Chapter One

An Introduction to Econometrics and Statistical Inference

CHAPTER OBJECTIVES

After reading this chapter, you will be able to:

1. Understand the steps involved in conducting an empirical research project.
2. Understand the meaning of the term *econometrics*.
3. Understand the relationship among populations, samples, and statistical inference.
4. Understand the important role that sampling distributions play in statistical inference.

A STUDENT'S PERSPECTIVE

empirical research project
A project that applies empirical analysis to observed data to provide insight into questions of theoretical interest.

Suppose that having enrolled in an introductory econometrics course you are wondering what you are going to learn and why doing so is going to be valuable to your future. What is the answer to this most important question?

This is exactly the situation encountered by all of our former students. The short answer is that you will learn to conduct an **empirical research project** from start to finish. The long answer is that you will acquire important skills that are highly valued in the modern labor market.

BIG PICTURE OVERVIEW

> *Hal Varian likes to say that the sexy job in the next ten years will be statisticians. After all, who would have guessed that computer engineers would be the cool job of the 90s? When every business has free and ubiquitous data, the ability to understand it and extract value from it becomes the complimentary scarce factor. It leads to intelligence, and the intelligent business is the successful business, regardless of its size. Data is the sword of the 21st century, those who wield it well, the Samurai.*

—Jonathan Rosenberg, senior vice president of product management at Google, February 16, 2009, http://googleblog.blogspot.com/2009/02/from-height-of-this-place.html

As the quote implies, empirical research skills are extremely valuable. The market value of possessing such skills is evidenced by the fact that in 2011, college graduates with economics degrees (one of the most empirically focused majors) were among the highest-paid nonengineering majors in terms of their median mid-career earnings, as shown in Figure 1.1.

FIGURE 1.1
The 20 Best-Paying College Degrees by Mid-Career Median Salary in 2011 according to Payscale.com

College Degree Mid-Career Median Salary
Petroleum engineering $155,000
Chemical engineering $109,000
Electrical engineering $103,000
Material science & engineering $103,000
Aerospace engineering $102,000
Physics $101,000
Applied mathematics $98,600
Computer engineering $101,000
Nuclear engineering $97,800
Biomedical engineering $97,800
Economics $94,700
Mechanical engineering $94,500
Statistics $93,800
Industrial engineering $93,100
Civil engineering $90,200
Mathematics $89,900
Environmental engineering $88,600
Management information systems $88,200
Software engineering $87,800
Finance $87,300

To maximize your chances of becoming a samurai in the world of empirical research, this textbook focuses on helping you acquire four skills: (1) the creative-thinking skills required to identify sufficiently interesting questions of interest and appropriate economic theories, (2) the data collection and management skills required to assemble and work with appropriate data sets, (3) the empirical skills required to perform valid statistical inference, and (4) the communication skills required to successfully convey the quality and importance of the work to potential readers.

We depict the goals of this text in Figure 1.2.

FIGURE 1.2
A Visual Depiction of the Goals of This Textbook

Goal → Learn to Conduct an Empirical Research Project

↑

We want to learn to do this

To do so, we need to acquire:

(1) Creative-thinking skills required to identify questions of interest and appropriate economic theories.
(2) Data collection and management skills required to assemble and work with appropriate data sets.
} **Chapter 2**

(3) Empirical skills required to perform valid statistical inference.
} **Chapters 3–15**

(4) Communication skills required to successfully convey the quality and importance of the work to the potential reader.
} **Chapter 16**

To get started, we need to introduce the individual steps involved in conducting an empirical research project.

1.1 UNDERSTAND THE STEPS INVOLVED IN CONDUCTING AN EMPIRICAL RESEARCH PROJECT

In broad terms, there are five steps to conducting a successful empirical research project: (1) determining the question of interest; (2) developing the appropriate theory relating to that answer; (3) collecting data that are appropriate for empirically investigating the answer; (4) implementing appropriate empirical techniques, correctly interpreting results, and drawing appropriate conclusions based on the estimated results; and (5) effectively writing up a summary of the first four steps that conveys to readers the value of the work that was completed.

Step 1: Determining the Question of Interest

In many ways, this is the most important step in producing a quality empirical research project. Without an intriguing question, even the most technically proficient project is likely to fall on deaf ears because the reader is likely to have no investment whatsoever in the outcome. Conversely, a project addressing a question of great interest is likely to grab and hold a reader's attention regardless of the degree of technical rigor with which it is executed.

Determining a sufficiently interesting question is, of course, the million-dollar question for researchers. Unfortunately, there is no clear answer as to how to best determine those questions that will be of the greatest interest to the greatest number of readers. Some individuals are gifted with the innate ability to identify such questions, while others have great difficulty mastering the art. In general, given that we are all more passionate about subjects in which we are personally invested, we suggest starting by identifying topics that you find particularly interesting. For instance, a restaurant server might be interested in analyzing the factors associated with receiving higher tips; a macroeconomics student might be interested in analyzing the factors associated with lower unemployment in certain states and countries; a football fan might be interested in analyzing the factors associated with a team making the playoffs in a given season.

Once we have identified a general topic, we need to determine an appropriate question related to that topic. How can we be sure that our chosen question is likely of sufficient interest to potential readers? A useful tip is to write down the proposed question and read it aloud to see whether it passes the sound test (i.e. does it sound like a question that will interest potential readers?). If it passes this test, then we can proceed to the next step in the process. If it does not, then we will need to repeat the process until we identify a question that does pass the test.

Step 2: Developing the Appropriate Theory Relating to the Answer

This step bears the fruit of previous coursework that has been completed. For instance, drawing on their accumulated knowledge, the server mentioned earlier might hypothesize that the couples dining out on weekend nights are more likely to tip well than couples dining out on weekday nights; the macroeconomics student might hypothesize that states with more educated citizens are likely to have lower unemployment rates; and the football fan might hypothesize that the greater a team's payroll, the better the team is likely to perform on the field.

Step 3: Collecting Data That Is Appropriate for Empirically Investigating the Answer

Before even attempting to perform empirical analysis, we *must* collect data that are appropriate for investigating our chosen question of interest. In practice, this step is likely as important as any other in determining the ultimate success of our project because without appropriate data, we will be unable to complete the project regardless how clever or important our chosen question. For this reason, learning how to collect (and eventually manage) data is a vitally important skill that every budding empirical researcher needs to develop. Starting to acquire these skills is the focus of Chapter 2 in this textbook.

Step 4: Implementing Appropriate Empirical Techniques, Correctly Interpreting Results, and Drawing Appropriate Conclusions

Once a sufficiently interesting question has been identified, the correct underlying theory has been determined, and the appropriate data have been collected, we must be able to properly execute the empirical analysis required to perform the necessary statistical

An Important Caveat about the Relative Importance of Each Step

While the text focuses almost exclusively on introducing the tools required to perform appropriate empirical analysis (step 4 in conducting an empirical research project), once those tools are acquired (which itself is a difficult and time-consuming process), steps 1 through 3 and step 5 generally require at least as much, if not more, actual time than step 4. There are several reasons for this. For one, identifying questions of sufficient interest and identifying and collecting appropriate data are often both quite difficult and time-consuming, particularly for those who are new to the practice. For another, once the appropriate data have been collected, the appropriate econometric analysis performed, and the correct conclusions drawn, producing an effective write-up takes time because doing so is often an iterative process that requires several drafts before successful completion.

For this reason, we encourage students tasked with conducting empirical research projects to start their work sooner rather than later. We cannot count the number of times that just before the project deadline, we have been contacted by students who have been stuck on one of the early steps (most often finding appropriate data for analyzing their chosen question) and have therefore been having great difficulty completing their work. Almost invariably, these students have commented that they now wish they had taken our advice about the relative importance of the other steps more seriously.

inference to answer the question of interest. Developing the skills required to successfully perform this step is the focus of Chapters 3 through 15.

Step 5: Effectively Writing Up a Summary of the Empirical Work

This is our opportunity to sell the significance and the quality of our work to potential readers. While this might sound somewhat easier than several of the preceding steps, its importance should not be discounted: Regardless of how well-executed steps 1 through 4 are, if the quality of those steps is not communicated sufficiently well in the write-up, then the reader is unlikely to be left with a positive view of the work. Given the importance of this step, as a capstone, Chapter 16 provides a detailed example of how to tie steps 1 through 4 together in an effective write-up.

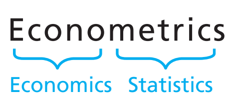
Before embarking on our journey, we need to introduce the term that is at the heart of the empirical tools that we discuss.

1.2 UNDERSTAND THE MEANING OF THE TERM *ECONOMETRICS*

econometrics
The application of statistical techniques to economic data.

In broad terms, **econometrics** is the intersection between the disciplines of economics and statistics. This should be evident from the term itself. Refer to Figure 1.3 below.

FIGURE 1.3
An Introduction to the Study of Econometrics



economics
The study of how people chose to use scarce resources.

statistics
The science of collection, organization, analysis, and interpretation of data.

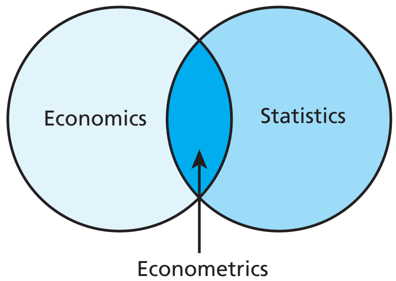
How do these root words combine to create the field of econometrics? And how is the study of econometrics potentially of great value? To answer these questions, it is important to consider the words in more detail. **Economics** is a social science examining the human behavior of "how individuals respond to incentives." **Statistics** is "the mathematics of the collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling."

This combination of words interacts as follows. To examine human behavior, economists develop scientific models known as *economic theory* that predict how individuals, firms, and so on, respond to incentives in different situations. While economic theories can certainly be developed to explain almost any human behavior, they are only valuable if they accurately explain how humans actually behave. This brings us to the intersection between economics and statistics. How might we determine whether a given economic theory is indeed consistent with actual human behavior? By collecting data that results from actual human behavior and applying appropriate empirical techniques to determine whether the observed behavior is consistent with the theoretical model.

Venn diagram
A tool for visually depicting the overlap between the total variation in two or more different factors of interest.

As an alternative way to think about the study of econometrics, we introduce the **Venn diagram** in Figure 1.4. Letting the lighter circle represent the entire discipline of economics and the slightly darker circle represent the entire discipline of statistics, the overlap between these two disciplines, which is demonstrated as the darkest area, is econometrics.

FIGURE 1.4
A Venn Diagram Explanation of Econometrics



What does Figure 1.4 tell us about the study of econometrics? On the one hand, it tells us that not all of economics involves the application of statistics and that not all of statistics is applied to economics. Indeed, it is the case that much of the work in both disciplines is performed independent from the other. While true, there is most definitely an important overlap between the two fields, and as indicated in Figure 1.4, that overlap forms the field of econometrics.

Note that while we couch the preceding discussion in terms of economics, we could just as easily place other social science and business disciplines in the green area, suggesting that the tools introduced in this text can be applied to questions of interest in other many other disciplines. Indeed, many of the data sets that we analyze touch on questions of importance to disciplines other than economics.

With this discussion in mind, we turn to the fundamental issue that lies at the heart of statistics.

1.3 UNDERSTAND THE RELATIONSHIP AMONG POPULATIONS, SAMPLES, AND STATISTICAL INFERENCE

By definition, *statistics* is

> *"the mathematics of the collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling."*
>
> www.thefreedictionary.com/statistics

This definition introduces three important terms that play a vital role in the study of econometrics—*population, sampling,* and *inference.* Given that these three terms form the very basis of statistical analysis, we need to consider their meanings in detail.

Populations and Samples

As discussed earlier, once we have developed a given economic theory, we can attempt to determine whether that theory is indeed consistent with actual human behavior by (1) collecting data resulting from actual human behavior and (2) using appropriate techniques to estimate the degree to which the observed behavior is consistent with the underlying theory. The question is, in order to adequately answer our chosen question, do we need to collect data on the observed behavior of all individuals subject to the given theory? This question becomes increasingly important when we consider that in many cases, a given economic theory affects potentially millions and millions of individuals. For instance, suppose that we wish to test the economic theory that reductions in marginal tax rates increase consumer spending in the United States. Given that there are more than 300 million citizens in the United States, it is important to ask whether determining if our theory is correct requires data on the level of spending at different marginal tax rates for every single one of those 300 million citizens. Fortunately for us, the answer is no. Understanding why not is where the three important terms, *population, sampling,* and *inference,* come into play.

population
The entire group of entities that we are interested in learning about.

sample
A subset or part of the population; what is used to perform statistical inference.

confidence
A statement about the chance that an event (or group of events) will occur.

Continuing with the marginal tax rate example, the 300 million plus U.S. citizens represent the **population** that we are interested in learning about. To answer the question of whether we need to collect data on the entire population to determine whether our economic theory is indeed consistent with observed human behavior, we need to consider the potential costs and benefits associated with collecting data for the entire population versus collecting data for only a given **sample** drawn from the population.

Returning to our question of the size of the sample that we must collect in order to learn about the population of interest, on the one hand, the greater the percentage of the population for which we collect data, the greater the **confidence** we have in our results. In fact, if we collect data on the entire population, we will have 100 percent confidence in our results, as we will be 100 percent certain that they accurately reflect the entire population. On the other hand, collecting data is expensive because every single observation collected requires expenditures of time and effort, which can potentially sum to a very, very significant expense for a larger population (i.e., there is a reason that the U.S. census is only conducted once every 10 years). Given these competing concerns, we would like to find a middle ground that balances the benefit of collecting data on a larger percentage of the population with the cost of assembling a very large sample. The question we must address is how large a sample we need to collect in order to have sufficient confidence that the conclusions we draw from the observed sample are likely to be accurate reflections of the population we are trying to learn about. As we will see, the answer is that for our conclusion to be considered valid, the number of observations in the sample drawn from the population need only be what many would consider a surprisingly small number.

statistical inference
The process of drawing conclusions from data that are subject to random variation.

It is important to note that the fact that we generally only observe a specific sample drawn from the population means that the population values we are interested in learning about are generally unobserved. Instead, the only information available to us is the statistics that are observed for the given sample drawn from the population. As a result, we generally need to use the statistics observed for the given sample drawn from the population as our best indication of the true (but unobserved) population parameters. The process of using statistics observed for a given sample as the best estimate of parameters that likely exist for the population of interest is referred to as **statistical inference.**

As a concrete example to help make the previous topics easy to visualize, we like to consider fishing for rainbow trout in a beautiful alpine lake, such as the one pictured in Figure 1.5.

FIGURE 1.5
A Beautiful Lake Containing Thousands of Rainbow Trout

representative sample
A sample indicative of what would be found with access to the entire population.

independent random sampling
A technique in which each individual is chosen entirely by chance and each member of the population has a known, but possibly nonequal, chance of being included in the sample.

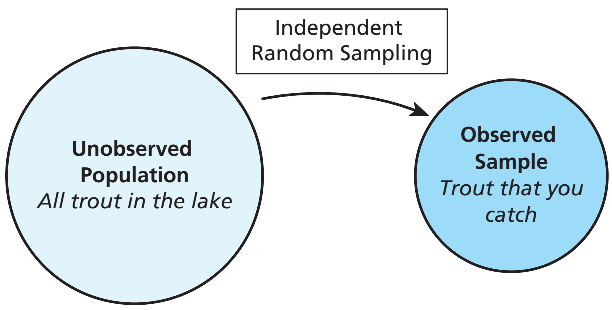
sampling with replacement
Takes place when a unit is drawn from a finite population, the value is recorded, and then returned to the sample.

To determine the typical size of rainbow trout in this lake, do we need to catch and measure all of the millions of rainbow trout that live there? No. We only need to catch a relatively small **representative sample** of rainbow trout to generate a valid estimate. This is true because as long as the sample of fish that we catch is truly representative of the population from which it is drawn, then we should be able to perform valid statistical inference on the basis of that sample (i.e., the typical size of the fish in the sample should provide an accurate indication of the typical size of the fish in the entire lake).

How do we know whether the specific sample drawn from the population is indeed representative of the population itself? One guarantee is that the chosen sample is collected through **independent random sampling,** which is ensured if we **sample with replacement** by returning each observation to the population once we have selected it and recorded its value. In the case of our rainbow trout example, we would do so by releasing all fish back into the lake once we catch and measure them. If we return each fish to the lake after catching it, then presumably there is an equal chance of catching a given fish every single time we cast our line and, as a result, the likelihood of catching a given fish with one cast should be independent of the likelihood of catching a given fish with a different cast. In practice, most sampling is conducted without replacement. This tends not to be an issue as long as the size of the population is much, much larger than the size of the sample because, in such cases, the probability of drawing two of the same observations is extremely small.

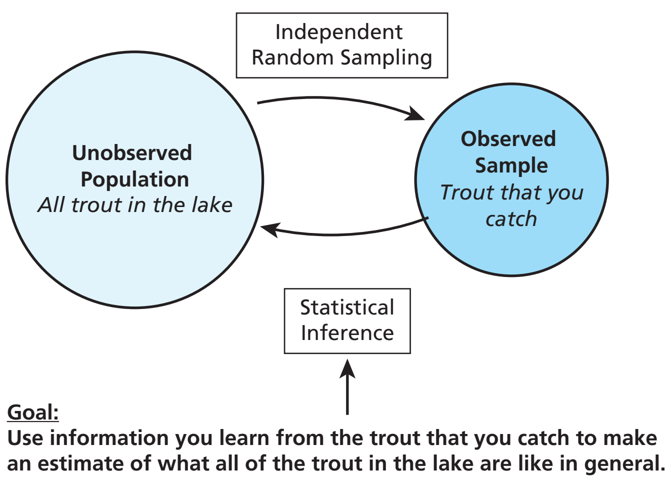
Visually, we depict the statistical inference process in Figure 1.6.

FIGURE 1.6
An Introduction to Independent Random Sampling



Once we have collected our specific independent random sample from the population, what do we do with it? We use the observed values for the sample as our best estimate of the values that are likely to exist within the entire population. As discussed earlier, we refer to this process as performing statistical inference, and we visually represent it in Figure 1.7.

FIGURE 1.7
A Visual Depiction Statistical Inference



Before continuing our discussion, we need to define a few important terms related to populations and samples that will be used throughout the remainder of the textbook. The true (but unobserved) value for the variable of interest that exists within the population is

parameter
A function that exists within the population.

statistic
A function that is computed from the sample data.

point estimate
A single valued statistic that is the best guess of a population parameter.

referred to as a **parameter.** The value for the variable of interest that is observed for the specific random sample drawn from the unobserved population is referred to as a sample **statistic.** Finally, because the sample statistics serve as our best estimates of the true population parameters, we often refer to sample statistics as being our best **point estimates** of the unobserved population parameters.

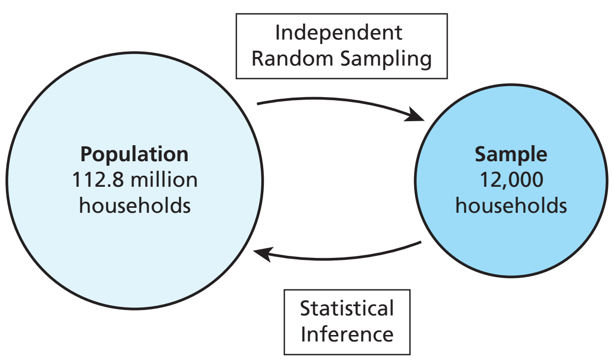
What does this discussion mean in practice? A simple example that is likely familiar to most readers provides context.

A Real-World Example of Statistical Inference: The Nielsen Ratings

Suppose that we own a retail business and that we are interested in purchasing televison advertising time in hopes of increasing revenue and, ultimately, profit. What information would we need to make an informed decision as to the time of day and the program during which we should purchase advertising time? The answer to this question is complicated by the fact that the number of consumers predisposed to favor our product is likely to vary by time of day and type of program, meaning that there are likely certain times that are more profitable than others for our ad buys. How can we determine which times are more profitable and which times are less profitable to advertise? By collecting a specific sample of a given size from the population and using information observed for that sample to draw conclusions about the population from which the sample is drawn.

Do we need to perform this calculation ourselves? No. In America, the task of collecting information on household viewing habits is performed by Nielsen Media Research that, according to the website www.nielsen.com/us, "tracks the television and media-viewing habits of homes across the country" and produces a statistic commonly referred to as the Nielsen Ratings. According to its website, there are currently roughly 112.8 million television households in the United States. Is it possible to observe the viewing habits of all of these 112.8 million households? No! So what does Nielsen Media Research do? It surveys the viewing habits of a randomly selected sample of roughly 12,000 households and uses observations from that sample to draw conclusions about the population of roughly 113 million households. This means that the company is attempting to determine the viewing habits of nearly 113 million television households based on observations of only 12,000 households, or 0.01 percent of the population. Figure 1.8 visually depicts this process.

FIGURE 1.8
A Visual Depiction of the Real-World Nielsen Ratings Example



A question that may occur to you at this point is how can such a small sample (12,000 households) provide valid information as to such a large population (more than 112 million households)? This is where the statistical inference process becomes rather amazing to many individuals. Based on the laws of statistics, if our data are collected through independent random sampling, then we can draw accurate conclusions about the population of interest from even a relatively small-seeming sample drawn from the population. Indeed, on its website, Nielsen Media Research highlights the importance of its sampling approach by stating that the first crucial step is to "scientifically select a group of households that mirrors the population at large . . . Selecting a representative sample of homes is vital to collecting data that mirrors the population's viewing habits . . . Nielsen TV families are a cross-section of households from all over the United States . . . We carefully draw our samples in a way that offers every American household with a television an equal chance of being selected."

While this discussion might seem rather complicated, the issues discussed are summarized quite nicely by the following joke:

> *A statistics professor was describing sampling theory to his class, explaining how a sample can be studied and used to generalize to a population. One of the students in the back of the room kept shaking his head. "What's the matter?" asked the professor. "I don't believe it," said the student, "why not study the whole population in the first place?" The professor continued explaining the ideas of random and representative samples. The student still shook his head. The professor launched into the mechanics of proportional stratified samples, randomized cluster sampling, the standard error of the mean, and the central limit theorem. The student remained unconvinced saying, "Too much theory, too risky, I couldn't trust just a few numbers in place of ALL of them." Attempting a more practical example, the professor then explained the scientific rigor and meticulous sample selection of the Nielsen television ratings which are used to determine how multiple millions of advertising dollars are spent. The student remained unimpressed saying, "You mean that just a sample of a few thousand can tell us exactly what over 112 MILLION people are doing?" Finally, the professor, somewhat disgruntled with the skepticism, replied, "Well, the next time you go to the campus clinic and they want to take a sample to perform a blood test . . . tell them that's not good enough . . . tell them to TAKE IT ALL!!"*

sampling distribution
The distribution of a sample statistic such as the sample mean.

Given the preceding discussion, an important question is how can we use the observed sample statistics to learn about the unobserved population parameters that we are attempting to estimate. The key to answering this question lies in our knowledge **sampling distributions.**

1.4 UNDERSTAND THE IMPORTANT ROLE THAT SAMPLING DISTRIBUTIONS PLAY IN STATISTICAL INFERENCE

notation

n number of observations in a sample

N number of observations in the population

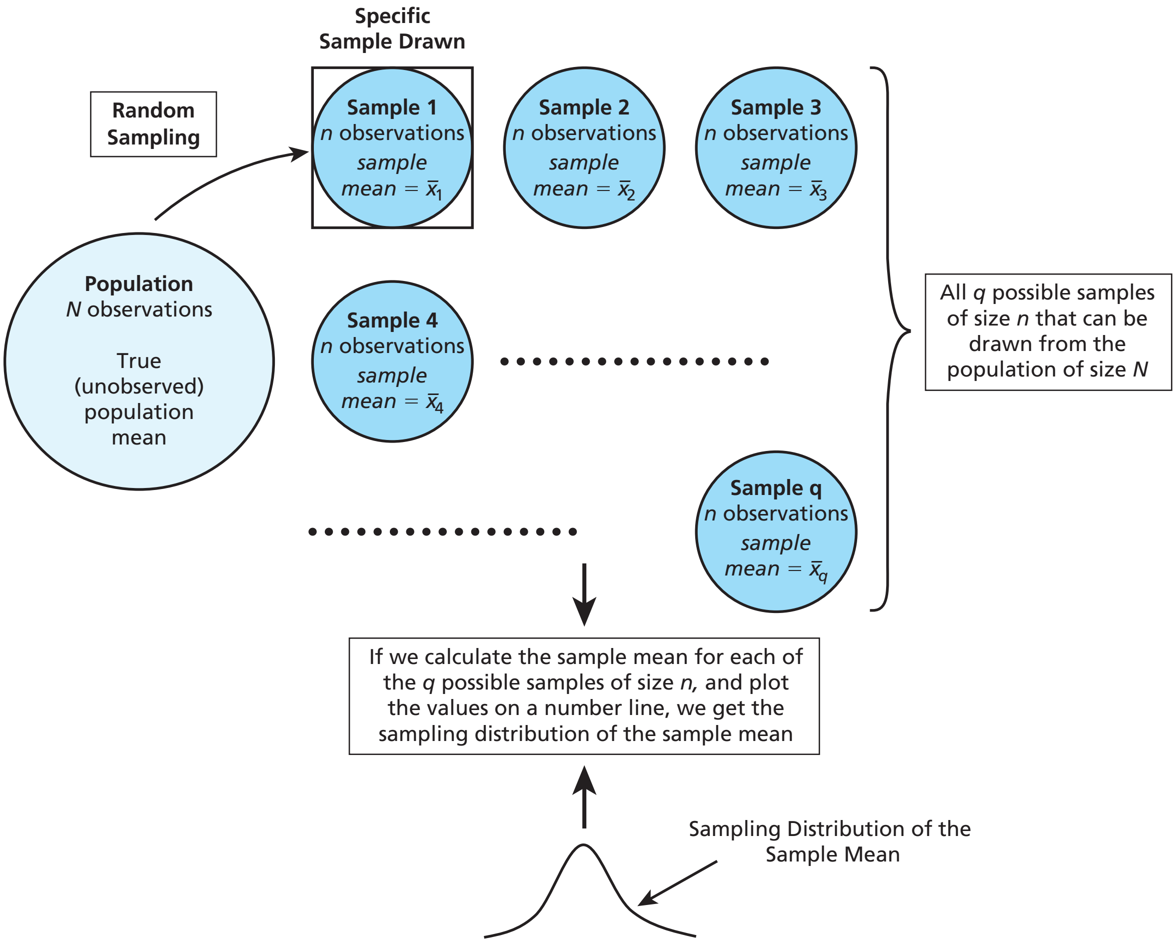
As we will discuss in Chapter 5, a key to understanding the statistical inference process is the realization that the specific independent random sample of size n that is drawn from the *unobserved* population of size N is but one of many, many potential samples of size n that could have been drawn from the population. Given that each of the many, many potential independent random samples of size n are likely to provide different information, an important concern is the degree to which each individual sample is likely to provide a valid estimate of the unobserved population. As Chapter 3 will (hopefully) make clear, the root of statistical inference is knowledge of the distribution that would result if we (1) collected all possible samples of size n that could be drawn from the unobserved population of size N, (2) calculated the value of a given statistic (say, the sample mean) for each of those samples, and (3) placed those values in order on the number line to create a distribution known as a sampling distribution.

Visually, we depict the process of constructing a sampling distribution in Figure 1.9.

Notice that while we need to consider what a given sampling distribution would look like in theory, in practice, we will never put forth the extreme effort required to construct an entire sampling distribution. Instead, we will only collect a specific, independent random sample from the population—sample 1 as boxed in Figure 1.9—and combine information from that sample with our theoretical knowledge of what the sampling distribution would look like (were we to construct it) to learn about the population.

To aid understanding of this important concept, in Chapters 3 through 8 we specifically choose examples for which we can collect data for the entire population and, rather than analyzing those populations themselves, we select and analyze data for specific, independent random samples of given sizes drawn from those populations. We believe that this approach is advantageous for the following reason: Because we actually observe the entire population (which is extremely rare in practice), we have advanced knowledge

FIGURE 1.9 **A Visual Depiction of the Construction of a Sampling Distribution**



of the population parameters that we wish to estimate. We find this potentially valuable for two reasons: (1) In Chapter 3, we are able to start constructing a sampling distribution by drawing multiple, independent random samples from the population and placing calculated values for those different samples on the number line, and (2) In Chapters 3 through 8, whenever appropriate, we are able to compare sample statistics between the one independent random sample on which we focus and the known population parameters that they are intended to estimate—an approach that helps us consider the potential strengths and weaknesses of the statistical inference process.

It is important to note that we follow this approach for illustration purposes only. Beyond our desire to create a concrete example for introducing the concepts of sampling distributions and statistical inference, there is absolutely no reason to randomly select a sample from a known population (if we are lucky to have access to the entire population) or to randomly draw a smaller sample from a larger sample (if we are not lucky enough to have access to the entire population).

ADDITIONS TO OUR EMPIRICAL RESEARCH TOOLKIT

In this chapter, we have introduced a number of tools that will prove valuable when performing empirical research. They include:

- The individual steps to conducting an empirical research projects.
- The relationship between population, samples, and statistical inference.
- A general understanding of the importance of sampling distributions.

OUR NEW EMPIRICAL TOOLS IN PRACTICE: USING WHAT WE HAVE LEARNED IN THIS CHAPTER

On the basis of our introduction to the material to be covered in the course, most of our students find themselves anxious to get started learning this most valuable subject.

LOOKING AHEAD TO CHAPTER 2

As discussed earlier, there are generally five steps associated with conducting a successful empirical research project. While step 4—implementing appropriate empirical techniques, correctly interpreting results, and drawing appropriate conclusions based on the estimated results—is the focus of the majority of this textbook, before being able to perform such analysis, we must first identify a question of interest, develop appropriate theory, and identify and collect appropriate data. Chapter 2 discusses these three steps in more detail.

Problems

1.1 What is something you are interested in for which you could either directly collect data or survey individuals to obtain data? Explain how you would accomplish steps 1, 2, and 3 in conducting an empirical research project for this idea.

1.2 For each of the following, state if it is a population or a sample. Why?

- a. U.S. census
- b. Unemployment rate
- c. Consumer confidence
- d. Housing prices in New York City

1.3 Answer the following:

- a. Flip a coin 10 times. Put a 1 each time the coin comes up heads and a 0 each time the coin comes up tails. Count the number of heads you obtained and divide by 10. What number did you get? This is your estimate of the probability of obtaining heads.
- b. Is the number you obtained in part (a) a parameter or a statistic?
- c. Now flip the coin 25 times. Put a 1 each time you obtain a heads and a 0 for tails. Count the number of heads you obtained and divide by 25. What number did you get?
- d. What is the chance you obtain a heads in a fair coin flip? How do you believe this number is obtained using your results from parts (a) and (c)?
- e. How does this process relate to the idea of a sampling distribution?

1.4 Answer the following:

- a. Roll a die 10 times. Put a 1 each time the die is a one and a 0 each time the die comes up any other number. Count the number of ones you obtained and divide by 10. What number did you get? This is your estimate of the probability of obtaining a one.
- b. Is the number you obtained in part (a) a parameter or a statistic?
- c. Now roll the die 25 times. Put a 1 each time the die comes up as a one and a 0 for any other number. Count the number of ones you obtained and divide by 25. What number did you get?
- d. What is the chance you obtain a one with a fair die? How do you believe this number is obtained using your results from parts (a) and (c)?
- e. How does this process relate to the idea of a sampling distribution?