variation in Democratic Thermometer Ratings in the Student Population



B. Low Variation in Democratic Thermometer Ratings in the Student Population



close to the population mean—it all depends on which cases were randomly selected. Because each student has an equal chance of being chosen for the sample, one sample might pick up a few more students who reside in the upper range of the distribution. Another sample from the same population might randomly choose a few more students from the lower range.

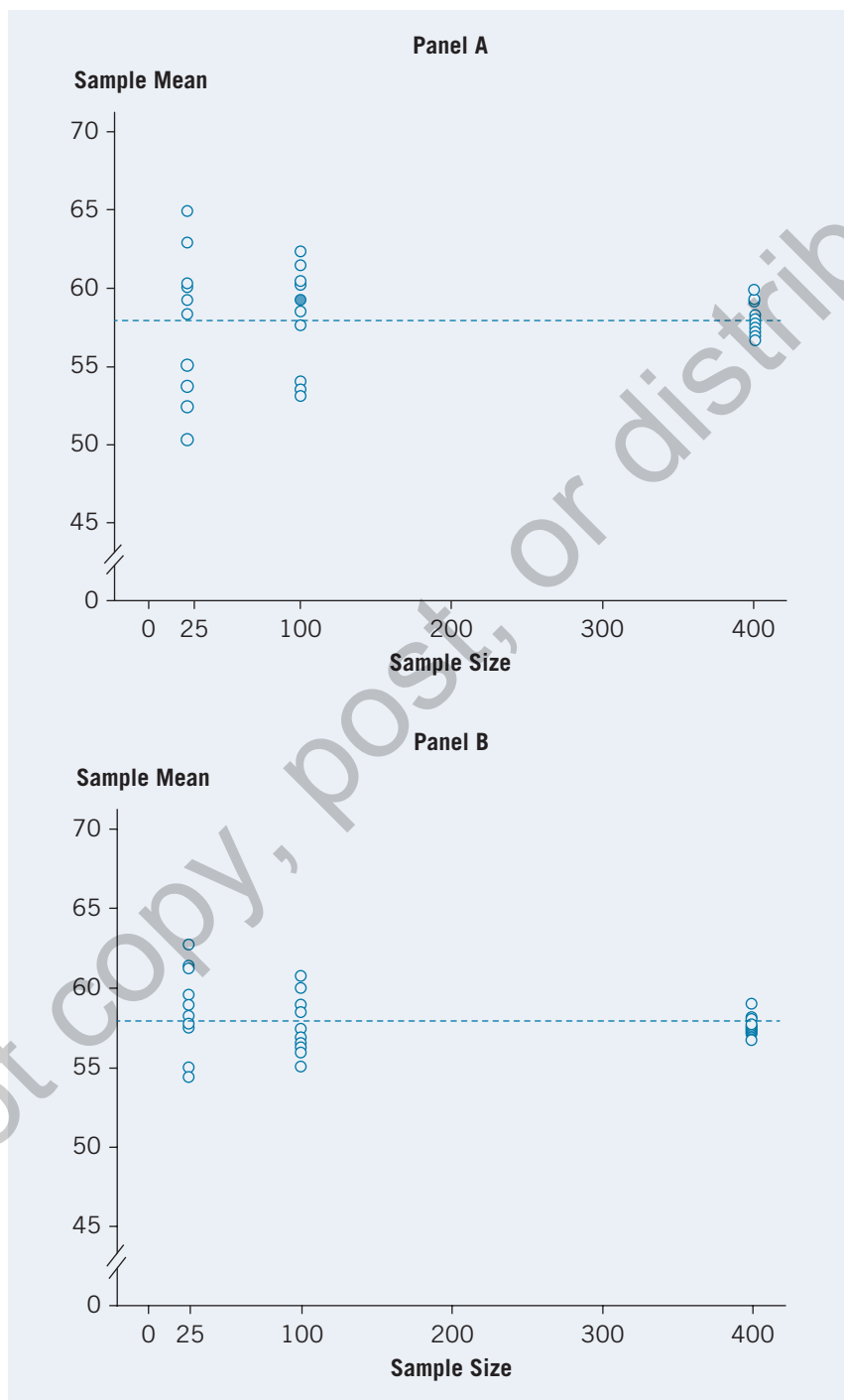
In contrast, visualize a general student population like the one depicted in panel B of Figure 6-6. Notice that the ratings are clustered around a well-defined center, with fewer cases at the extremes of the scale. Since variation in the population characteristic is low, the variation component of random sampling error is low. A random sample taken from the population would produce a sample mean that is close to the population mean. What is more, repeated sampling from the same population would produce sample mean after sample mean that are close to the population mean—and close to each other.

The population means are the same in both scenarios, a mean rating of 58. However, student ratings are more spread out in population A—the distribution has a higher standard deviation—than in population B. The distribution of Democratic Party feeling thermometer ratings in population A has a mean equal to 58 and a standard deviation equal to 24.8. The distribution of Democratic Party feeling thermometer ratings in population B has a mean equal to 58 and a standard deviation equal to 17.8. Having artificially created these realistic populations, we can ask the computer to draw random samples of different sizes from each population.⁶ We can then calculate and record the mean Democratic rating obtained from each sample. As we will demonstrate, a statistic computed on a random sample from population A will have a higher amount of random sampling error than will a statistic computed on a random sample drawn from population B.

The results of this simulation are presented in Figure 6-7. All the sample means displayed in panel A are based on the same student population—a population in which $\mu = 58$ and $\sigma = 24.8$. All the sample means displayed in panel B were drawn from a student population in which $\mu = 58$ and $\sigma = 17.8$. The dashed horizontal line in each panel shows the location of the true population mean, the parameter being estimated by the sample means. For each population, the computer drew ten random samples of $n = 25$, ten random samples of $n = 100$, and ten random samples of $n = 400$. So, by scanning from left to right within each panel, you can see the effect of sample size on random sampling error. By moving between panel A and panel B, you can see the effect of the standard deviation on random sampling error. (So that we don't lose track of the initial sample mean of 59, it appears as a solid dot in the $n = 100$ group in panel A.)

Consider the set of sample means with the largest error component, the samples of $n = 25$ in panel A. Even though three or four of these sample means come fairly close to the population mean of 58, most are wide off the mark, ranging in value from the chilly (a mean Democratic rating of 50) to the balmy (a mean rating of 65). A small sample size, combined with a dispersed population parameter, equals a lot of random error. As we move across panel A to the ten sample means based on $n = 100$, we get a tighter grouping and less wildness, but even here the means range from about 53 to 62. The samples of $n = 400$ return much better precision. Four of the ten sample means hit the population mean almost exactly. Plainly enough, as sample size increases, error declines. By comparing panel A with panel B, we can see the effect of the population standard deviation on random sampling error. For example, notice that the ten samples of $n = 25$ in panel B generate sample statistics that are about as accurate as those produced by the samples of $n = 100$ in panel A. When less dispersion exists in the population parameter, a smaller sample can sometimes yield relatively accurate statistics. Naturally, just as in panel A, increases in sample size bring the true population mean into clearer focus. At $n = 400$ in panel B,

Figure 6-7 Sample Means from High Variation and Low Variation Populations



Note: Hypothetical data. The hypothetical sample mean of 59 is represented by the solid dot in the $n = 100$ group in panel A. The dashed horizontal line shows the location of the true population mean ($\mu = 58$).

six of the ten sample means are within a few tenths of a point of the true population mean. A larger sample, combined with lower dispersion, equals less random error and greater confidence in a sample statistic.

Standard Error of a Sample Proportion

The general principles of random sampling error also apply to sample proportions. Many of the variables of interest to political researchers are not measured at the interval level. Rather, they are measured at the nominal or ordinal level. When you analyze a variable measured at the nominal or interval level, you can describe its distribution of values with a frequency distribution table. You can calculate the proportion of observations that have each value of the variable, but you cannot calculate the variable's mean or standard deviation.⁷ The variation component of the standard error of proportion formula reflects the lower level of measurement. The proportion of a sample falling into one category of a nominal or ordinal variable is denoted by the ordinary letter p . Given a proportion p , the variation component equals the square root of the product of p and $1 - p$. The standard error of a sample proportion is directly related to this variance component and inversely related to the square root of the sample size (n):

$$\text{Standard error of sample proportion} = \frac{\sqrt{p \times (1-p)}}{\sqrt{n}}$$

Consider a random sample of whether college students ($n = 100$) voted in the last presidential election. Suppose that 72 of those surveyed answered, "Yes, I voted in the last presidential election," and the remaining 28 answered, "No, I did not vote in the last presidential election." What is the sample proportion of voters?

$$\text{Sample proportion of voters} = \frac{\text{Number answering "yes"}}{\text{Sample size}} = \frac{72}{100} = .72$$

The sample proportion of voters, p , is .72. This is a descriptive sample statistic. Based on this sample, our best estimate of voting in the general student population is .72, but we know sample statistics have random sampling error. If the proportion of voters is .72, the proportion of nonvoters must be .28 (the answer choices were simply yes and no). How closely does the sample proportion, .72, estimate the proportion of voters in the student population? What is the standard error of the observed sample statistic, .72?

In this example, the proportion of students falling into the "Yes, I voted in the last presidential election" category is .72, so $p = .72$. The proportion of a sample falling into all other categories of the variable (in this case, the "No" responses) would be equal to $1 - .72$, or .28. Sometimes, the value of $1 - p$ is denoted by the letter q . This proportion, q , is equal to one minus p , or $q = 1 - p$. This information—the value of p , the value of q , and the sample size—permits us to estimate the standard error of the sample proportion, p .

$$\text{Standard error of proportion} = \frac{\sqrt{p \times (1-p)}}{\sqrt{n}} = \frac{\sqrt{.72 \times .28}}{\sqrt{100}} = \frac{.45}{100} = .045$$

Based on this sample, we'd estimate the proportion of voters in the student population to be .72, but we would recognize the standard error of this estimate

is .045. This is equivalent to estimating that 72 percent of students vote with a standard error of 4.5 percent. The standard error of proportion formula tells how much noise there is in our estimate of the population proportion.

Based on the sample described in this section, we found a .72 sample proportion with a .045 standard error. What if we instead estimated the proportion of students who don't vote? We've already seen that the proportion of nonvoters must be .28. How much random sampling error can we expect in our estimate of student nonvoters? If the sample proportion, p , we're interested in is .28, then $1 - p$ equals .72. The numerator of the standard error of proportion formula is $\sqrt{.28 \times .72}$, which is equal to $\sqrt{.72 \times .28}$. So the standard error of the proportion of nonvoters is also .045. Whenever there are two values of the variable, like yes and no, the proportion estimated in each category will have the same standard error.

There is a final statistical caveat. We have just demonstrated how the central limit theorem and normal distribution can be applied in determining the standard error of a sample proportion. Under most circumstances, this method works quite well. However, we know that normal estimation works best for sample proportions closer to .50, and it begins to lose its applicability as p approaches .00 or 1.00. How can one know if normal estimation may be used? Here is a general procedure. Multiply p by the sample size, and multiply q by the sample size. If both numbers are 10 or higher, then normal estimation will work fine. (Actually, if both numbers are 5 or more, normal estimation will still work.) The researchers in this example are on solid inferential ground, since $100 \times .72 = 72$ and $100 \times .28 = 28$. Trouble could begin to brew, however, if the researchers were to subset their sample into smaller groups—subdividing, say, on the basis of gender or class rank—and were then to make inferences from these smaller subsamples.

CONFIDENCE INTERVALS

When a researcher uses a sample statistic to estimate a population parameter, the researcher should expect the statistic to deviate from the parameter by an amount known as standard error. Because of the random probability of obtaining a sample statistic that departs dramatically from the population parameter, statisticians never talk about certainty. They talk, instead, about confidence and probability. The normal distribution and standard errors play an essential role in making probabilistic inferences. A **probability** is defined as the likelihood of the occurrence of an event or set of events.

The researcher's most common tool for expressing how much confidence he or she has that a sample statistic reflects the true population parameter is the **95 percent confidence interval** (95% CI). To understand it, recall the thought experiment where you know the population parameter and have researchers repeatedly try to estimate the parameter from random samples. The 95% CI is defined as the interval around a sample statistic that would contain the true population parameter in 95 percent of repeated samples. Less frequently, researchers use 90% and 99% CIs, which would be the intervals around sample statistics that would contain the true population parameters in 90 and 99 percent of repeated samples, respectively.

A 95% CI consists of two numbers: the interval's lower boundary and the interval's upper boundary. Thanks to the central limit theorem, we know that the boundaries of the 95% CI are defined by the sample statistic minus 1.96 standard errors at the lower end and the sample statistic plus 1.96 standard errors at the upper end. Therefore, the sample statistic is always exactly the midpoint of the 95% CI.

$$95\% \text{ CI lower boundary} = \text{Sample statistic} - (1.96 \times \text{Standard error})$$

$$95\% \text{ CI upper boundary} = \text{Sample statistic} + (1.96 \times \text{Standard error})$$

These formulas for finding the lower and upper boundaries of a 95% CI work for sample proportions, regardless of sample size. They work for sample means too, provided the sample size is large. When we estimate sample means from small samples, some adjustments that we discuss in the last section in this chapter are necessary.

Let's assume again that researchers estimate the proportion of students who voted in the last presidential election from a random sample of $n = 100$ and get a sample proportion of .80. We already know how to calculate the standard error of this estimate: $\sqrt{.8 \times .2 \div 100} = \sqrt{.0016} = .03$. Let's use this information to find the 95% CI for the sample proportion.

$$\begin{aligned} 95\% \text{ CI lower boundary} &= .80 - (1.96 \times .04) \\ &= .80 - .08 \\ &= .72 \end{aligned}$$

$$\begin{aligned} 95\% \text{ CI upper boundary} &= .80 + (1.96 \times .04) \\ &= .80 + .08 \\ &= .88 \end{aligned}$$

The 95% CI in this example is [.72, .88]. We can report the CI's upper and lower boundaries as a pair of values in square brackets, separated by a comma. Square brackets signify that the interval includes the boundaries, whereas rounded parentheses signify that the boundaries are just outside the CI.

The 95 percent standard of confidence is widely applied, but its precise boundaries can be a bit tedious to calculate quickly. Therefore, it is customary to round off 1.96 to 2.0, resulting in a useful shortcut. To find the 95% CI for a sample statistic, multiply the standard error by 2. Subtract this number from the sample statistic to find the lower confidence boundary. Add this number to the sample statistic to find the upper confidence boundary. Applying this shortcut, the 95% CI for the proportion of student voters is .80 plus/minus 0.08 (which is 2 times .04). This yields the same confidence interval as before, rounding to two decimal places.⁸

We would use the same approach to find the 90% CI or the 99% CI of the proportion of student voters in this $n = 100$ sample. The only difference is rather than marking boundaries at 1.96 standard errors around the sample statistic (2 standard errors if we're using the rule of thumb), the 90% CI is the sample mean ± 1.64 standard errors and the 99% CI is the sample mean ± 2.58 standard errors. With a sample proportion of .80 and a standard error equal to .04, the 90% CI would be [.73, .87] and the 99% CI would be [.70, .90]. The 90% CI is narrower than its 95 percent counterpart because it demands less confidence that the interval contains the population parameter; the 99 percent CI is wider than the 95% CI because it demands more confidence. To summarize:

$$90\% \text{ CI} = \text{Sample statistic} \pm (1.64 \times \text{Standard error})$$

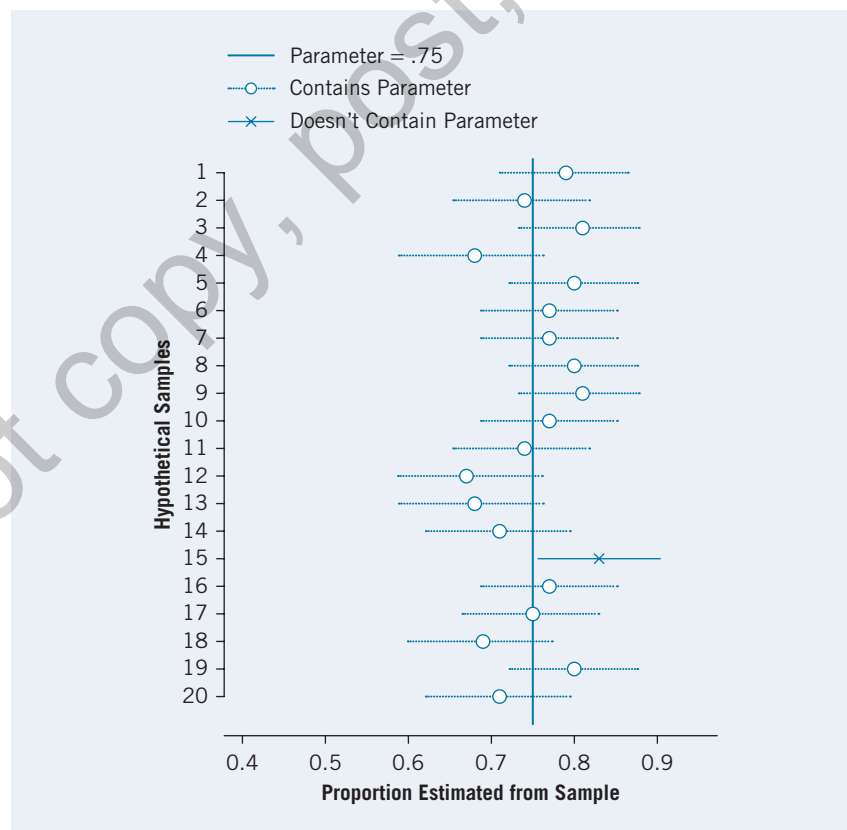
$$95\% \text{ CI} = \text{Sample statistic} \pm (1.96 \times \text{Standard error})$$

$$99\% \text{ CI} = \text{Sample statistic} \pm (2.48 \times \text{Standard error})$$

What exactly do confidence intervals tell us? This has been a source of philosophical debate among political methodologists. As a general proposition, there is a 95 percent probability that a 95% CI contains the true population parameter. However, when you estimate a sample statistic and find the lower and upper boundaries of its 95% CI, that specific interval either contains that true population parameter or it does not. Imagine being blindfolded and randomly picking a mean from Figure 6-1's distribution of sample proportions. There's a 95 percent probability that a 95% CI around the proportion you pick will contain the true population proportion (which we said is .75), but once you make your selection, the probability that proportion's 95% CI contains the true population parameter is either 0 or 1. To see this distinction more clearly, consider a series of twenty hypothetical 95% CIs (Figure 6-8) produced from a population proportion of .75 and sample size $n = 100$.

Each horizontal line segment in Figure 6-8 shows a sample proportion and its corresponding 95% CI. As you can see, over repeated samples, 95 percent of these confidence intervals contain the true population parameter—in this case, the population proportion .75 represented by the vertical gray line. The fifteenth sample's CI misses the true value completely, however; the probability that this fifteenth sample's CI contains the population parameter is 0. When you conduct analysis using sample data, you don't know whether the CI for your sample statistic contains the population parameter, but you can state that 95% CIs have a .95 probability of containing the population parameter.

Figure 6-8 95 Percent Confidence Intervals from Twenty Samples



SAMPLE SIZE AND THE MARGIN OF ERROR OF A POLL

As consumers of popular media, we are more likely to encounter percentages or proportions than arithmetic means. Sometimes, of course, the sampling procedures used by media-based organizations are questionable, and their reported results should be consumed with a grain (or more) of salt. Reputable pollsters always report some measure of random sampling error (normally following the form “margin of error $\pm \dots$ ”). The margin of error of a poll is an alternative way of communicating the boundaries of the 95% CI. You are now well equipped to interpret such percentages when you encounter them in news stories. Understanding the relationship between a poll’s sample size and its margin of error is also useful for planning survey research.

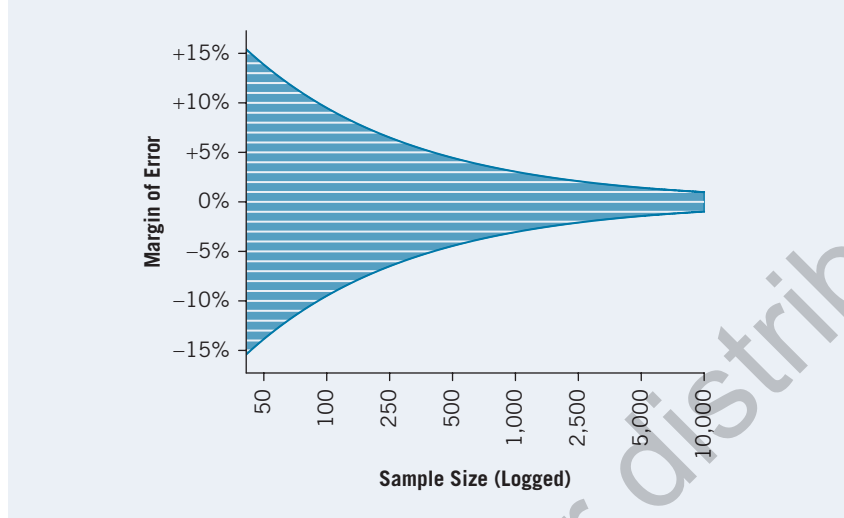
Because of the nonlinear relationship between sample size and random sampling error, samples that seem rather small nonetheless carry an acceptable amount of random error. Consider three samples: $n = 400$, $n = 1,600$, and $n = 2,500$. The sample size component of the smallest sample size is the square root of 400, which is equal to 20. So, for a sample of this size, we would calculate random sampling error by dividing the variation component by 20. Random sampling error for the next sample would be the variation component divided by the square root of 1,600, which is equal to 40. So, by going from a sample size of 400 to a sample size of 1,600, we can cut random sampling in half, from $\sqrt{pq}/20$ to $\sqrt{pq}/40$. Thus, if resources permit, obtaining a sample of $n = 1,600$ would be a smart move. Random sampling error for the largest sample would be equal to the variation component divided by the square root of 2,500, which is equal to $\sqrt{pq}/50$. Boosting the sample size by 900 cases—from 1,600 to 2,500—occasions a modest increase in the sample size component, from 40 to 50. The benefits of increasing sample size are subject to diminishing returns; each person added to the sample reduces random sampling error less than the last person did.

In order to use the standard error of proportion formula to determine how large a sample must be to achieve a desired margin of error, the researcher must make an educated guess about the expected sample proportion. The sample proportion (p) and its complement (q) appear in the numerator of the standard error of proportion formula, \sqrt{pq}/\sqrt{n} . To be on the safe side, one can assume that public opinion on the poll question will be as varied as can be, with $p = .5$ and $q = .5$. Recall from Chapter 2 that a nominal or ordinal-level variable has high dispersion when each value has the same number of cases. (The value of \sqrt{pq} is also greatest when p and q equal $.5$.) If one makes this assumption for survey planning purposes, the margin of error of the poll (in proportion form) will be $\pm 1.96 \times \sqrt{.25}/\sqrt{n}$. This equation can be simplified to find the required sample size for a desired margin of error:

$$\text{Sample size needed} = \frac{.98}{(\text{Margin of error})^2}$$

This equation is sometimes further simplified as $n = 1/ME^2$, where ME is the desired margin of error expressed as a proportion. If the researcher wants to ensure a poll will yield no worse than a margin of error of $\pm 2\%$, the researcher needs a sample with 2,450 respondents ($n = .98/.02^2$). Figure 6-9 provides a graphic representation of the relationship between sample size and the margin of error of a poll. Note that the x -axis values are displayed on a log base-10 scale (for a definition and discussion of logarithms see Chapter 9).

Figure 6-9 Sample Size and Percentage Margin of Error



Sophisticated sampling is an expensive undertaking, and survey designers must balance the cost of drawing larger samples against the payoff in precision. For this reason, most of the surveys you see and read about have sample sizes in the 1,500 to 2,000 range, an acceptable comfort range for estimating a population parameter. Samples this size strike a reasonable balance between precision and cost. Samples this large can estimate population parameters with a 3 percent margin of error with a reasonable expenditure of time and money. At the same time, there is no rule setting minimum and maximum sample sizes. Internet services have dramatically decreased the cost of conducting polls, so we might expect to see larger samples with lower margins of error become the new standard.

INFERENCE WITH SMALL BATCHES: THE STUDENT'S *t*-DISTRIBUTION

In this section, we discuss an important qualification to our preceding discussion of statistical inferences about sample means. As we noted above, to determine the standard error of the sample mean—the degree to which the sample mean varies, by chance, from the population mean—the researcher needs to know the population standard deviation (signified by σ , the numerator of the standard error of mean formula). If that population parameter is unavailable, as it usually is, then the researcher at least needs an estimate of the population standard deviation. In most realistic research situations, the analyst has a random sample—and that's it. Just as the researcher can use the sample mean to estimate the population mean, the researcher can use the sample standard deviation to estimate the population standard deviation in calculating the standard error of the mean. So, in practice, the standard error of the sample mean is calculated as follows:

$$\text{Standard error of sample mean} = \frac{\text{Sample standard deviation}}{\sqrt{n}}$$

For fair-sized samples, substituting the sample standard deviation for σ works fine and the researcher can expect sample means to follow a normal distribution on repeated samples.⁹ Remember, though, that the central limit theorem only applies when the sample size is “large enough” and doesn’t apply when the sample size is too small. Take another look at the second and third columns of Figure 6-2. If the sample sizes are large enough, the expected sampling distributions take the signature bell-curve shape, but when $n = 5$ or $n = 15$, the expected sampling distributions are not smooth and symmetrical. When you are analyzing smaller samples, or if you have divided up a large sample into smaller subsamples for separate analyses, the exact properties of the normal distribution may no longer be applied in making inferences about means. This small sample problem requires us to adjust how we make inferences a bit to avoid overconfidence.

When the sample size is “too small” for the central limit theorem and the normal distribution, we instead use a similar distribution, the Student’s t -distribution (see Box 6-1). The **Student’s t -distribution** is a probability distribution that can be used for making inferences about a population mean when the sample size is relatively small. This special qualification does not apply to making inferences about proportions.¹⁰

When we estimate the mean value of a variable in a small sample, we use the standard error of mean formula above, using the sample standard deviation as our best estimate of the population standard deviation. In other words, we calculate the standard error of the mean the same way, no matter what size sample we’re analyzing. But we take a different approach when we use the standard error of the mean to make inferences about the expected sampling distribution and to calculate confidence intervals. We expect the mean of a sample to follow the Student’s t -distribution, not a normal distribution.

Normal distributions always have the signature “bell-curve” shape; the Student’s t -distribution is more varied in appearance. It might help to think of the t -distribution as a normal distribution that’s learned to relax (it may have had a beer or two). All t -distributions have lower peaks and thicker tails than the normal distribution (see Figure 6-11). The shape of the Student’s t -distribution depends on its degrees of freedom. **Degrees of freedom** are a statistical property of a large family of distributions, including the Student’s t -distribution. The number of degrees of freedom is equal to the sample size, n , minus the number of parameters being estimated by the sample. If we are using $n = 15$ observations to estimate a population mean, we would have $n - 1$ degrees of freedom: $15 - 1 = 14$ degrees of freedom. The smaller the sample size, the lower the degrees of freedom and the more relaxed the t -distribution. As sample size and degrees of freedom increase, t -distributions look more and more like the normal distribution, so much so that they are quickly indistinguishable. As you can see in Figure 6-11, the t -distribution with four degrees of freedom ($df = 4$) has a noticeably lower peak than the standard normal distribution and greater area in the tails of the distribution (compare the height of the lines below -2 standard errors and above $+2$ standard errors). As degrees of freedom increase, so too does the t -distribution’s peak, lowering its right and left tails. With just 39 degrees of freedom, the difference between a t -distribution and the normal distribution is barely noticeable.

When we make inferences assuming a normal-shaped sampling distribution, we do not have to worry about the size of the sample, so we calculate a value of Z and then find the area of the curve above that value. In estimation using the Student’s t -distribution, however, the sample size determines the shape of the distribution. Because the specific shape of the Student’s t -distribution depends on the sample size, the critical values listed in Table 6-2 look different from those listed in Table 6-1, which showed the area in the tails of the normal distribution for different values of Z .

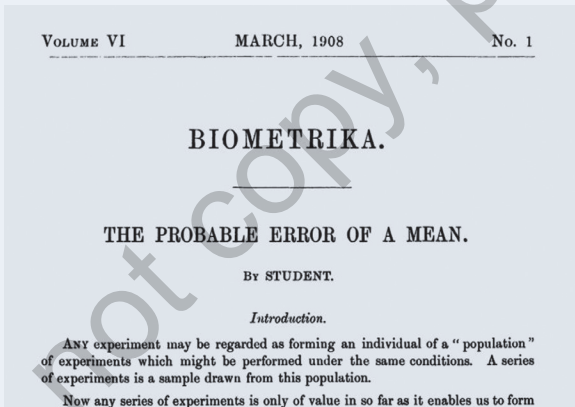
BOX 6-1

Why Is It Called Student's t -Distribution?

As a brewer in the Arthur Guinness & Son Brewery in Dublin, Ireland in the early 1900s, William Gosset experimented with different methods of brewing what was the most popular beer in the world at that time. How could he know whether manipulating some variable in the brewing process significantly affected the results? Gosset began to realize that the fundamental difference between the experiments he conducted and those he studied in school was sample size. Producing just one batch of a new recipe meant making a lot of beer, so thousands of independent observations of an experimental manipulation were not practical. However, this also meant that Gosset's experiments did not satisfy the minimum assumptions for making statistical inferences using the central limit theorem and the normal distribution.

Gosset discovered a new probability distribution that helped him make proper inferences about means based on small sample sizes. It was a profound discovery with implications far beyond beer making, but there was a problem with publishing it. To protect its trade secrets, the Guinness brewery didn't allow its employees to publish papers about their research. Fortunately, Gosset persisted and reached a compromise with his employer: rather than publish the paper under his own name, he would publish it anonymously as "Student" (see Figure 6-10). The probability distribution we use to make inferences about means based on small sample sizes became known as Student's t -distribution. Gosset worked for Guinness until his death at age 61.

Figure 6-10 Student and His Famous Paper



William Gosset ("Student")

Source: Gosset: User Wujaszek on pl.wikipedia [Public domain].

As you know, the shape of a t -distribution depends on degrees of freedom. Varying degrees of freedom are listed in the first column of Table 6-2, including several seen in Figure 6-11. Now examine the columns of Table 6-2, under the heading "Critical Value for Significance Level (Two-Tailed)." The columns are labeled with different percentages: 10%, 5%, 1%, and .1%. The entries in each column are values of t (the t -distribution's standardized counterpart to the normal distribution's Z -values).

Figure 6-11 *t*-Distributions Compared to a Standard Normal Distribution

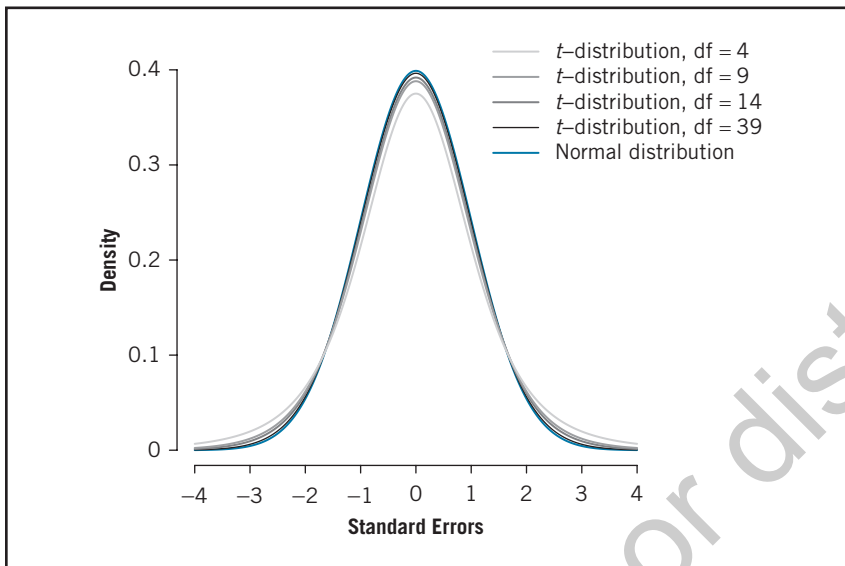


Table 6-2 Critical Values of Student's *t*-Distribution

Degrees of Freedom (<i>df</i>)	Critical Value for Significance Level (Two-Tailed)			
	10%	5%	1%	.1%
4 [†]	2.13	2.78	4.60	8.61
5	2.02	2.57	4.03	6.87
6	1.94	2.45	3.71	5.96
7	1.90	2.37	3.50	5.41
8	1.86	2.31	3.36	5.04
9 [†]	1.83	2.26	3.25	4.78
10	1.81	2.23	3.17	4.59
11	1.80	2.20	3.11	4.44
12	1.78	2.18	3.05	4.32
13	1.77	2.16	3.01	4.22
14 [†]	1.76	2.15	2.98	4.14
15	1.75	2.13	2.95	4.07

(Continued)

Table 6-2 (Continued)

Degrees of Freedom (<i>df</i>)	Critical Value for Significance Level (Two-Tailed)			
	10%	5%	1%	.1%
16	1.75	2.12	2.92	4.01
17	1.74	2.11	2.90	3.97
18	1.73	2.10	2.88	3.92
19	1.73	2.09	2.86	3.88
20	1.73	2.09	2.85	3.85
21	1.72	2.08	2.83	3.82
22	1.72	2.07	2.82	3.79
23	1.71	2.07	2.81	3.77
24	1.71	2.06	2.80	3.75
25	1.71	2.06	2.79	3.73
26	1.71	2.06	2.78	3.71
27	1.70	2.05	2.77	3.69
28	1.70	2.05	2.76	3.67
29	1.70	2.05	2.76	3.66
30	1.70	2.04	2.75	3.65
40	1.68	2.02	2.70	3.55
60	1.67	2.00	2.66	3.46
90	1.66	1.99	2.63	3.40
100	1.66	1.98	2.63	3.39
120	1.66	1.98	2.62	3.37
1,000	1.65	1.96	2.58	3.30
Normal (<i>Z</i>)	1.64	1.96	2.58	3.29

*These Student's *t*-distributions are plotted in Figure 6-11.

Each cell tells you the absolute value of *t* beyond which lies 10, 5, 1 and .1 percent of a *t*-distribution. For example, the top-most cell in the 10% column says that, with four degrees of freedom, 10 percent of a *t*-distribution lies beyond *t*-values of ± 2.13 . Reading across the row, you can see that 5 percent of the same *t*-distribution lies beyond ± 2.78 , 1 percent of the distribution falls past ± 4.01 , and the interval needed to capture 99.9 percent of the distribution's area stretches way out to ± 8.61 . That sets a very wide boundary for random sampling error.

When we make inferences about means with small samples, the boundaries of our 95% CIs are not fixed at ± 1.96 standard errors. When the population standard deviation is unknown and the sample size is small, the t -distribution sets wider boundaries than ± 1.96 standard errors and, therefore, permits less confidence in the accuracy of a sample statistic. How much wider the 95% CI is than ± 1.96 standard errors depends on how large a sample is being used for making inferences. As sample sizes increase, the t -distribution adjusts these boundaries accordingly, narrowing the limits of random sampling error and allowing more confidence in the measurements made from the sample.

Although the terminology used to describe the t -distribution is different from that used to describe the normal distribution, the procedures for drawing inferences about a population parameter are essentially the same. To illustrate, let's assume again that researchers record feeling thermometer values from a random sample with $n = 15$, get a sample mean of 58, and a sample standard deviation of 24.8. We already know how to calculate the standard error of this estimate: $24.8/\sqrt{100} = 2.48$. With only fifteen observations in the sample, we don't want to be overly confident in our estimate of the population mean. In a repeated sample, we would expect this sample mean to follow a t -distribution with 14 degrees of freedom (see the $df = 14$ distribution in Figure 6-11). How much wider do we set the boundaries of a 95% CI when we expect this sampling distribution instead of a normal distribution? Refer to Table 6-2 and locate 14 degrees of freedom and the 5% significance level; 2.15 is the critical value to use. With a small sample like this, we use 2.15 as the critical value (rather than 1.96). Let's use this information to find the 95% CI for a sample mean.

$$\begin{aligned} \text{95\% CI lower boundary} &= 58 - (2.15 \times 2.48) \\ &= 58 - 5.33 \\ &= 52.67 \end{aligned}$$

$$\begin{aligned} \text{95\% CI upper boundary} &= 58 + (2.15 \times 2.48) \\ &= 58 + 5.33 \\ &= 63.33 \end{aligned}$$

The 95% CI in this example is [52.67, 63.33]. It is wider than the 95% CI we would get if we had used a large sample and the normal distribution.¹¹ To calculate the 90% or 99% CI, we would take the same approach, adjusting the critical value to correspond to 10 percent and 1 percent significance for a t -distribution with 14 degrees of freedom (referring again to Table 6-2, those critical values are 1.76 and 2.98).

The signature of the Student's t -distribution is that it adjusts the confidence interval, depending on the size of the sample. Indeed, notice what happens to the t -values in the 5% column as the sample size (and thus degrees of freedom) increases. As sample size increases, the critical value of t that marks the 5 percent boundary begins to decrease. More degrees of freedom mean less random sampling error and, thus, more confidence in the sample statistic. For comparison purposes, the value of Z that is associated with the 5 percent benchmark, $Z = 1.96$, appears in the bottom cell of the 5% column of Table 6-2. Notice that the critical value of t for a large sample (degrees of freedom = 1,000) rounds to ± 1.96 . So, for samples with many degrees of freedom, the Student's t -distribution closely approximates the normal distribution.

With samples as small as 40 observations, the t -distribution so closely resembles the normal distribution that the empirical rule's convenient benchmarks for 1, 2, and 3 standard errors are useful. Recall that the 95% CI can be determined quickly by multiplying the standard error by 2, then subtracting this number from the sample

mean to find the lower confidence boundary and adding this number to the sample mean to find the upper confidence boundary. This is a good rule, because it works well in most situations. As you can see from Table 6-2, even for fairly small samples—those having 40 or more degrees of freedom—the shortcut will provide an adequate estimate of the 95% CI.

SUMMARY

Just how accurately does a sample statistic estimate a population parameter? You now know that the answer is a resounding “It depends on three factors.” First, it depends on whether the sample was randomly selected from the population. As discussed in Chapter 4, by ensuring that each member of the population has an equal chance of being included in the sample, the researcher eliminates “bad” error, systematic error, from the sample statistic. A random sample permits the researcher to estimate the amount of “good” error, random error, contained in the sample statistic. Second, it depends on the size of the random sample. Larger samples yield better estimates than smaller samples. But you now understand why samples that seem small can nonetheless provide a solid basis for inference. Third, it depends on the amount of variation in the population. You are now familiar with a key measure of variation for interval-level variables, the standard deviation. And you know how the standard deviation works together with sample size in bracketing the confidence interval for a sample mean.

Many symbols and terms were discussed in this chapter. Table 6-3 provides a list, arranged in roughly the order in which these terms and symbols were introduced. Let’s review them. The population mean (μ) is the parameter of chief concern—the measure of central tendency that the researcher is most interested in estimating. Variation around the population mean is determined by the standard deviation (σ), the measure of dispersion that summarizes how much clustering or spread exists in the population. The relative position of any number in a list can be expressed as a Z score, the number of standard units that the number falls above or below the mean. The standard deviation and the standard error are standard units of deviation from the mean that can be used to convert raw values into Z scores. By knowing Z and applying the inferential properties of the normal distribution, we can decide how closely the sample mean, symbolized by \bar{x} , approximates the population mean.

Table 6-3 Terms and Symbols and the Roles They Play in Inference

Term or Symbol (Pronunciation)	What It Is or What It Does	What Role It Plays in Sampling and Inference
μ (“mew”)	Population mean	Usually μ is unknown and is estimated by \bar{x}
N	Population size	Usually N is unknown and assumed to be very large
σ (“sigma”)	Population standard deviation	Measures variation in a population characteristic. The variation component of random sampling error
Z score	Converts raw deviations from μ into standard units	Defines the tick marks of the normal distribution; 68 percent of the distribution lies between $Z = -1$ and $Z = +1$; 95 percent of the distribution lies between $Z = -1.96$ and $Z = +1.96$
\bar{x} (“x bar”)	Sample mean	Sample statistic that estimates μ

Term or Symbol (Pronunciation)	What It Is or What It Does	What Role It Plays in Sampling and Inference
n	Sample size	The sample size component of random sampling error is \sqrt{n}
s	Sample standard deviation	Substitutes for σ as the variation component of random sampling error when σ is unknown
Standard error of the sample mean	Measures how much \bar{x} departs, by chance, from μ	Random sampling error Equal to σ/\sqrt{n} , if σ is known Equal to s/\sqrt{n} , if σ is unknown
95 percent confidence interval (95% CI)	The interval in which 95 percent of all possible values of \bar{x} will fall by chance	Defined by $\bar{x} \pm 1.96$ standard errors in normal estimation. Can usually be determined by the shortcut rule: $\bar{x} \pm 2$ standard errors in all estimation
p	Proportion of a sample falling into one value of a nominal or ordinal variable	Sample estimate of a population proportion
q	Proportion of a sample falling into all other values of a nominal or ordinal variable	Equal to $1 - p$
Standard error of a sample proportion	Measures how much p departs, by chance, from a population proportion	Defined by \sqrt{pq} / \sqrt{n} . Ordinarily can be applied in finding the 95 percent confidence interval of p , using normal estimation

We know that the sample mean will be equal to the population mean, plus any random sampling error that was introduced in drawing the sample. The size of this error, termed the standard error of the sample mean, is determined by σ and the sample size (n). Again applying the normal distribution, the researcher can estimate the 95% CI for \bar{x} , the boundaries within which 95 percent of all possible sample means will fall by chance. Z scores are directly applied here. By multiplying the standard error by $Z = 1.96$ —or rounding up to 2 by the

simplifying shortcut—the researcher can figure the probable boundaries of the true population mean. In practice, the population standard deviation is rarely known, so the researcher uses the sample standard deviation, denoted by s , as a stand-in for σ , and then applies the Student's t -distribution. As you know, much political research, especially survey research, involves nominal and ordinal variables. In this chapter we also discussed how normal estimation may be usefully applied in estimating the 95% CI for a sample proportion.

Take a closer look. edge.sagepub.com/pollock

KEY TERMS

95 percent confidence interval (p. 184)
census (p. 168)
central limit theorem (p. 172)

cumulative density (p. 176)
degrees of freedom (p. 189)
empirical rule (p. 174)
inferential statistics (p. 168)

normal distribution (p. 174)
population (p. 168)
population parameter (p. 168)
probability (p. 184)

random sample (p. 169)

random sampling error (p. 169)

sample (p. 168)

sample statistic (p. 169)

sampling distribution (p. 170)

standard error (p. 169)

standard normal

distribution (p. 176)

standardization (p. 174)

Student's *t*-distribution (p. 189)

Z score (p. 174)

Screencasts



Chapter 6, Pt. 1



Chapter 6, Pt. 2



Chapter 6, Pt. 3



Chapter 6, Pt. 4

EXERCISES

- Each of the following proportions is based on survey responses. For each proportion, use the ± 2 shortcut to determine the 95% CI.
 - When asked if they are a government employee, .20 said “yes” ($n = 121$).
 - When asked if marijuana should be legal or illegal, .47 said “legal” ($n = 100$).
 - Of the individuals in a survey, .40 still live in the same city they lived in when they were 16 years of age ($n = 225$).
- In response to a GSS question in 2006 about the number of hours spent per day watching television, the responses by the fifteen subjects who identified themselves as Buddhist were 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 3, 4, 4, and 5. For these fifteen subjects, the mean number of hours spent per day watching television is 1.93 hours.
 - What is the standard error of this mean estimate?
 - Using the ± 2 shortcut, calculate the 95% CI.
- A recent survey asked college freshmen to correctly identify as many U.S. Supreme Court justices as possible. Of the 816 students who responded to this survey, the mean number of justices correctly identified was 1.81 with a standard deviation of 1.98.
 - What is the standard error of this mean estimate?
 - Using the ± 2 shortcut, calculate the 95% CI.
- Professor Kingsfield’s class follows very strict grading guidelines. Assume that student averages in Kingfield’s class are normally distributed with a mean of 62 points and a standard deviation of 12 points. Professor Kingsfield wants letter grades in his class to have a very specific distribution: 68 percent of students get Cs, 13.5 percent get Ds, 13.5 percent get Bs, 2.5 percent receive Fs, and the top 2.5 percent get As. What should be the numeric range for each letter grade? Refer to Figure 6-5 for help answering this question.
 - Students with ____ or above get As.
 - Students with averages between ____ and ____ get Bs.
 - Students with averages between ____ and ____ get Cs.
 - Students with averages between ____ and ____ get Ds.
 - Students with ____ or below get Fs.
- Some states allow voters to pass laws through ballot initiatives. In most states that allow ballot initiatives, a simple majority is all that’s needed. Before backing a ballot initiative, a political organization wants to know whether the majority of a state’s voters would be in favor of it. The organization wants to know what size random sample is required to estimate the percentage of voters who favor the initiative with different margins of error. Find the sample size needed to estimate support for the initiative with 5, 3, 1, and 0 percent margins of error. If any of these sample sizes can’t be calculated, explain why.
- Sociologists have conducted much interesting research on gender stereotypes in American society. A curious aspect of stereotypes is that people tend to perceive differences between groups to be greater than they actually are.

This suggests, for example, that when asked about the heights of men and women, survey respondents would tend to perceive men to be taller than women. Suppose you wanted to test this notion that individuals perceive a greater height difference between men and women than exists in the population. In the population, men, on average, are 5 inches taller than women. So the true population difference between men and women is 5 inches.

You obtain a random sample of 400 individuals. For each respondent, you record his or her perceptions of the difference between male and female heights. In your sample, you find that the mean difference in perceived heights is 6 inches. So respondents perceive that men are 6 inches taller than women. The sample standard deviation is 4 inches.

- A. Calculate the standard error of the sample mean.
- B. Using the ± 2 shortcut, calculate the 95% CI.
- C. If the true gender difference is 5 inches, can you infer from your sample that

individuals perceive a greater difference than actually exists? Explain.

7. The sheriff is concerned about speeders on a certain stretch of county road. The sheriff, a retired statistician, sets up a radar device and, over a long period of time, obtains data on the entire population of vehicles using the road. The mean vehicle speed: 50 miles per hour.
 - A. The sheriff cracks down on speeders. Following the crackdown, the sheriff takes a random sample ($n = 100$) of vehicle speeds on the roadway. The sample data: sample mean, 47 mph; sample standard deviation, 8 mph. (i) Using the sample standard deviation as a substitute for the population standard deviation, what is the standard error of the sample mean? (ii) Using the ± 2 shortcut, what is the 95% CI of the sample mean?
 - B. A skeptical county commissioner claims that the crackdown had no effect and that average speed on the roadway is still 50 mph. (i) Is the skeptic on solid statistical ground? (ii) Explain how you know.

NOTES

1. The terms *population characteristic* and *population parameter* are synonymous and are used interchangeably in this chapter.
2. Computational methods that simulate data to approximate solutions are sometimes called Monte Carlo simulations (a reference to the origins of methods to study games of chance played in Monte Carlo's famous casinos). For example, rather than analytically deriving the probability of winning a game of Solitaire, you could have a computer play many hands of Solitaire and estimate the probability of winning.
3. Any number of distributions of thermometer scores in the population could produce these parameters. We don't need to assume thermometer scores have a defined distribution in the population.
4. Unlike sample statistics, which are represented by ordinary letters, population parameters are always symbolized by Greek letters. A population mean is symbolized by the Greek letter μ (pronounced "mew").
5. The sample size is denoted by a lowercase n . By contrast, a population size is denoted by an uppercase N .
6. The simulations presented in Figure 6-6 and Figure 6-7 were created using the Stata program, *bxmodel* (version 1.2, January 31, 2006), written by Philip B. Ender, Statistical Computing and Consulting UCLA, Academic Technology Services.
7. If you transform a nominal- or ordinal-level variable into a dummy variable with values coded as 0 and 1, the dummy variable's mean equal the proportion of 1s (p) and the dummy variable's standard deviation is nearly the same as the square root of the product of p and $1 - p$ (the latter value is slightly smaller).
8. Statisticians typically detest imprecision, but this shortcut rule is acceptable because it is conservative; that is, it widens slightly the bandwidth of the random error associated with the 95% CI.
9. Normal estimation may be used for samples of $n = 100$ or more. See David Knoke, George W. Bohrnstedt, and Alisa Potter Mee, *Statistics for Social*

Data Analysis, 4th ed. (Belmont, Calif.: Wadsworth/Thomson, 2002), 128.

10. The small sample issue doesn't come up when we make inferences about proportions because we know, based on how the variable is defined, that the only possible variable values in the population are 0 and 1.

The standard error of proportion formula does not systematically underestimate standard error when the sample sizes are small the way the standard error of mean formula does.

11. With a large sample, the same sample mean and standard deviation, the 95% CI is [53.14, 62.86].

Do not copy, post, or distribute