CHAPTER 7

Sampling Design

Learning Objectives

After reading this chapter, you should understand . . .

1 The two premises on which sampling theory is based.

2 The characteristics of accuracy and precision for measuring sample validity.

3 The two categories of sampling techniques and the variety of sampling techniques within each category.

4 The six questions that must be answered to develop a sampling plan.

5 The critical issues and formulas that determine the appropriate sample size.

6 The various sampling techniques and when each is used.

Bringing Research to Life

From 100 feet away, Eric Burbidge—a recent hire by CityBus—saw that the approaching number 99 bus was filthy. Forty riders were docilely lined up to board—one was a letter carrier, and a collection of other blue-collar men and women comprised the rest. Some were idly thumbing through newspapers—Good! He'd learned in college that the ideal subject was thoughtful, articulate, rational, and above all, cooperative. Some were chatting, and two were ruminating slowly on hero sandwiches, which he supposed they had picked up at the tavern across the street from the bus station.

Burbidge swept past the queued passengers, taking care not to make premature eye contact, and brusquely rapped his clipboard against the bus's folding door. With a whoosh the driver snapped the door open, and Burbidge heaved himself up onto the bus. "Good evening, driver. I'm Eric Burbidge, from headquarters."

"Figures. We don't see many suits on Route 99."

"I'm here to conduct a scientific survey to determine newspaper readership, if any, of riders on this route."

"Yessir," murmured the driver, without enthusiasm.

"You must know from your employees' newsletter that the corporation is soon to announce a restructuring of the route system and schedule, pursuant to which we shall have to purchase advertising space in the leading media to reveal our new route structure, maps, and schedules."

"Won't that be a mess," muttered the driver, in the same flat tone. "And, yessir, I read your newsletter. For 15 years I have read it."

"You may have noticed that this route runs north-to-south equidistant between the twin cities . . ."

"The route runs north-south, with one city on the east and the other on the west, yessir. I caught on to that long ago, sir. Clean splits the line between East City and West City and never gets closer to one than the other. If you want to get to one city or the other, you transfer to another bus. My bus just goes straight north. Just like it shows on the route map, sir."

"As corporate research director," said Burbidge, gracefully emphasizing his point by pointing a forefinger roofward, "I take nothing for granted. I need scientific methodology to test my hypothesis that readership of newspapers on this route would be equally divided between the *East City Gazette* and the *West City Tribune.* To that very end I have prepared this survey for your riders, which with your complete cooperation I have selected to pass out along this route, on this day, at this time."

"Newspapers, you say," said the driver, showing a flicker of involvement. "I could tell you quite a lot about those newspapers . . ."

"Not required! This is a scientific survey of riders on Route 99, and anecdotal information is not appropriate. What I require of you is to let the passengers up onto the bus so I may give them these pencils and surveys, and then to refrain from swerving or unnecessarily agitating the bus while they are filling out my surveys. Do you think you can manage that?"

"I'd better turn on the inside lights, don't you think?"

"Well, of course, turn on the lights. That goes without saying."

The passengers boarded slowly, exchanging pleasantries with the driver. They were quite a little clan, Burbidge could see.

The bus rolled steadily northward, and Burbidge was pleased to see that riders hunched over his one-page survey, though he was struck by how long it took them to answer the simplest of questions. As each completed the survey, he or she shyly shuffled forward and proffered it to Burbidge without comment.

Just when he sensed everything was going as well as might be expected, two men stood up, one in the front by the driver, one in the rear, and spread their legs to straddle the aisle. A ball of newsprint was tossed into the aisle, and the passengers began whooping and batting the paper forward and back. "What is this, driver?" demanded Burbidge.

"Hockey. They are playing hockey. The idea is to knock the ball between this guy's legs or that guy's in the back."

"Aren't you going to stop them?"

"No harm done. These are a friendly bunch. Big sports fans. In fact," said the driver, whose voice had gained enthusiasm, "the East City Club plays pro hockey tonight. So, when I clean out the bus, most of the newspapers will be the *East City Gazette.*" He rattled on, contributing to Burbidge's annoyance, explaining that the riders liked to study the night's pro game in advance, the better to discuss it among themselves, so newsstand sales were brisk in the terminal, but only for the newspaper that did the better job of covering the sport du jour. "Of course, tomorrow night there is pro basketball in West City, and most of the riders will pick up the *West City Tribune* at the newsstands."

"That's impossible to accept," shouted Burbidge. "Such behavior would bias my scientific survey, which asks for the paper they most recently purchased!"

"Well, you had better accept it. 'Cause I've been cleaning out this bus for five years. It's the *Gazette* before hockey, and the *Trib* before basketball. This is a hockey crowd tonight, which means extra *Gazettes*."

Burbidge was mortified and hoped he wasn't revealing his distress to the driver. But the driver added, "Course, in the mornings these folks bring the papers that the newsboys toss on their lawns, so you don't see such a situation."

"By choosing between sampling this route—morning or evening—I get a systematically different set of results," mumbled Burbidge, more to himself than to the driver. "And by choosing a hockey night or a basketball night, I further distort the results."

"Naturally," agreed the driver, with irritating enthusiasm.

The driver had now warmed to his exposition of trash-can sociological research. "Of course, by reading our company newsletter, as I do, I know that by the time you announce the new routes and schedules, we will be finished with hockey and basketball and into the baseball season." Could Burbidge detect irony in the driver's chuckle?

"If you are furnishing me information that is at all reliable, driver, then the generalizability of my survey would be contingent on whether I sample in the morning or evening, on the professional sports schedule on a particular night, and on the season of the year."

"That's part of it, yessir."

"Part of it?"

"Well, yessir . . . you see, most of these folks on the 5:15 bus are East City folks, and most of the people on the 5:45 bus are West City folks, so your outcomes would naturally depend on whether you took a survey on the 5:15 or 5:45 bus."

"How can this be?"

"Well, you see, the 5:15 bus on this route—the one you are on now—gets to Boght Corners at 5:55, and the East City folks transfer onto the eastbound bus that is waiting there for them. Most of the West City folks hang out in the bar across the street from the terminal, get on the 5:45 bus, and rendezvous at 6:25 with the westbound bus at Boght Corners."

"I see. I see." He was reconsidering the wisdom of this survey. "AND IS THERE ANYTHING ELSE you would care to share with me?"

"Only that the 5:45 boarders don't read the newspaper much at all, as they have been watching sports on the TV in the bar, are lightly soused, some of them, and can't read the small print because I don't turn on the inside lights."

"I might as well have stayed home," Burbidge muttered in despair.

The driver stopped the bus and swiveled to face Burbidge. "Wouldn't that have been a pity, sir, as I would have been deprived of this excellent lesson in scientific research."

The Nature of Sampling

Most people intuitively understand the idea of sampling. One taste from a drink tells us whether it is sweet or sour. If we select a few employment records out of a complete set, we usually assume our selection reflects the characteristics of the full set. If some of our staff favors a flexible work schedule, we infer that others will also. These examples vary in their representativeness, but each is a sample.

The basic idea of **sampling** is that by selecting some of the elements in a population, we may draw conclusions about the entire population. A **population element** is the subject on which the measurement is being taken. It is the unit of study. While an element may be a person, it can just as easily be something else. For example, each office worker questioned about a flexible work schedule is a population element, and each business account analyzed is an element of an account population. A **population** is the total collection of elements about which we wish to make some inferences. All office workers in the firm compose a population of interest; all 4,000 files define a population of interest. A **census** is a count of all the elements in a population. If 4,000 files define the population, a census would obtain information from every one of them.

In studying customer satisfaction with the service operation for MindWriter, the population of interest to Jason and Myra is all individuals who have had a laptop repaired while the CompleteCare program has been in effect. The population element is any one individual interacting with the service program.

Why Sample?

There are several compelling reasons for sampling, including: (1) lower cost, (2) greater accuracy of results, (3) greater speed of data collection, and (4) availability of population elements.

Lower Cost The economic advantages of taking a sample rather than a census are massive. Consider the cost of taking a census. In 2000, due to a Supreme Court ruling requiring a census rather than statistical sampling techniques, the U.S. Bureau of the Census increased its 2000 Decennial Census budget estimate by $1.723 billion to $4.512 billion.[1] Is it any wonder that researchers in all types of organizations ask, "Why should we spend thousands of dollars interviewing all 4,000 employees in our company if we can find out what we need to know by asking only a few hundred?"

Greater Accuracy of Results Deming argues that the quality of a study is often better with sampling than with a census. He suggests, "Sampling possesses the possibility of better interviewing (testing), more thorough investigation of missing, wrong, or suspicious information, better supervision, and better processing than is possible with complete coverage."[2] Research findings substantiate this opinion. More than 90 percent of the total survey error in one study was from nonsampling sources and only

10 percent or less was from random sampling error.[3] The U.S. Bureau of the Census shows its confidence in sampling by taking sample surveys to check the accuracy of its census. And while it is politically correct to take a census of the population, we know that segments of the population are seriously undercounted.

Greater Speed of Data Collection Sampling's speed of execution reduces the time between the recognition of a need for information and the availability of that information. For every disgruntled customer that the MindWriter CompleteCare program generates, several prospective customers will move away from MindWriter to a competitor's laptop. So fixing the problems within the CompleteCare program will not only keep current customers coming back but will also discourage prospective customers from defecting to competitive brands due to negative word-of-mouth.

Availability of Population Elements Some situations require sampling. When we test the breaking strength of materials, we must destroy them; a census would mean complete destruction of all materials. Sampling is also the only process possible if the population is infinite.

SNAPSHOT Terrorism and the News

News organizations often conduct or commission polls to determine whether stories are potentially newsworthy or to get a sense of public opinion regarding some issue. Every network news broadcast was gathering "man-on-the-street" reactions (nonprobability samples) on Tuesday, September 11, 2001, immediately following the terrorist attacks on the World Trade Center and the Pentagon. However, given the magnitude of the story, a probability study was needed to understand the depth of various issues. Gallup conducted one such telephone poll with a national probability sample of U.S. adults, with a margin of error of ±4 percent. Commissioned by CNN and *USA Today,* Gallup was to assess "American support for the general idea of military action against the groups or nations responsible." In response, 92 percent of Americans supported some type of military action. Simultaneously, two polls commissioned by ABC News and the *Washington Post,* conducted by TNS Intersearch on September 11–12, indicated 93 percent and 94 percent support for military action.

In all of these polls, Americans also said they were willing to bide their time, finding the right time and place to retaliate. In the poll conducted by TNS Intersearch, the support was strong, even when respondents were apprised of the potential consequences of the action. An NBC News/*Wall Street Journal* poll conducted by Hart/Teeter on September 13 showed 83 percent of adults supported "forceful military action." Each of these polls surveyed between 556 and 618 adults with precision factors of ±4 to ±4.5 percent. A review of polls dating back to 1986 in the ABC News poll vault shows that early support for military action following an incident tends to hold through the duration of military action. However, the basis for such responses, going back to the retaliation in Libya, has been Americans' recollection of prompt, tactical, and decisive strikes. In a prolonged campaign, with accumulating American casualties and the likelihood of continued urban terror, support for continued engagement may wane. Only history will tell if citizen support for a military response to this historic terrorist attack will demonstrate the same resolve.

www.abcnews.com

www.msnbc.com

www.cnn.com

www.wsj.com

www.usatoday.com

www.washingtonpost.com

www.intersearch.tnsofres.com

www.hartresearch.com

www.gallup.com

Sample versus Census The advantages of sampling over census studies are less compelling when the population is small and the variability within the population is high. Two conditions are appropriate for a census study: A census is (1) *feasible* when the population is small and (2) *necessary* when the elements are quite different from each other.[4] When the population is small and variable, any sample we draw may not be representative of the population from which it is drawn. The resulting values we calculate from the sample are incorrect as estimates of the population values. Consider North American manufacturers of stereo components. Fewer than 50 companies design, develop, and manufacture amplifier and loudspeaker products at the high end of the price range. The size of this population suggests a census is feasible. The diversity of their product offerings makes it difficult to accurately sample from this group. Some companies specialize in speakers, some in amplifier technology, and others in compact disk transports. Choosing a census in this situation is appropriate.

What Is a Good Sample?

The ultimate test of a sample design is how well it represents the characteristics of the population it purports to represent. In measurement terms, the sample must be valid. Validity of a sample depends on two considerations: accuracy and precision.

Accuracy Accuracy is the degree to which bias is absent from the sample. When the sample is drawn properly, some sample elements underestimate the population values being studied and others overestimate them. Variations in these values offset each other; this counteraction results in a sample value that is generally close to the population value. For these offsetting effects to occur, however, there must be enough elements in the sample, and they must be drawn in a way to favor neither overestimation nor underestimation.

An accurate (unbiased) sample is one in which the underestimators and the overestimators are balanced among the members of the sample. There is no **systematic variance** with an accurate sample. Systematic variance has been defined as "the variation in measures due to some known or unknown influences that 'cause' the scores to lean in one direction more than another."[5] Homes on the corner of the block, for example, are often larger and more valuable than those within the block. Thus, a sample that selects corner homes only will cause us to overestimate home values in the area. Burbidge learned that in selecting Route 99 for his newspaper readership sample, the time of the day, day of the week, and season of the year of the survey dramatically reduced the accuracy and validity of his sample.

Even the large size of some samples cannot counteract systematic bias. The classic example of a sample with systematic variance was the *Literary Digest* presidential election poll in 1936, in which more than 2 million persons participated. The poll predicted Alfred Landon would defeat Franklin Roosevelt for the presidency of the United States. Your memory is correct; we've never had a president named Alfred Landon. We discovered later that the poll drew its sample from telephone owners who were in the middle and upper classes—at the time, the bastion of the Republican party—while Roosevelt appealed to the much larger working class that didn't own phones and typically voted for the Democratic party candidate.

Precision A second criterion of a good sample design is *precision of estimate.* No sample will fully represent its population in all respects. The numerical descriptors that describe samples may be expected to differ from those that describe populations because of random fluctuations inherent in the sampling process. This is called **sampling error** and reflects the influences of chance in drawing the sample members. Sampling error is what is left after all known sources of systematic variance have been accounted for. In theory, sampling error consists of random fluctuations only, although some unknown systematic variance may be included when too many or too few sample elements possess a particular characteristic. Let's say we draw a sample from an alphabetical list of MindWriter owners who are having their laptops currently serviced by the Complete-Care program. We insert a survey response card in a sample of returned laptops. Eighty percent of those surveyed had their laptops serviced by Max Jensen. And we know from our exploratory study that Max had more complaint letters about his work than any other technician. We have failed to truly randomize the sample with our alphabetical listing of possible sample elements and thus have increased our sampling error.

Precision is measured by the *standard error of estimate,* a type of standard deviation measurement; the smaller the standard error of estimate, the higher is the precision of the sample. The ideal sample design produces a small standard error of estimate. However, not all types of sample design provide estimates of precision, and samples of the same size can produce different amounts of error variance.

Types of Sample Design

The researcher makes several decisions when designing a sample. These are represented in Exhibit 7–1. The sampling decisions flow from two decisions made in the formation of the management-research question hierarchy: the nature of the management question and the specific investigative questions that evolve from the research question.

A variety of sampling techniques is available. The one the researcher should select depends on the requirements of the project, its objectives, and the funds available. In the discussion that follows, we will use three examples:

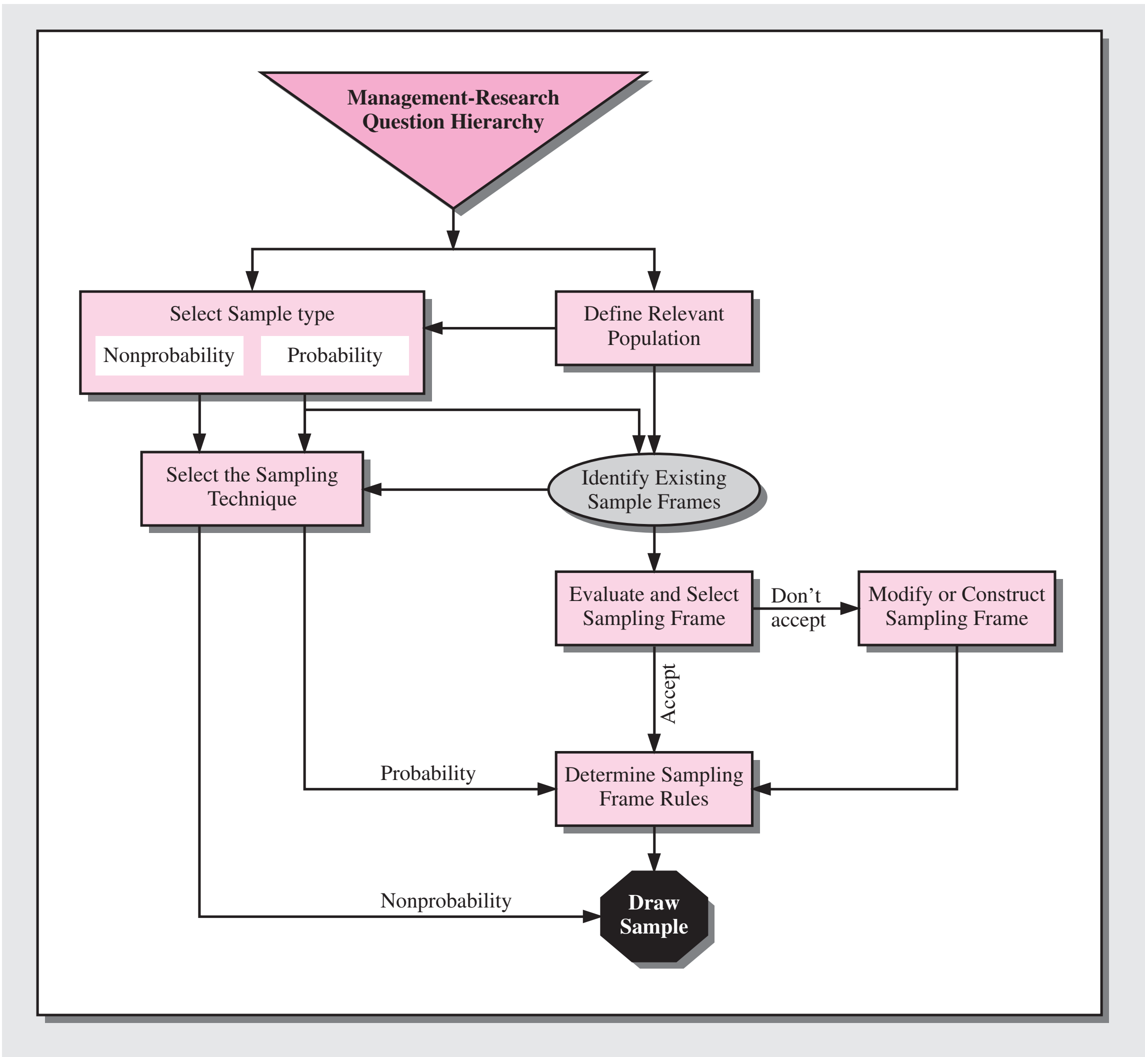
- The CityBus study introduced in the vignette at the beginning of this chapter.
- The continuing MindWriter service satisfaction study.
- A study of the feasibility of starting a dining club near the campus of Metro University.

The researchers at Metro U are exploring the feasibility of creating a dining club whose facilities would be available on a membership basis. To launch this venture, they will need to make a substantial investment. Research will allow them to reduce many risks. Thus the research question is: "Would a membership dining club be a viable enterprise?" Some investigative questions that flow from the research question include:

1. Who would patronize the club and on what basis?
2. How many would join the club under various membership and fee arrangements?
3. How much would the average member spend per month?
4. What days would be most popular?
5. What menu and service formats would be most desirable?
6. What lunch times would be most popular?
7. Given the proposed price levels, how often per month would each member have lunch or dinner?
8. What percent of the people in the population say they would join the club, based on the projected rates and services?

We use the last three investigative questions for examples and focus specifically on questions 7 and 8 for assessing the project's risks. First we will digress with other

EXHIBIT 7–1 Sampling Design within the Research Process



information and examples on sample design, coming back to Metro U in the next section.

In decisions of sample design, the representation basis and the element selection techniques, as shown in Exhibit 7–2, classify the different approaches.

Representation The members of a sample are selected on a probability basis or by another means. **Probability sampling** is based on the concept of *random selection*—a controlled procedure that assures that each population element is given a known nonzero chance of selection.

In contrast, **nonprobability sampling** is arbitrary (nonrandom) and subjective. Each member does not have a known nonzero chance of being included. Allowing

EXHIBIT 7–2 **Types of Sampling Designs**

	Representation Basis	
Element Selection	**Probability**	**Nonprobability**
Unrestricted	Simple random	Convenience
Restricted	Complex random	Purposive
	Systematic	Judgment
	Cluster	Quota
	Stratified	Snowball
	Double	

interviewers to choose sample elements "at random" (meaning "as they wish" or "wherever they find them") is not random sampling. Only probability samples provide estimates of precision. While we are not told how Burbidge selected the riders of Bus 99 as his sample, it's clear that he did not use probability sampling techniques.

Element Selection Whether the elements are selected individually and directly from the population—viewed as a single pool—or when additional controls are imposed, element selection may also classify samples. If each sample element is drawn individually from the population at large, it is an *unrestricted sample*. Restricted sampling covers all other forms of sampling.

Probability Sampling

The unrestricted, simple random sample is the simplest form of probability sampling. Since all probability samples must provide a known nonzero chance of selection for each population element, the **simple random sample** is considered a special case in which each population element has a *known and equal chance* of selection. In this section, we use the simple random sample to build a foundation for understanding sampling procedures and choosing probability samples. Exhibit 7–3 provides an overview of the steps involved in choosing a random sample.

Steps in Sampling Design

There are several decisions to be made in securing a sample. Each requires unique information. While the questions presented here are sequential, an answer to one question often forces a revision to an earlier one. In this section we will consider the following:

1. What is the relevant population?
2. What are the parameters of interest?
3. What is the sampling frame?
4. What is the type of sample?
5. What size sample is needed?
6. How much will it cost?

EXHIBIT 7–3 How to Choose a Random Sample

Selecting a **random sample** is accomplished with the aid of computer software, a table of random numbers, or a calculator with a random number generator. Drawing slips out of a hat or ping-pong balls from a drum serves as an alternative *if every element in the sampling frame has an equal chance of selection.* Mixing the slips (or balls) and returning them between every selection ensures that every element is just as likely to be selected as any other.

A table of random numbers (such as Appendix F, Table F–10) is a practical solution when no software program is available. Random number tables contain digits that have no systematic organization. Whether you look at rows, columns, or diagonals, you will find neither sequence nor order. Table F–10 in Appendix F is arranged into 10 columns of five-digit strings, but this is solely for readability.

Assume the researchers want a special sample from a population of 95 elements. How will the researcher begin?

1. **Assign each element within the sampling frame a unique number** from 01 to 95.
2. **Identify a random start from the random number table** (drop a pencil point-first onto the table with closed eyes. Let's say the pencil dot lands on the eighth column from the left and 10 numbers down from the top of Table F–10, marking the five digits 05067).
3. **Determine how the digits in the random number table will be assigned to the sampling frame** to choose the specified sample size (researchers agree to read the first two digits in this column downward until 10 are selected).
4. **Select the sample elements from the sampling frame** (05, 27, 69, 94, 18, 61, 36, 85, 71, and 83 using the above process. The digit 94 appeared twice and the second instance was omitted; 00 was omitted because the sampling frame started with 01).

Other approaches to selecting digits are endless; Horizontally right to left, bottom to top, diagonally across columns, and so forth. Computer selection of a simple random sample will be more efficient for larger projects.

SNAPSHOT — The Right Sample for Studying Sex Education

The Henry J. Kaiser Family Foundation (KFF), "an independent philanthropy focusing on the major health care issues facing the nation," is "primarily an operating organization that develops and runs its own research and communications programs." It recently released a study of principals, teachers, students, and their parents that challenges the "convention that Americans are reluctant to have sexual health issues taught in school; [rather] the surveys show that most parents, along with educators and students themselves, would expand sex education courses and curriculum."

How do you conduct a study on this sensitive topic? KFF chose Princeton Survey Research Associates to do several series of phone surveys. Interviews with "313 principals, 1,001 teachers of sex education classes, [and] 1,501 pairs of students and [their] parents were conducted February 2 through May 23, 1999." The principals and teachers were recruited to represent "all public, middle junior, and senior high schools enrolling grades 7 through 12 in the continental United States. They were randomly and proportionately selected from a national database of public schools by type of school." Random digit dialing was used to identify households with children between 11–19 years of age who were enrolled in public schools in grades 7 through 12. Once the student was identified, the interviewer asked to speak with the male parent or guardian (followed by the female guardian, if the male guardian was unavailable). The parent was surveyed first, followed by the student, during the same contact if possible. "At least 15 attempts were made to complete an interview at every sample school or household." At the 95 percent confidence level, the error interval was ±3 for students-parents and teachers, and ±6 percent for principals. Participation rate was 54 percent for students-parents, 72 percent for teachers, and 41 percent for principals. What do you think of the sampling done for this study? You can link to a full study report from the KFF website.

www.kff.org

www.psra.com

What Is the Relevant Population? The definition of the population may be apparent from the management problem or the research question(s) but often it is not. Is the population for the dining club study at Metro University defined as "full-time day students on the main campus of Metro U"? Or should the population include "all persons employed at Metro U"? Or should townspeople who live in the neighborhood be included? Without knowing the target market chosen for the new venture, it is not obvious which of these is the appropriate sampling population.

There also may be confusion about whether the population consists of individuals, households, or families, or a combination of these. If a communication study needs to measure income, then the definition of the population element as individual or household can make quite a difference. In an observation study, a sample population might be nonpersonal: displays within a store or any ATM a bank owns or all single-family residential properties in a community. Good operational definitions are critical in choosing the relevant population.

Assume the Metro University Dining Club is to be solely for the students and employees on the main campus. The researchers might define the population as "all currently enrolled students and employees on the main campus of Metro U." However, this does not include family members. They may want to revise the definition to make it "current students and employees of Metro U, main campus, and their families."

In the nonprobability sample, Burbidge seems to have defined his relevant population as any rider of the CityBus system. He presumes he has an equal need to determine newspaper readership of both regular and infrequent CityBus riders, so that he might reach them with information about the new route structure, maps, and schedules. He can, however, easily reach regular riders by distributing information about the new routes via display racks on the bus for a period of time before the new routes are implemented. Infrequent riders, then, are the real population of interest of his newspaper readership study.

What Are the Parameters of Interest? **Population parameters** are summary descriptors (e.g., incidence proportion, mean, variance) of variables of interest in the population. **Sample statistics** are descriptors of the relevant variables computed from sample data. Sample statistics are used as estimators of population parameters. The sample statistics are the basis of our inferences about the population. Depending on how measurement questions are phrased, each may collect a different type of data (see Exhibit 7–4). Each different type of data also generates different sample statistics.

We discuss data types in greater detail in Chapter 8.

Sampling from pools of special respondents, such as doctors and other professionals with time-starved occupations, often involves hiring firms that specialize in such recruiting.

EXHIBIT 7–4 Parameter of Interest and Type of Data

Parameter of Interest	Type of Data	Example Scale
Attendance at a special event	Nominal	Participation in a promotion (yes, no)
Percent of patrons who order their steak cooked rare, regardless of health warnings	Ordinal	Doneness of meat proportion (well done, medium, rare)
Mean temperature of ideal vacation location	Interval	Temperature in degrees
Average number of store visits per month	Ratio	Actual number of store visits

When the variables of interest in the study are measured on interval or ratio scales, we use the sample mean to estimate the population mean and the sample standard deviation to estimate the population standard deviation. Asking Metro U affiliates to reveal their frequency of eating on or near campus (less than 5 times per week, greater than 5 but less than 10 times per week, or greater than 10 times per week) would provide an interval data estimator. In MindWriter, the rating of service by CompleteCare (excellent, good, fair, etc.) would be an example of an interval data estimator. Asking the CityBus riders about their number of days of ridership during the past seven days would result in ratio data.

We discuss proportion estimators in more detail later in this chapter.

When the variables of interest are measured on nominal or ordinal scales, we use the sample proportion of incidence to estimate the population proportion and the *pq* to estimate the population variance. The **population proportion of incidence** "is equal to the number of elements in the population belonging to the category of interest, divided by the total number of elements in the population."[6] Proportion measures are necessary for nominal data and are widely used for other measures as well. The most frequent proportion measure is the percentage. In the Metro U study, examples of nominal data are the proportion of a population that expresses interest in joining the club (for example, 30 percent; therefore *p* is equal to .3 and *q,* those not interested, equals .7) or the proportion of married students who report they now eat in restaurants at least five times a month. The CityBus study seeks to determine whether East City or West City has the most riders on Bus 99. MindWriter might want to know if men or women have experienced the most problems with laptop model 9000. These measures for CityBus and MindWriter would result in nominal data. Exhibit 7–5 indicates population parameters of interest for our three example studies.

There may also be important subgroups in the population about whom we would like to make estimates. For example, we might want to draw conclusions about the extent of dining club use that could be expected from married students versus single students, residential students versus commuter students, and so forth. Such questions have a strong impact on the nature of the sampling frame we accept, the design of the sample, and its size. Burbidge should be more interested in reaching infrequent rather than regular CityBus riders with the newspaper advertising he plans. And in the MindWriter study, Jason may be interested in comparing the responses of those who experienced poor service and those who experienced excellent service through the CompleteCare program.

EXHIBIT 7–5 Sample Population Parameters

Example	Population Parameter of Interest (type data)	Scale
CityBus	Frequency of ridership within 7 days (interval data)	More than 10 times, 6 to 10 times, 5 or fewer times
	Proportion of East City vs. West City riders (nominal data)	Actual percentage
MindWriter	Perceived quality of service (interval data)	Excellent, good, fair, poor
	Proportion of gender of Laptop 9000 customers with problems (nominal data)	Percent male, female
Metro U	Frequency of eating on or near campus within 7 days (ratio data)	Actual eating experiences
	Proportion of students/employees expressing interest (nominal data)	Actual percentage interested

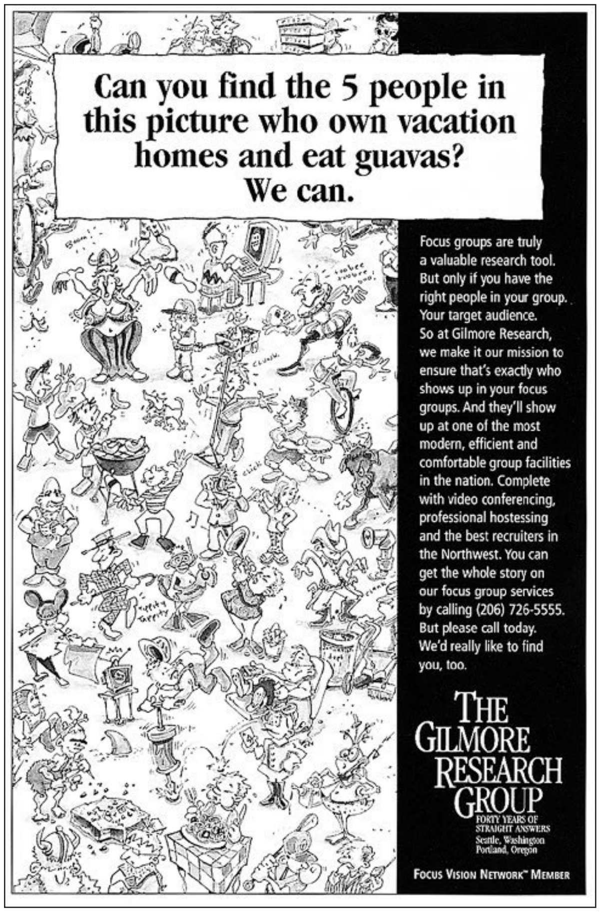
What Is the Sampling Frame? The **sampling frame** is closely related to the population. It is *the list of elements from which the sample is actually drawn.* Ideally, it is a complete and correct list of population members only. Jason should find limited problems obtaining a sampling frame of CompleteCare service users as MindWriter has maintained a database of all calls coming into the Call Center and all serial numbers of laptops serviced. As a practical matter, however, the sampling frame often differs from the theoretical population.

For the dining club study, the Metro U directory would be the logical first choice as a sampling frame. Directories are usually accurate when published in the fall, but suppose the study is being done in the spring. The directory will contain errors and omissions because some people will have withdrawn or left since the directory was published, while others will have enrolled or been hired. Usually university directories don't mention the families of students or employees. Just how much inaccuracy one can tolerate in choosing a sampling frame is a matter of judgment. You might use the directory anyway, ignoring the fact that it is not a fully accurate list. However, if the directory is a year old, the amount of error might be unacceptable. One way to make the sampling frame for the Metro U study more representative of the population would be to secure a supplemental list of the new students and employees as well as a list of the withdrawals and terminations from Metro U's registrar and human resources databases. You could then add and delete information from the original directory. Or, if their privacy policies permit, you might just request a current listing from each of these offices and use these lists as your sampling frame.

A greater distortion would be introduced if a branch campus population were included in the Metro U directory. This would be an example of a too-inclusive frame—that is, a frame that includes many elements other than the ones in which we are interested. A university directory that includes faculty and staff retirees is another example of a too-inclusive sampling frame.

We discuss screening procedures in Chapter 12.

Often you have to accept a sampling frame that includes people or cases beyond those in whom you are interested. You may have to use a telephone directory to draw a sample of business telephone numbers. Fortunately, this is easily resolved. You draw a sample from the larger population and then use a screening procedure to eliminate those who are not members of the group you wish to study.



The Gilmore Research Group takes a humorous approach to a nagging research problem: Getting the right people to participate in a research study—whether for a focus group, a pilot test, or a probability study. www.gilmore-research.com

The Metro U dining club survey is an example of a sampling frame problem that is readily solved. Often one finds this task much more of a challenge. Suppose you need to sample the members of an ethnic group, say, Asians residing in Little Rock, Arkansas. There is probably no directory of this population. While you may use the general city directory, sampling from this too-inclusive frame would be costly and inefficient as Asians represent only a small fraction of Little Rock's population. The screening task would be monumental. Since ethnic groups frequently cluster in certain neighborhoods, you might identify these areas of concentration and then use a reverse area telephone or city directory, which is organized by street address, to draw the sample. Burbidge had a definite problem as no sample frame of CityBus riders existed. While some regular riders used monthly passes, infrequent riders usually paid cash for their fares. It might have been possible for Burbidge to have anticipated this and to have developed over time a listing of customers. Bus drivers could have collected relevant contact information over a month, but the cost of contacting customers via phone or mail would have been much more expensive than the self-administered intercept approach Burbidge chose for data collection. One sampling frame available to Burbidge was a list of bus routes. This list would have allowed him to draw a probability sample using a cluster sampling technique. We discuss more complex sampling techniques later in this chapter.

What Is the Type of Sample? The researcher faces a basic choice: a probability or nonprobability sample. With a probability sample, a researcher can make probability-based confidence estimates of various parameters that cannot be made with nonprobability samples. Choosing a probability sampling technique has several consequences. A researcher must follow appropriate procedures, so that:

- Interviewers or others cannot modify the selections made.
- Only those selected elements from the original sampling frame are included.
- Substitutions are excluded except as clearly specified and controlled according to predetermined decision rules.

Despite all due care, the actual sample achieved will not match perfectly the sample that is originally drawn. Some people will refuse to participate, and others will be difficult, if not impossible, to find. The latter represent the well-known "not-at-home" problem and require that enough callbacks be made to ensure they are adequately represented in the sample.

With personnel records available at a university and a population that is geographically concentrated, a probability sampling method is possible in the dining club study. University directories are generally available, and the costs of using a simple random sample would not be great here. Then, too, since the researchers are thinking of a major investment in the dining club, they would like to be confident they have a representative sample. The same analysis holds true for MindWriter: A sample frame is readily available, making a probability sample possible and likely.

While the probability cluster sampling technique was available to him, it is obvious that Burbidge chose nonprobability sampling, arbitrarily choosing Bus 99 as a judgment sample and attempting to survey everyone riding the bus during the arbitrary times in which he chose to ride.

What Size Sample Is Needed? Much folklore surrounds this question. The most pervasive myths are summarized as follows:

- A sample must be large or it is not representative.
- A sample should bear some proportional relationship to the size of the population from which it is drawn.

In reality, how large a sample should be is a function of the variation in the population parameters under study and the estimating precision needed by the researcher. A sample of 400 may sometimes be appropriate, while a sample of more than 2,000 may be required in other circumstances; in another case, perhaps a sample of only 40 is needed.

Some principles that influence sample size include:

MANAGEMENT Tip

- The greater the dispersion or variance within the population, the larger the sample must be to provide estimation precision.
- The greater the desired precision of the estimate, the larger the sample must be.
- The narrower the interval range, the larger the sample must be.
- The higher the confidence level in the estimate, the larger the sample must be.
- The greater the number of subgroups of interest within a sample, the greater the sample size must be, as each subgroup must meet minimum sample size requirements.
- If the calculated sample size exceeds 5 percent of the population, sample size may be reduced without sacrificing precision.

Since researchers can never be 100 percent certain a sample reflects its population, they must decide how much precision they need. Precision is measured by (1) the interval range in which they would expect to find the parameter estimate and (2) the degree of confidence they wish to have in that estimate.

The size of the probability sample needed can be affected by the size of the population, but only when the sample size is large compared with the population. This so-called *finite adjustment factor* enters the calculation when the calculated sample is 5 percent or more of the population. The net effect of the adjustment is to reduce the size of the sample needed to achieve a given level of precision.[7]

Other considerations often weigh heavily on the sample size decision. The conditions under which the sample is being conducted may suggest that only certain sampling techniques are feasible. One type of sample may be inappropriate because we have no list of population elements and must therefore sample geographic units. This is what happened to Burbidge in his choice of Bus 99. Since various designs have differing statistical and economic efficiencies, the choice of design will also affect the size of the sample.

The researcher also may be interested in making estimates concerning various subgroups of the population; then the sample must be large enough for each of these subgroups to meet the desired level of precision. One achieves this in simple random sampling by making the total sample large enough to ensure that each critical subgroup meets the minimum size criterion. In more complex sampling procedures, the smaller subgroups are sampled more heavily, and then the parameter estimates drawn from these subgroups are weighted. Depending on the variety of causes of MindWriter complaints about repair service, Jason and Myra's sample size might be controlled by the number of models of laptops that MindWriter services or the actual problem serviced.

How Much Will It Cost? Cost considerations influence decisions about the size and type of sample and also the data collection methods. Almost all studies have some budgetary constraint, and this may encourage a researcher to use a nonprobability

SNAPSHOT — Sampling the Census

For the 2000 census, the U.S. Bureau of the Census proposed for the first time that the every-ten-year census substitute statistical sampling to help adjust for the undercounting of some population segments. This undercounting includes renters, as well as immigrant and minority groups, who have been less responsive to past census requests for information. In January 1999, by a 5–4 decision, the U.S. Supreme Court banned sampling in census results for the purpose of federal apportionment, but allowed it for counting population cohorts for business research purposes and for districting within states.

Before the 2010 census, Republicans hope to ban sampling altogether because undercounted groups traditionally vote as Democrats. However, Democrats could vote to change the wording of the Census Act as amended in 1976, thus incorporating sampling for all uses. Researchers are most concerned about whether the sampling procedure could incorporate errors in the results, rendering the results at least as flawed as the data now produced and making long-term comparative studies difficult.

Census results are used as a quality control in selecting units for sample cells. For the first time, the 2000 census allowed Americans to "identify themselves as more than one race, leaving the door open to seemingly endless combinations of racial and ethnic identities." According to early census releases, approximately 7 million people identify themselves as multiracial. While less than 3 percent of the overall U.S. population, these individuals are younger than the population as a whole—42 percent are under 18 years of age. Drawing a nationally representative sample will never be the same again. Visit the census site on the Internet to discover more.

www.census.gov

sample. Probability sample surveys incur list costs for sample frames, callback costs, and a variety of other costs that are not necessary when more haphazard or arbitrary methods are used. But when the data collection method is changed, the amount and type of data that can be obtained also change. Note the effect of a $2,000 budget on sampling considerations:

Simple random sampling: $25 per interview; 80 completed interviews.

Geographic cluster sampling: $20 per interview; 100 completed interviews.

Self-administered questionnaire: $12 per respondent; 167 completed instruments.

Telephone interviews: $10 per respondent; 200 completed interviews.[8]

For CityBus the cost of sampling riders' newspaper preferences to discover where to run the route-reconfiguration announcements must be significantly less than the cost of running ads in both East City and West City dailies. Thus the nonprobability judgment sampling procedure that Burbidge used was logical from a budget standpoint. The investment required to open the dining club at Metro U also justifies the more careful probability approach taken by the students. For MindWriter, an investment in CompleteCare has already been made; Myra needs to be highly confident that her recommendations to her supervising manager to change CompleteCare procedures and policies are on target and thoroughly supported by the data collected. These considerations justify MindWriter's probability sampling approach.

Complex Probability Sampling

Simple random sampling is often impractical. Reasons include (1) it requires a population list (sampling frame) that is often not available; (2) it fails to use all the information about a population, thus resulting in a design that may be wasteful; and (3) it may be expensive to implement in both time and money. These problems have led to the development of alternative designs that are superior to the simple random design in statistical and/or economic efficiency.

A more efficient sample in a statistical sense is one that provides a given precision (standard error of the mean or proportion) with a smaller sample size. A sample that is economically more efficient is one that provides a desired precision at a lower dollar cost. We achieve this with designs that enable us to lower the costs of data collecting, usually through reduced travel expense and interviewer time.

In the discussion that follows, four alternative probability sampling approaches are considered: (1) systematic sampling, (2) stratified sampling, (3) cluster sampling, and (4) double sampling.

Systematic Sampling

A versatile form of probability sampling is **systematic sampling.** In this approach, every *k*th element in the population is sampled, beginning with a random start of an element in the range of 1 to *k*. The *k*th element is determined by dividing the sample size into the population size to obtain the skip pattern applied to the sampling frame. The major advantage of systematic sampling is its simplicity and flexibility. It is easier to instruct field workers to choose the dwelling unit listed on every *k*th line of a listing sheet than it is to use a random numbers table. With systematic sampling, there is no need to number the entries in a large personnel file before drawing a sample. To draw a systematic sample merely do the following:

MANAGEMENT Tip

- Identify the total number of elements in the population.
- Identify the sampling ratio (k = total population size divided by size of the desired sample).
- Identify the random start.
- Draw a sample by choosing every kth entry.

Invoices or customer accounts can be sampled by using the last digit or a combination of digits of an invoice or customer account number. Time sampling is also easily accomplished. Systematic sampling would be an appropriate technique for MindWriter's CompleteCare program evaluation.

While systematic sampling has some theoretical problems, from a practical point of view it is usually treated as a simple random sample. When similar population elements are grouped within the sampling frame, systematic sampling is statistically more efficient than a simple random sample. This might occur if the listed elements are ordered chronologically, by size, by class, and so on. Under these conditions, the sample approaches a proportional stratified sample. The effect of this ordering is more pronounced on the results of cluster samples than for element samples and may call for a proportional stratified sampling formula.[9]

A concern with systematic sampling is the possible *periodicity* in the population that parallels the sampling ratio. In sampling days of the week, a 1 in 7 sampling ratio would give biased results. A less obvious case might involve a survey in an area of apartment houses where the typical pattern is eight apartments per building. Many systematic sampling fractions, such as 1 in 8, could easily oversample some types of apartments and undersample others. The only protection against this is constant vigilance by the researcher.

Another difficulty may arise when there is a *monotonic trend* in the population elements. That is, the population list varies from the smallest to the largest element or vice versa. Even a chronological list may have this effect if a measure has trended in one direction over time. Whether a systematic sample drawn under these conditions provides a biased estimate of the population mean or proportion depends on the initial random draw. Assume that a list of 2,000 commercial banks is created, arrayed from the largest to the smallest, from which a sample of 50 must be drawn for analysis. A sampling ratio of 1 to 40 (begun with a random start at 16) drawing every 40th bank would exclude the 15 largest banks and give a small-size bias to the findings. Ways to deal with this concern include:

MANAGEMENT Tip

- Randomize the population before sampling.
- Change the random start several times in the sampling process.
- Replicate a selection of different samples.

Stratified Sampling

Most populations can be segregated into several mutually exclusive subpopulations, or strata. The process by which the sample is constrained to include elements from each of the segments is called **stratified random sampling.** University students can be divided by their class level, school or major, gender, and so forth. After a population is divided into the appropriate strata, a simple random sample can be taken within each stratum. The sampling results can then be weighted and combined into appropriate population estimates.

There are three reasons why a researcher chooses a stratified random sample: (1) to increase a sample's statistical efficiency, (2) to provide adequate data for analyzing the various subpopulations, and (3) to enable different research methods and procedures to be used in different strata.[10]

Stratification is usually more efficient statistically than simple random sampling and at worst it is equal to it. With the ideal stratification, each stratum is homogeneous internally and heterogeneous with other strata. This might occur in a sample that includes members of several distinct ethnic groups. In this instance, stratification makes a pronounced improvement in statistical efficiency.

It is also useful when the researcher wants to study the characteristics of certain population subgroups. Thus, if one wishes to draw some conclusions about activities in the different classes of a student body, stratified sampling would be used. Stratification is also called for when different methods of data collection are applied in different parts of the population. This might occur when we survey company employees at the home office with one method but must use a different approach with employees scattered over the country.

If data are available on which to base a stratification decision, how shall we go about it?[11] The ideal stratification would be based on the primary variable under study. If the major concern is to learn how often per month patrons would use the dining club, then one would like to stratify on this expected number of use occasions. The only difficulty with this idea is that if we knew this information, we would not need to conduct the study. We must, therefore, pick a variable for stratifying that we believe will correlate with the frequency of club use per month, something like work or class schedule as an indication of when a sample element might be near campus at mealtime.

Researchers often have several important variables about which they want to draw conclusions. A reasonable approach is to seek some basis for stratification that correlates well with the major variables. It might be a single variable (class level), or it might be a compound variable (class by gender). In any event, we will have done a good stratifying job if the stratification base maximizes the difference among strata means and minimizes the within-stratum variances for the variables of major concern.

SNAPSHOT IRI's Wal-Mart Solution

2001 ushered in a new year in retailing information for U.S. consumer packaged goods (CPG) manufacturers. Prior to August 1, Information Resources, Inc. (IRI), which purchased point-of-sale data from Wal-Mart and resold it in numerous forms to manufacturers, was restricted by contract from providing a client manufacturer with specific competitors' data. When the contract expired, Wal-Mart decided it would no longer sell its point-of-sale data. So, what does a syndicated research provider do when it loses access to the data of the world's largest retailer? It changes its sampling design.

Beginning in September, IRI introduced its InfoScan® Advantage service to monitor Wal-Mart activity with consumer panel data. But IRI needed to change the design of its existing panel to better reflect Wal-Mart's customers. IRI expanded its panel almost 20 percent—to 65,000 households—added more Hispanic and African-American households, and drew more households from rural counties. While the data will not be as detailed as in the past, Ed Kuehnle, IRI's group president of IRI North American, claims "InfoScan® Advantage will provide IRI customers [CPG manufacturers like Johnson & Johnson, PepsiCo and Procter & Gamble] with the most comprehensive and in-depth information possible from the available data sources." IRI panelists record purchases with calculator-size scanners. IRI provides a view of food, drug, mass merchandise, and Wal-Mart outlets in one integrated database. "This intelligence enhances the ability of IRI clients to reduce risk in new product introductions, optimize marketing investments. and effectively execute [marketing plans] at retail."

www.infores.com

The more strata used, the closer you come to maximizing interstrata differences (differences between stratum) and minimizing intrastratum variances (differences within a given stratum). You must base the decision partially on the number of subpopulation groups about which you wish to draw separate conclusions. Costs of stratification also enter the decision. There is little to be gained in estimating population values when the number of strata exceeds six.[12]

The size of the strata samples is calculated with two pieces of information: (1) how large the total sample should be and (2) how the total sample should be allocated among strata. In deciding how to allocate a total sample among various strata, there are proportionate and disproportionate options.

Proportionate versus Disproportionate Sampling In **proportionate stratified sampling,** each stratum is properly represented so the sample drawn from it is proportionate to the stratum's share of the total population. This approach is more popular than any of the other stratified sampling procedures. Some reasons for this include:

- It has higher statistical efficiency than will a simple random sample.
- It is much easier to carry out than other stratifying methods.
- It provides a self-weighting sample; the population mean or proportion can be estimated simply by calculating the mean or proportion of all sample cases, eliminating the weighting of responses.

On the other hand, proportionate stratified samples often gain little in statistical efficiency if the strata measures and their variances are similar for the major variables under study.

Any stratification that departs from the proportionate relationship is **disproportionate.** There are several disproportionate allocation schemes. One type is a judgmentally determined disproportion based on the idea that each stratum is large enough to secure adequate confidence levels and interval range estimates for individual strata.

A researcher makes decisions regarding disproportionate sampling, however, by considering how a sample will be allocated among strata. One author states,

> In a given stratum, take a larger sample if the stratum is larger than other strata; the stratum is more variable internally; and sampling is cheaper in the stratum.[13]

If one uses these suggestions as a guide, it is possible to develop an optimal stratification scheme. When there is no difference in intrastratum variances and when the costs of sampling among strata are equal, the optimal design is a proportionate sample.

While disproportionate sampling is theoretically superior, there is some question as to whether it has wide applicability in a practical sense. If the differences in sampling costs or variances among strata are large, then disproportionate sampling is desirable. It has been suggested that "differences of several-fold are required to make disproportionate sampling worthwhile."[14]

The process for drawing a stratified sample is:

MANAGEMENT Tip

- Determine the variables to use for stratification.
- Determine the proportions of the stratification variables in the population.
- Select proportionate or disproportionate stratification based on project information needs and risks.
- Divide the sampling frame into separate frames for each stratum.
- Randomize the elements within each stratum's sampling frame.
- Follow random or systematic procedures to draw the sample.

Cluster Sampling

In a simple random sample, each population element is selected individually. The population can also be divided into groups of elements with some groups randomly selected for study. This is **cluster sampling.** Cluster sampling differs from stratified sampling in several ways.

Stratified Sampling	Cluster Sampling
1. We divide the population into a few subgroups, each with many elements in it. The subgroups are selected according to some criterion that is related to the variables under study.	1. We divide the population into many subgroups, each with a few elements in it. The subgroups are selected according to some criterion of ease or availability in data collection.
2. We try to secure homogeneity within subgroups and heterogeneity between subgroups.	2. We try to secure heterogeneity within subgroups and homogeneity between subgroups, but we usually get the reverse.
3. We randomly choose elements from within each subgroup.	3. We randomly choose a number of the subgroups, which we then typically study in depth.

When done properly, cluster sampling also provides an unbiased estimate of population parameters. Two conditions foster the use of cluster sampling: (1) the need for more economic efficiency than can be provided by simple random sampling and (2) the frequent unavailability of a practical sampling frame for individual elements.

Statistical efficiency for cluster samples is usually lower than for simple random samples chiefly because clusters are usually homogeneous. Families in the same block (a typical cluster) are often similar in social class, income level, ethnic origin, and so forth.

While statistical efficiency in most cluster sampling may be low, economic efficiency is often great enough to overcome this weakness. The criterion, then, is the net relative efficiency resulting from the trade-off between economic and statistical factors. It may take 690 interviews with a cluster design to give the same precision as 424 simple random interviews. But if it costs only $5 per interview in the cluster situation and $10 in the simple random case, the cluster sample is more attractive ($3,450 versus $4,240).

Area Sampling Much research involves populations that can be identified with some geographic area. When this occurs, it is possible to use **area sampling,** the most important form of cluster sampling. This method overcomes both the problems of high sampling cost and the unavailability of a practical sampling frame for individual elements. Area sampling methods have been applied to national populations, county populations, and even smaller areas where there are well-defined political or natural boundaries.

Suppose you want to survey the adult residents of a city. You would seldom be able to secure a listing of such individuals. It would be simple, however, to get a detailed city map that shows the blocks of the city. If you take a sample of these blocks, you are also taking a sample of the adult residents of the city.

Design In designing cluster samples, including area samples, we must answer several questions:

1. How homogeneous are the clusters?
2. Shall we seek equal or unequal clusters?
3. How large a cluster shall we take?
4. Shall we use a single-stage or multistage cluster?
5. How large a sample is needed?

1. Clusters are homogeneous. This contributes to low statistical efficiency. Sometimes one can improve this efficiency by constructing clusters to increase intracluster variance. In the dining club study, the students might have constructed clusters that included members from all classes. In area sampling, they could combine adjoining blocks that contain different income groups or social classes. Area cluster sections do not have to be contiguous, but the cost savings is lost if they are not near each other.

2. A cluster sample may be composed of clusters of equal or unequal size. The theory of clustering is that the means of sample clusters are unbiased estimates of the population mean. This is more often true when clusters are equal. It is often possible to construct artificial clusters that are approximately equal, but natural clusters, such as households in city blocks, often vary substantially. While one can deal with clusters of unequal size, it may be desirable to reduce or counteract the effects of unequal size. There are several approaches to this:

MANAGEMENT Tip

- Combine small clusters and split large clusters until each approximates an average size.
- Stratify clusters by size and choose clusters from each stratum.
- Stratify clusters by size and then subsample, using varying sampling fractions to secure an overall sampling ratio.

In this latter case, we may seek an overall sampling fraction of 1/60 and desire that subsamples contain five elements each. One group of clusters might average about 10 elements per cluster. In the "10 elements per cluster" stratum, we might choose 1 in 30 of the clusters and then subsample each chosen cluster at a 1/2 rate to secure the overall 1/60 sampling fraction. Among clusters of 120 elements, we might select clusters at a 1/3 rate and then subsample at a 1/20 rate to secure the 1/60 sampling fraction.[15]

3. The third question concerns the size of the cluster. There is no *a priori* answer to this question. Even with single-stage clusters, say, of 5, 20, or 50, it is not clear which size is superior. Some have found that in studies using single-stage clusters, the optimal cluster size is no larger than the typical city block.[16] Comparing the efficiency of the above three cluster sizes requires that we discover the different costs for each size and estimate the different variances of the cluster means.

4. The fourth question concerns whether to use a single-stage or a multistage cluster design. For most area sampling, especially large-scale studies, the tendency is to use multistage methods.

There are four reasons that justify subsampling, in preference to the direct creation of smaller clusters and their selection in one-stage cluster sampling:

- Natural clusters may exist as convenient sampling units yet may be larger than the desired economic size.
- We can avoid the cost of creating smaller clusters in the entire population and confine it to the selected sampling units.

- The effect of clustering . . . is often less in larger clusters. For example, a compact cluster of four dwellings from a city block may bring into the sample similar dwellings, perhaps from one building; but four dwellings selected separately can be spread around the dissimilar sides of the block.
- The sampling of compact clusters may present practical difficulties. For example, independent interviewing of all members of a household may seem impractical.[17]

5. The fifth question concerns how large a sample is needed—that is, how many subjects must be interviewed or observed. The answer to this question depends heavily on the specific cluster design, and these details can be complicated. Unequal clusters and multistage samples are the chief complications, and their statistical treatment is beyond the scope of this book.[18] Here we will treat only single-stage samples with equal-size clusters (called *simple cluster sampling*). It is analogous to simple random sampling. The simple random sample is really a special case of simple cluster sampling. We can think of a population as consisting of 20,000 clusters of one student each, or 2,000 clusters of 10 students each, and so on. The only difference between a simple random sample and a simple cluster sample is the size of cluster. Since this is so, we should expect that the calculation of a probability sample size would be the same for both types.

Double Sampling

It may be more convenient or economical to collect some information by sample and then use this information as the basis for selecting a subsample for further study. This procedure is called **double sampling, sequential sampling,** or **multiphase sampling.** It is usually found with stratified and/or cluster designs. The calculation procedures are described in more advanced texts.

Double sampling can be illustrated by the dining club example. You might use a telephone survey or another inexpensive survey method to discover who would be interested in joining such a club and the degree of their interest. You might then stratify the interested respondents by degree of interest and subsample among them for intensive interviewing on expected consumption patterns, reactions to various services, and so on. Whether it is more desirable to gather such information by one-stage or two-stage sampling depends largely on the relative costs of the two methods.

Because of the wide range of sampling designs available, it is often difficult to select an approach that meets the needs of the research question and helps to contain the costs of the project. To help with these choices, Exhibit 7–6 may be used to compare the various advantages and disadvantages of probability sampling. Nonprobability sampling techniques are covered in the next section. They are used frequently and offer the researcher the benefit of low cost. However, they are not based on a theoretical framework and do not operate from statistical theory; consequently, they produce selection bias and nonrepresentative samples. Despite these weaknesses, their widespread use demands their mention here.

Nonprobability Sampling

Any discussion of the relative merits of probability versus nonprobability sampling clearly shows the technical superiority of the former. In probability sampling, researchers use a random selection of elements to reduce or eliminate sampling bias. Under such conditions, we can have substantial confidence that the sample is representative of the population from which it is drawn. In addition, with probability sample

EXHIBIT 7–6 Comparison of Probability Sampling Designs

Type	Description	Advantages	Disadvantages
Simple random	Each population element has an equal chance of being selected into the sample. Sample drawn using random number table/generator.	Easy to implement with automatic dialing (random digit dialing) and with computerized voice response systems.	Requires a listing of population elements. Takes more time to implement. Uses larger sample sizes. Produces larger errors. Expensive.
Systematic	Selects an element of the population at a beginning with a random start and following the sampling fraction selects every *k*th element.	Simple to design. Easier to use than the simple random. Easy to determine sampling distribution of mean or proportion. Less expensive than simple random.	Periodicity within the population may skew the sample and results. If the population list has a monotonic trend, a biased estimate will result based on the start point.
Stratified	Divides population into subpopulations or strata and uses simple random on each strata. Results may be weighted and combined.	Researcher controls sample size in strata. Increased statistical efficiency. Provides data to represent and analyze subgroups. Enables use of different methods in strata.	Increased error will result if subgroups are selected at different rates. Expensive. Especially expensive if strata on the population have to be created.
Cluster	Population is divided into internally heterogeneous subgroups. Some are randomly selected for further study.	Provides an unbiased estimate of population parameters if properly done. Economically more efficient than simple random. Lowest cost per sample, especially with geographic clusters. Easy to do without a population list.	Often lower statistical efficiency (more error) due to subgroups being homogeneous rather than heterogeneous.
Double (sequential or multiphase)	Process includes collecting data from a sample using a previously defined technique. Based on the information found, a subsample is selected for further study.	May reduce costs if first stage results in enough data to stratify or cluster the population.	Increased costs if indiscriminately used.

designs, we can estimate an interval range within which the population parameter is expected to fall. Thus, we can not only reduce the chance for sampling error but also estimate the range of probable sampling error present.

With a subjective approach like nonprobability sampling, the probability of selecting population elements is unknown. There are a variety of ways to choose persons or cases to include in the sample. Often we allow the choice of subjects to be made by field workers on the scene. When this occurs, there is greater opportunity for bias to enter the sample selection procedure and to distort the findings of the study. Also, we cannot estimate any range within which to expect the population parameter. Given the technical advantages of probability sampling over nonprobability sampling, why would anyone choose the latter? There are some practical reasons for using these less precise methods.

Practical Considerations

We may use nonprobability sampling procedures because they satisfactorily meet the sampling objectives. While a random sample will give us a true cross section of the population, this may not be the objective of the research. If there is no desire or need to generalize to a population parameter, then there is much less concern about whether the sample fully reflects the population. Often researchers have more limited objectives. They may be looking only for the range of conditions or for examples of dramatic variations. This is especially true in exploratory research where one may wish to contact only certain persons or cases that are clearly atypical. Burbidge would have likely wanted a probability sample if the decision resting on the data was the actual design of the new CityBus routes and schedules. However, the decision of where and when to place advertising announcing the change is a relatively low-cost one in comparison.

Additional reasons for choosing nonprobability over probability sampling are cost and time. Probability sampling clearly calls for more planning and repeated callbacks to ensure that each selected sample member is contacted. These activities are expensive. Carefully controlled nonprobability sampling often seems to give acceptable results, so the investigator may not even consider probability sampling. Burbidge's results from Bus 99 would generate questionable data, but he seemed to realize the fallacy of many of his assumptions once he spoke with Bus 99's driver—something he should have done during exploration prior to designing the sampling plan.

While probability sampling may be superior in theory, there are breakdowns in its application. Even carefully stated random sampling procedures may be subject to careless application by the people involved. Thus, the ideal probability sampling may be only partially achieved because of the human element.

It is also possible that nonprobability sampling may be the only feasible alternative. The total population may not be available for study in certain cases. At the scene of a major event, it may be infeasible to even attempt to construct a probability sample. A study of past correspondence between two companies must use an arbitrary sample because the full correspondence is normally not available.

In another sense, those who are included in a sample may select themselves. In mail surveys, those who respond may not represent a true cross section of those who receive the questionnaire. The receivers of the questionnaire decide for themselves whether they will participate. There is some of this self-selection in almost all surveys because every respondent chooses whether or not to be interviewed.

Methods

Convenience

Nonprobability samples that are unrestricted are called **convenience samples.** They are the least reliable design but normally the cheapest and easiest to conduct. Researchers or field workers have the freedom to choose whomever they find, thus the name *convenience.* Examples include informal pools of friends and neighbors, people responding to a newspaper's invitation for readers to state their positions on some public issue or a TV reporter's "man-on-the-street" intercept interviews, or using employees to evaluate the taste of a new snack food.

While a convenience sample has no controls to ensure precision, it may still be a useful procedure. Often you will take such a sample to test ideas or even to gain ideas about a subject of interest. In the early stages of exploratory research, when you are seeking guidance, you might use this approach. The results may present evidence that is so overwhelming that a more sophisticated sampling procedure is unnecessary. In an interview with students concerning some issue of campus concern, you might talk to 25 students selected sequentially. You might discover that the responses are so overwhelmingly one-sided that there is no incentive to interview further.

Purposive Sampling A nonprobability sample that conforms to certain criteria is called *purposive sampling.* There are two major types—judgment sampling and quota sampling.

Judgment sampling occurs when a researcher selects sample members to conform to some criterion. In a study of labor problems, you may want to talk only with those who have experienced on-the-job discrimination. Another example of judgment sampling occurs when election results are predicted from only a few selected precincts that have been chosen because of their predictive record in past elections. Burbidge chose Bus 99 because the current route between East City and West City led him to believe that he could get a representation of both East City and West City riders.

When used in the early stages of an exploratory study, a judgment sample is appropriate. When one wishes to select a biased group for screening purposes, this sampling method is also a good choice. Companies often try out new product ideas on their employees. The rationale is that one would expect the firm's employees to be more favorably disposed toward a new product idea than the public. If the product does not pass this group, it does not have prospects for success in the general market.

Quota sampling is the second type of purposive sampling. We use it to improve representativeness. The logic behind quota sampling is that certain relevant characteristics describe the dimensions of the population. If a sample has the same distribution on these characteristics, then it is likely to be representative of the population regarding other variables on which we have no control. Suppose the student body of Metro U is 55 percent female and 45 percent male. The sampling quota would call for sampling students at a 55 to 45 percent ratio. This would eliminate distortions due to a nonrepresentative gender ratio. Burbidge could have improved his nonprobability sampling by considering time of day and day of week variations and choosing to distribute surveys to Bus 99 riders at various times, thus creating a quota sample.

In most quota samples, researchers specify more than one control dimension. Each should meet two tests: It should (1) have a distribution in the population that we can estimate, and (2) be pertinent to the topic studied. We may believe that responses to a question should vary, depending on the gender of the respondent. If so, we should seek proportional responses from both men and women. We may also feel that undergraduates differ from graduate students, so this would be a dimension. Other dimensions, such as the student's academic discipline, ethnic group, religious affiliation, and social group affiliation, also may be chosen. Only a few of these controls can be used. To illustrate, suppose we consider the following:

Gender: Two categories—male, female.

Class level: Two categories—graduate, undergraduate.

College: Six categories—Arts and Science, Agriculture, Architecture, Business, Engineering, other.

Religion: Four categories—Protestant, Catholic, Jewish, other.

Fraternal affiliation: Two categories—member, nonmember.

Family social-economic class: Three categories—upper, middle, lower.

In an extreme case, we might ask an interviewer to find a male undergraduate business student who is Catholic, a fraternity member, and from an upper-class home. All combinations of these six factors would call for 288 such cells to consider. This type of control is known as *precision control.* It gives greater assurance that a sample will be representative of the population. However, it is costly and too difficult to carry out with more than three variables.

When we wish to use more than three control dimensions, we should depend on *frequency control.* With this form of control, the overall percentage of those with each characteristic in the sample should match the percentage holding the same characteristic in the population. No attempt is made to find a combination of specific characteristics in a single person. In frequency control, we would probably find that the accompanying sample array is an adequate reflection of the population:

	Population	Sample
Male	65%	67%
Married	15	14
Undergraduate	70	72
Campus resident	30	28
Independent	75	73
Protestant	39	42

Quota sampling has several weaknesses. First, the idea that quotas on some variables assume a representativeness on others is argument by analogy. It gives no assurance that the sample is representative of the variables being studied. Often, the data used to provide controls may also be outdated or inaccurate. There is also a practical limit on the number of simultaneous controls that can be applied to ensure precision. Finally, the choice of subjects is left to field workers to make on a judgmental basis. They may choose only friendly looking people, people who are convenient to them, and so forth.

Despite the problems with quota sampling, it is widely used by opinion pollsters and marketing and other researchers. Probability sampling is usually much more costly and time-consuming. Advocates of quota sampling argue that while there is some danger of systematic bias, the risks are usually not that great. Where predictive validity has been checked (e.g., in election polls), quota sampling has been generally satisfactory.

Snowball This design has found a niche in recent years in applications where respondents are difficult to identify and are best located through referral networks.

In the initial stage of **snowball sampling,** individuals are discovered and may or may not be selected through probability methods. This group is then used to locate others who possess similar characteristics and who, in turn, identify others. Similar to a reverse search for bibliographic sources, the "snowball" gathers subjects as it rolls along. Various techniques are available for selecting a nonprobability snowball with provisions for error identification and statistical testing. Let's consider a brief example.

The high end of the U.S. audio market is composed of several small firms that produce ultra-expensive components used in recording and playback of live performances. A risky new technology for improving digital signal processing is being contemplated by one firm. Through its contacts with a select group of recording engineers and electronics designers, the first-stage sample may be identified for interviewing. Subsequent interviewees are likely to reveal critical information for product development and marketing.

Variations on snowball sampling have been used to study drug cultures, teenage gang activities, power elites, community relations, insider trading, and other applications where respondents are difficult to identify and contact.

Close-Up

Applying Concepts

In the Metro University Dining Club study, we explore probability sampling and the various concepts used to design the sampling process.

Exhibit 7–7 shows the Metro U Dining Club study population (N = 20,000) consisting of five subgroups based on their preferred lunch times. The values 1 through 5 represent the preferred lunch times of 11:00 A.M., 11:30 A.M., 12:00 noon, 12:30 P.M., and 1:00 P.M. The frequency of response (f) in the population distribution, shown beside the population subgroup, is what would be found if a census of the elements was taken. Normally, population data are unavailable or are too costly to obtain. We are pretending omniscience for the sake of the example.

Now assume we sample 10 elements from this population without knowledge of the population's characteristics. We use a sampling procedure from a statistical software program, a random number generator, or a table of random numbers. Our first sample (n_1 = 10) provides us with the frequencies shown below sample n_1 in Exhibit 7–7. We also calculate a mean score, X_1 = 3.0, for this sample. This mean would place the average preferred lunch time at 12:00 noon. The mean is a *point estimate* and our best predictor of the unknown population mean, μ (the arithmetic average of the population). Assume further that we return the first sample to the population and draw a second, third, and fourth sample by the same procedure. The frequencies, means, and standard deviations are as shown in the exhibit. As the data suggest, each sample shares some similarities with the population, but none is a perfect duplication because no sample perfectly replicates its population.

We cannot judge which estimate is the true mean (accurately reflects the population mean). However, we can estimate the interval in which the true μ will fall by using any of the samples. This is accomplished by using a formula that computes the **standard error of the mean.**

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

where

$\sigma_{\bar{X}}$ = Standard error of the mean or the standard deviation of all possible $\bar{X}$s

σ = Population standard deviation

n = Sample size

The standard error of the mean measures the standard deviation of the distribution of sample means. It varies directly with the standard deviation of the population from which it is drawn: If the standard deviation is reduced by 50 percent, the standard error will also be reduced by 50 percent. It also varies inversely with the square root of the sample size. If the square root of the sample size is doubled, the standard error is cut by one-half, provided the standard deviation remains constant.

EXHIBIT 7–7 Random Samples of Preferred Lunch Times

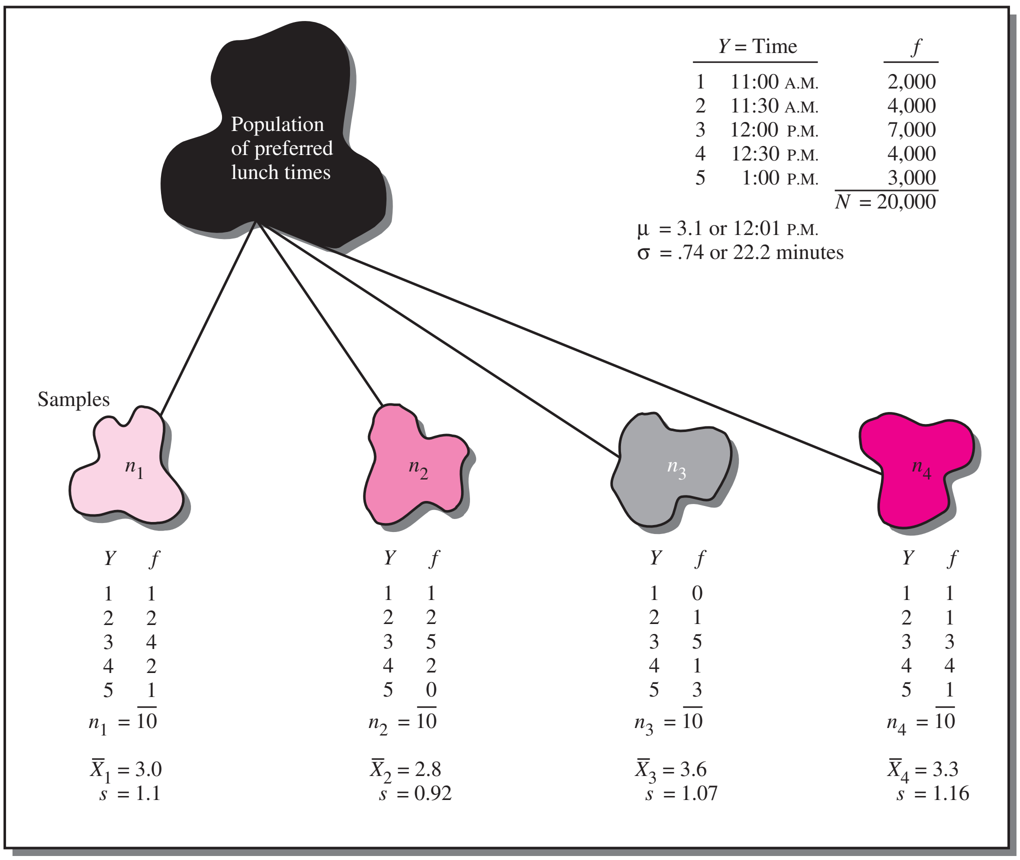
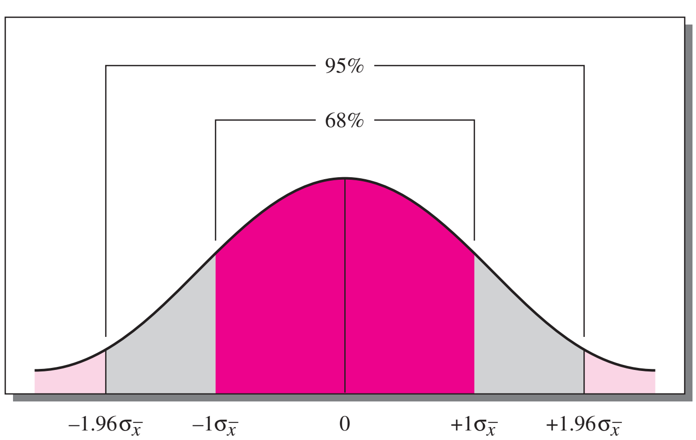


EXHIBIT 7–8 Confidence Levels and the Normal Curve



	Reducing the Standard Deviation	Doubling the Square Root of Sample Size
	$\sigma_{\bar{X}} = \frac{.74}{\sqrt{10}} = .234$	$\sigma_{\bar{X}} = \frac{.8}{\sqrt{25}} = .16$
$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}}$	$\sigma_{\bar{X}} = \frac{.37}{\sqrt{10}} = .117$	$\sigma_{\bar{X}} = \frac{.8}{\sqrt{100}} = .08$

Let's now examine what happens when we apply sample data (n_1) from Exhibit 7–7 to the formula. The sample standard deviation will be used as an unbiased estimator of the population standard deviation.

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}}$$

where

s = Standard deviation of the sample, n_1

$n_1 = 10$

$\bar{X}_1 = 3.0$

$s_1 = 1.15$

Substituting into the equation:

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{1.15}{\sqrt{10}} = .36$$

How does this improve our prediction of μ from $\bar{X}$? The standard error creates the interval range that brackets the point estimate. In this example, μ is predicted to be 3.0 or 12:00 noon (the mean of n_1) ± .36. This range may be visualized on a continuum:

We would expect to find the true μ between 2.64 and 3.36—between 11:49 A.M. and 12:11 P.M. (If 2 = 11:30 A.M. and .64 (30 minutes) = 19.2 minutes, then 2.64 = 11:30 A.M. + 19.2 minutes, or 11:49 A.M.) Since we assume omniscience for this illustration, we know the population average value is 3.1. Further, because standard errors have characteristics like other standard scores, we have 68 percent confidence in this estimate—that is, one standard error encompasses ±1 Z or 68 percent of the area under the normal curve (see Exhibit 7–8). Recall that the area under the curve also represents the confidence estimates that we make about our results. The combination of the interval range and the degree of confidence creates the **confidence interval.** To improve confidence to 95 percent, multiply the standard error of .36 by ±1.96 (Z), since 1.96 Z covers 95 percent of the area under the curve (see Exhibit 7–9). Now, with 95 percent confidence, the interval in which we would find the true mean increases to ±.70 (from 2.3 to 3.7 or from 11:39 A.M. to 12:21 A.M.).

Parenthetically, if we compute the standard deviation of the distribution of sample means [3.0, 2.8, 3.6, 3.3], we will discover it to be .35. Compare this to the standard error from the original calculation (.36). The result is consistent

True mean = 3.1

2.64 — 3.00 — 3.36

11:49 A.M. — $\bar{X}$ — 12:11 P.M.

EXHIBIT 7–9 Standard Errors Associated with Areas under the Normal Curve

Standard Error (Z)	Percent of Area*	Approximate Degree of Confidence
1.00	68.27	68%
1.65	90.10	90
1.96	95.00	95
3.00	99.73	99

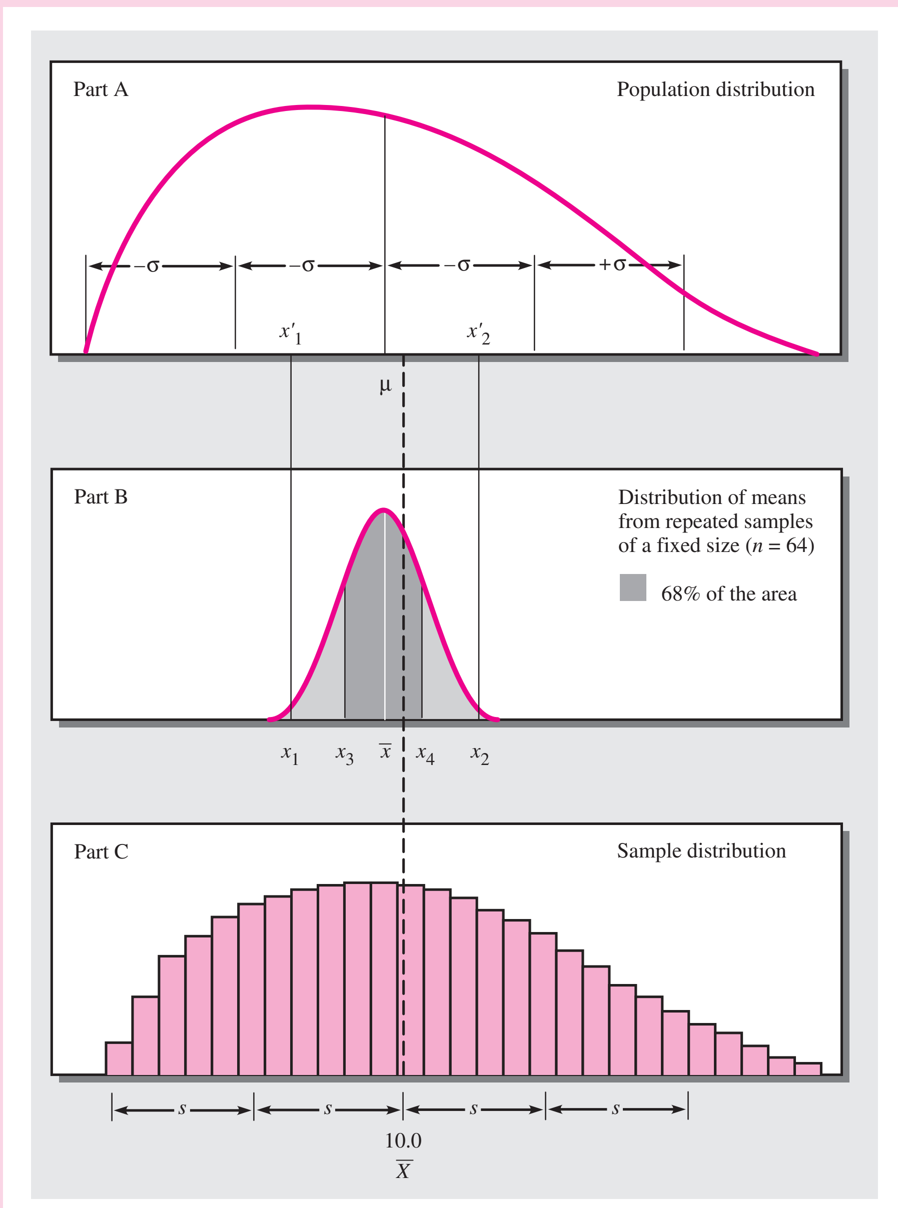
*Includes both tails in a normal distribution.

with the second definition of the standard error: the standard deviation of the distribution of sample means (n_1, n_2, n_3, and n_4). Now let's return to the dining club example and apply some of these concepts to the researchers' problem.

If the researchers were to interview all the students and employees in the defined population, asking them, "How many times per month would you eat at the club?" they would get a distribution something like that shown in Part A of Exhibit 7–10. The responses would range from zero to as many as 30 lunches per month with a μ and σ.

However, they cannot take a census, so μ and σ remain unknown. By sampling, the researchers find the

EXHIBIT 7–10 A Comparison of Population Distribution, Sample Distribution, and Distribution of Sample Means of Metro U Dining Club Study



Note: The distributions in these figures are not to scale, but this fact is not critical to our understanding of the dispersion relationship depicted.

mean to be 10.0 and the standard deviation to be 4.1 eating experiences (how often they would eat at the club per month). Turning to Part C of Exhibit 7–10, three observations about this sample distribution are consistent with our earlier illustration. First, it is shown as a histogram; it represents a frequency distribution of empirical data, while the smooth curve of Part A is a theoretical distribution. Second, the sample distribution (Part C) is similar in appearance but is not a perfect duplication of the population distribution (Part A). Third, the mean of the sample differs from the mean of the population.

If the researchers could draw repeated samples as we did earlier, they could plot the mean of each sample to secure the solid line distribution found in Part B. According to the **central limit theorem,** for sufficiently large samples ($n = 30$), the sample means will be distributed around the population mean approximately in a normal distribution. Even if the population is not normally distributed, the distribution of sample means will be normal if there is a large enough set of samples.

Estimating the Interval for the Metro U Dining Club Sample Any sample mean will fall within the range of the distribution extremes shown in Part B of Exhibit 7–10. We also know that about 68 percent of the sample means in this distribution will fall between x_3 and x_4 and 95 percent will fall between x_1 and x_2.

If we project points x_1 and x_2 up to the population distribution (Part A of Exhibit 7–10) at points x'_1 and x'_2, we see the interval where any given mean of a random sample of 64 is likely to fall 95 percent of the time. Since we will not know the population mean from which to measure the standard error, we infer that there is also a 95 percent chance that the population mean is within two standard errors of the sample mean (10.0). This inference enables us to find the sample mean, mark off an interval around it, and state a confidence likelihood that the population mean is within this bracket.

Because the researchers are considering an investment in this project, they would want some assurance that the population mean is close to the figure reported in any sample they take. To find out how close the population mean is to the sample mean, they must calculate the standard error of the mean and estimate an interval range within which the population mean is likely to be.

Given a sample size of 64, they still need a value for the standard error. Almost never will one have the value for the standard deviation of the population (σ), so we must use a proxy figure. The best proxy for σ is the standard deviation of the sample (s). Here the standard deviation ($s = 4.1$) was obtained from a pilot sample:

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{4.1}{\sqrt{64}} = .51$$

If one standard error of the mean is equal to 0.51 visits, then 1.96 standard errors (95 percent) are equal to 1.0 visit. The students can estimate with 95 percent confidence that the population mean of expected visits is within 10.0 ± 1.0 visit, or from 9.0 to 11.0 meal visits per month. We discuss pilot tests as part of the pretest phase in Chapter 12.

Changing Confidence Intervals The above estimate may not be satisfactory in two ways. First, it may not represent the degree of confidence the researchers want in the interval estimate, considering their financial risk. They might want a higher degree of confidence than the 95 percent level used here. By referring to a table of areas under the normal curve, they can find various other combinations of probability. Exhibit 7–11 summarizes some of those more commonly used. Thus, if the students want a greater confidence in the probability of including the population mean in the interval range, they can move to a higher standard error, say, $\bar{X} \pm 3\ \sigma_{\bar{X}}$. Now the population mean lies somewhere between 10.0 ± 3 (0.51) or from 8.47 to 11.53. With 99.73 percent confidence, we can say this interval will include the population mean.

We might wish to have an estimate that will hold for a much smaller range, for example, 10.0 ± 0.2. To secure this smaller interval range, we must either (1) accept a lower level of confidence in the results or (2) take a sample large enough to provide this smaller interval with the higher desired confidence level.

If one standard error is equal to 0.51 visits, then 0.2 visits would be equal to 0.39 standard errors (0.2/0.51 = .39). Referring to a table of areas under the normal curve (Appendix F, Table F–1), we find that there is a 30.3 percent chance that the true population mean lies within ± 0.39 standard errors of 10.0. With a sample of 64, the sample

EXHIBIT 7–11 Estimates Associated with Various Confidence Levels in the Metro U Dining Club Study

Approximate Degree of Confidence	Interval Range of Dining Visits per Month
68%	μ is between 9.48 and 10.52 visits
90	μ is between 9.14 and 10.86 visits
95	μ is between 8.98 and 11.02 visits
99	μ is between 8.44 and 11.56 visits

mean would be subject to so much error variance that only 30 percent of the time could the researchers expect to find the population mean between 9.8 and 10.2. This is such a low level of confidence that the researchers would normally move to the second alternative; they would increase the sample size until they could secure the desired interval estimate and degree of confidence.

Calculating the Sample Based on Critical Investigative Questions The researchers have selected two investigative question constructs as critical—"frequency of patronage" and "interest in joining"—because they believe both to be crucial to making the correct decision on the Metro U Dining Club opportunity. The first requires a point estimate, the second a proportion. By way of review, decisions needed and decisions made by Metro U researchers are summarized in Exhibit 7–12.

With reference to precision, the 95 percent confidence level is often used, but more or less confidence may be needed in light of the risks of any given project. Similarly, the size of the interval estimate for predicting the population parameter from the sample data should be decided. When a smaller interval is selected, the researcher is saying that precision is vital, largely because inherent risks are high. For example, on a five-point measurement scale, one-tenth of a point is a very high degree of precision in comparison to a one-point interval. Given that a patron could eat up to

EXHIBIT 7–12 Metro U Sampling Design Decisions on "Meal Frequency" and "Joining" Constructs

Decision Issues	Metro U Decisions: "Meal Frequency" (interval, ratio data)	Metro U Decisions: "Joining" (nominal, ordinal data)
1. The precision desired and how to quantify it:		
• The confidence researcher wants in the estimate (selected based on risk).	95% confidence ($Z = 1.96$)	95% confidence ($Z = 1.96$)
• The size of the interval estimate the researcher will accept (selected based on risk).	± .5 meals per month per person	± .10 (10 percent)
2. The expected dispersion in the population for the question used to measure precision:	0 to 30 meals	0 to 100%
• Sample mean.		
• Standard deviation.	4.1 meals	
• Sample proportion of population with the given attribute being measured.		30%
• Measure of the sample dispersion.		$pq = .30(1 - .30) = 0.21$
3. Whether a finite population adjustment should be used.	No	No
4. Estimate of standard deviation of population:		
• Standard error of mean.	.5/1.96 = 2.55	
• Standard error of the proportion.		.10/1.96 = 0.051
5. Sample size formula	Formula from page 209	Formula from page 210
6. Sample size	$n = 259$*	$n = 96$

*Because both investigative questions were of interest, the researcher would use the larger of the two sample sizes calculated, $n = 259$, for the study.

30 meals per month at the dining club (30 days times one meal per day), anything less than one meal per day would be asking for a high degree of precision in the Metro U study. The high risk of the Metro U study warrants the 0.5 meal precision selected.

The next factor that affects the size of the sample for a given level of precision is the population dispersion. The smaller the possible dispersion, the smaller will be the sample needed to give a representative picture of population members. If the population's number of meals ranges from 18 to 25, a smaller sample will give us an accurate estimate of the population's average meal consumption. However, with a population dispersion ranging from 0 to 30 meals consumed, a larger sample is needed for the same degree of confidence in the estimates. Since the true population dispersion of estimated meals per month eaten at Metro U Dining Club is unknowable, the standard deviation of the sample is used as a proxy figure. Typically, this figure is based on any of the following:

- Previous research on the topic.
- A pilot test or pretest of the data instrument among a sample drawn from the population.
- A rule of thumb (one-sixth of the range based on six standard deviations within 99.73 percent confidence).

If the range is from 0 to 30 meals, the rule-of-thumb method produces a standard deviation of five meals. The researchers want more precision than the rule-of-thumb method provides, so they take a pilot sample of 25 and find the standard deviation to be 4.1 meals.

A final factor affecting the size of a random sample is the size of the population. When the size of the sample exceeds 5 percent of the population, the finite limits of the population constrain the sample size needed. A correction factor is available in that event.

The sample size is computed for the first construct, meal frequency, as follows:

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}}$$

$$\sqrt{n} = \frac{s}{\sigma_{\bar{X}}}$$

$$n = \frac{s^2}{\sigma_{\bar{X}}}$$

$$n = \frac{(4.1)^2}{(.255)^2}$$

$$n = 258.5 \text{ or } 259$$

where

$\sigma_{\bar{X}} = 0.255\ (0.5/1.96)$

If the researchers are willing to accept a larger interval range (± 1 meal), and thus a larger amount of risk, then they can reduce the sample size to $n = 65$.

Calculating the Sample Size for the Proportions Question The second key question concerning the dining club study was: "What percentage of the population says it would join the dining club, based on the projected rates and services?" In business, we often deal with proportion data. An example is a CNN poll that projects the percentage of people who expect to vote for or against a proposition or a candidate. This is usually reported with a margin of error of ±5 percent.

In the Metro U study, a pretest answers this question using the same general procedure as before. But instead of the arithmetic mean, with proportions, it is p (the proportion of the population that has a given attribute)[19]—in this case, interest in joining the dining club. And instead of the standard deviation, dispersion is measured in terms of $p \times q$ (in which q is the proportion of the population not having the attribute, and $q = (1 - p)$). The measure of dispersion of the sample statistic also changes from the standard error of the mean to the standard error of the proportion σ_p.

We calculate a sample size based on this data by making the same two subjective decisions—deciding on an acceptable interval estimate and the degree of confidence. Assume that from a pilot test, 30 percent of the students and employees say they will join the dining club. We decide to estimate the true proportion in the population within 10 percentage points of this figure ($p = 0.30 \pm 0.10$). Assume further that we want to be 95 percent confident that the population parameter is within ± 0.10 of the sample proportion. The calculation of the sample size proceeds as before:

± 0.10 = Desired interval range within which the population proportion is expected (subjective decision).

$1.96\ \sigma_p$ = 95 percent confidence level for estimating the interval within which to expect the population proportion (subjective decision).

$\sigma_p = 0.051$ = Standard error of the proportion (0.10/1.96).

pq = Measure of sample dispersion (used here as an estimate of the population dispersion).

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

$$n = \frac{pq}{\sigma_p^2}$$

$$n = \frac{.03 \times .07}{(.051)^2}$$

$$n = 81$$

The sample size of 81 persons is based on an infinite population assumption. If the sample size is less than 5 percent of the population, there is little to be gained by using a finite population adjustment. The students interpreted the data found with a sample of 81 chosen randomly from the population as: "We can be 95 percent confident that 30 percent of the respondents would say they would join the dining club with a margin of error of ±10 percent."

Previously, the researchers used pilot testing to generate the variance estimate for the calculation. Suppose this is not an option. Proportions data have a feature concerning the variance that is not found with interval or ratio data. The *pq* ratio can never exceed 0.25. For example, if p = 0.5, then q = 0.5, and their product is 0.25. If either p or q is greater than 0.5, then their product is smaller than 0.25 (0.4 × 0.6 = 0.24, and so on). When we have no information regarding the probable p value, we can assume that p = 0.5 and solve for the sample size.

$$n = \frac{pq}{\sigma_p^2}$$

$$n = \frac{0.25}{(.051)^2}$$

$$n = 96$$

If we use this maximum variance estimate in the dining club example, we find the sample size needs to be 96 persons.

SUMMARY

1 Sampling is based on two premises. One is that there is enough similarity among the elements in a population that a few of these elements will adequately represent the characteristics of the total population. The second premise is that while some elements in a sample underestimate a population value, others overestimate this value. The result of these tendencies is that a sample statistic such as the arithmetic mean is generally a good estimate of a population mean.

2 A good sample has both accuracy and precision. An accurate sample is one in which there is little or no bias or systematic variance. A sample with adequate precision is one that has a sampling error that is within acceptable limits for the study's purpose.

3 A variety of sampling techniques is available. They may be classified by their representation basis and element selection techniques as shown in the accompanying table.

Probability sampling is based on random selection—a controlled procedure that ensures that each population element is given a known nonzero chance of selection. In contrast, nonprobability selection is "not random." When each sample element is drawn

	Representation Basis	
Element Selection	**Probability**	**Nonprobability**
Unrestricted	Simple random	Convenience
Restricted	Complex random	Purposive
	Systematic	Judgment
	Cluster	Quota
	Stratified	Snowball
	Double	

individually from the population at large, it is unrestricted sampling. Restricted sampling covers those forms of sampling in which the selection process follows more complex rules.

4 The simplest type of probability approach is simple random sampling. In this design, each member of the population has an equal chance of being included in a sample. In developing a probability sample, six procedural questions need to be answered:

1. What is the relevant population?
2. What are the parameters of interest?
3. What is the sampling frame?
4. What is the type of sample?
5. What size sample is needed?
6. How much will it cost?

Two kinds of estimates of a population parameter are made in probability sampling. First we make a point estimate that is the single best estimate of the population value. Then we make an interval estimate that covers the range of values within which we expect the population value to occur, with a given degree of confidence. All sample-based estimates of population parameters should be stated in terms of a confidence interval.

5 The specifications of the researcher and the nature of the population determine the size of a probability sample. These requirements are largely expressed in the following questions:

- What is the degree of confidence we want in our parameter estimate?
- How large an interval range will we accept?
- What is the degree of variance in the population?
- Is the population small enough that the sample should be adjusted for finite population?

Cost considerations are also often incorporated into the sample size decision.

6 Complex sampling is used when conditions make simple random samples impractical or uneconomical. The four major types of complex random sampling discussed in this chapter are systematic, stratified, cluster, and double sampling. Systematic sampling involves the selection of every *k*th element in the population, beginning with a random start between elements from 1 to *k*. Its simplicity in certain cases is its greatest value.

Stratified sampling is based on dividing a population into subpopulations and then randomly sampling from each of these strata. This method usually results in a smaller total sample size than would a simple random design. Stratified samples may be proportionate or disproportionate.

In cluster sampling, we divide the population into convenient groups and then randomly choose the groups to study. It is typically less efficient from a statistical viewpoint than the simple random because of the high degree of homogeneity within the clusters. Its great advantage is its savings in cost—if the population is dispersed geographically—or in time. The most widely used form of clustering is area sampling, in which geographic areas are the selection elements.

At times it may be more convenient or economical to collect some information by sample and then use it as a basis for selecting a subsample for further study. This procedure is called double sampling.

Nonprobability sampling also has some compelling practical advantages that account for its widespread use. Often probability sampling is not feasible because the population is not available. Then, too, frequent breakdowns in the application of probability sampling discount its technical advantages. You find also that a true cross section is often not the aim of the researcher. Here the goal may be the discovery of the range or extent of conditions. Finally, nonprobability sampling is usually less expensive to conduct than is probability sampling.

Convenience samples are the simplest and least reliable forms of nonprobability sampling. Their primary virtue is low cost. One purposive sample is the judgmental sample in which one is interested in studying only selected types of subjects. The other purposive sample is the quota sample. Subjects are selected to conform to certain predesignated control measures that secure a representative cross section of the population. Snowball sampling uses a referral approach to reach particularly hard-to-find respondents.

KEY TERMS

EXAMPLES

Company	Scenario	Page
ABC News	Commissioned TNS Intersearch to conduct numerous opinion polls following the September 11, 2001, attacks on the World Trade Center and the Pentagon.	180
CityBus*	A city transit authority is determining how to best promote a new route structure to current and potential riders.	BRTL and throughout
CNN	With *USA Today,* commissioned opinion polling after the September 11, 2001, attacks on the World Trade Center and the Pentagon.	180
Gallup Organization	Conducted opinion polling after the September 11, 2001, attacks on World Trade Center and the Pentagon for CNN and *USA Today.*	180
Hart/Teeter	Conducted opinion polling after the September 11, 2001, attacks on the World Trade Center and the Pentagon for NBC and *The Wall Street Journal.*	180

Henry J. Kaiser Family Foundation (KFF)	Conducted a benchmark study on U.S. public and private school sex education among teens, their parents, teachers, and principals.	185
Information Resources Inc. (IRI)	Syndicated research supplier to CPG manufacturers forced to redesign its sampling and research design when Wal-Mart chose not to renegotiate its contract to supply point-of-sale data; introduced InfoScan® Advantage.	194
Literary Digest	Published a 1936 presidential election poll that falsely predicted a Republican victory.	181
Metro University*	A study conducted to determine the feasibility of starting a membership-only dining club near campus.	182 and throughout
MindWriter*	A laptop manufacturer conducting a service program satisfaction study.	182 and throughout
NBC News	Commissioned Hart/Teeter to conduct opinion polls after the September 11, 2001, attacks on the World Trade Center and the Pentagon.	180
Princeton Survey Research Associates	The survey research organization commissioned by the Henry J. Kaiser Family Foundation for its study on sex education in America.	185
TaylorNelson Sofres (TNS) Intersearch	Conducted opinion polling after the September 11, 2001, attacks on the World Trade Center and the Pentagon for ABC News and *The Washington Post.*	180
U.S. Bureau of the Census	Substituting statistical sampling in the decennial census.	179, 191
USA Today	Commissioned Gallup to conduct opinion polling after the September 11, 2001, attacks on the World Trade Center and the Pentagon.	180
The Wall Street Journal	With NBC News, commissioned Hart/Teeter to conduct opinion polling after the September 11, 2001, attacks on the World Trade Center and the Pentagon.	180
The Washington Post	Commissioned TNS Intersearch to conduct opinion polling following the September 11, 2001, attacks on the World Trade Center and the Pentagon.	180

*Due to the confidential and proprietary nature of most research, the names of some companies have been changed.

DISCUSSION QUESTIONS

Terms in Review

1. Distinguish between

a. Statistic and parameter.

b. Sample frame and population.

c. Restricted and unrestricted sampling.

d. Standard deviation and standard error.

e. Simple random and complex random sampling.

f. Convenience and purposive sampling.

g. Sample precision and sample accuracy.

h. Systematic and error variance.

i. Variable and attribute parameters.

j. Point estimate and interval estimate.

k. Proportionate and disproportionate samples.

2. Under what kind of conditions would you recommend

a. A probability sample? A nonprobability sample?

b. A simple random sample? A cluster sample? A stratified sample?

c. Using the finite population adjustment factor?

d. A disproportionate stratified probability sample?

3. You plan to conduct a survey using unrestricted sampling. What subjective decisions must you make?

4. You draw a random sample of 300 employee records from the personnel file and find that the average years of service per employee is 6.3, with a standard deviation of 3.0 years.

a. What percentage of the workers would you expect to have more than 9.3 years of service?

b. What percentage would you expect to have more than 5.0 years of service?

Making Research Decisions

5. Your task is to interview a representative sample of attendees for the large concert venue where you work. The new season schedule includes 200 live concerts featuring all types of musicians and musical groups. Since neither the number of attendees nor their descriptive characteristics are known in advance, you decide on nonprobability sampling. Based on past seating configurations, you can calculate the number of tickets that will be available for each of the 200 concerts. Thus, collectively, you will know the number of possible attendees for each type of music. From attendance research conducted at concerts held during the previous two years, you can obtain gender data on attendees by type of music. How would you conduct a reasonably reliable nonprobability sample?

6. A manufacturer of precision gaskets makes gaskets in two grades: military and consumer automobile. In military applications, the precise gasket thickness is far more critical than in consumer automobile applications. The production run for military applications is very small, whereas the production run for consumer applications is very large. Explain how these facts affect decisions in sample design, confidence intervals, and sample size.

7. You wish to take an unrestricted random sample of undergraduate students at Cranial University to ascertain their levels of spending per month for food purchased off campus and eaten on the premises where purchased. You ask a test sample of nine students about their food expenditures and find that on the average they report spending $20, with two-thirds of them reporting spending from $10 to $30. What size sample do you think you should take? (Assume your universe is infinite.)

8. You wish to adjust your sample calculations to reflect the fact that there are only 2,500 students in your population. How does this additional information affect your estimated sample size in question 7?

9. Your large firm is facing its first union negotiation. Your superior wants an accurate evaluation of the morale of its large number of computer technicians. What size sample would you draw if it was to be an unrestricted sample?

Bringing Research to Life

10. Design an alternative nonprobability sample that will be more representative of infrequent and potential riders for the CityBus project.

11. How would you draw a cluster sample for the CityBus project?

From Concept to Practice

12. Using Exhibit 7–6 as your guide, for each sampling technique describe the sampling frame for a study of employers' skill needs in new hires using the industry in which you are currently working or wish to work.

WWW Exercises

Visit our website for Internet exercises related to this chapter at www.mhhe.com/business/cooper8

CASES*

A GEM OF A STUDY

GOODYEAR'S AQUATRED

CALLING UP ATTENDANCE

CAN THIS STUDY BE SAVED?

INQUIRING MINDS WANT TO KNOW—NOW!

KNSD SAN DIEGO

MATCH WITS WITH JASON

NCR: TEEING UP A NEW STRATEGIC DIRECTION

OUTBOARD MARINE CORPORATION

PEBBLE BEACH COMPANY

STATE FARM: DANGEROUS INTERSECTIONS

THE CATALYST FOR WOMEN IN FINANCIAL SERVICES

VOLKSWAGEN'S BEETLE

*All cases indicating a video icon are located on the Instructor's Videotape Supplement. All nonvideo cases are in the case section of the textbook. All cases indicating a CD icon offer a data set, which is located on the accompanying CD.

REFERENCE NOTES

1. United States Department of Commerce, Press Release CB99-CN.22, June 2, 1999. http://www.census.gov/Press-Release/www/1999/cb99.html.
2. W. E. Deming, *Sample Design in Business Research* (New York: Wiley, 1960), p. 26.
3. Henry Assael and John Keon, "Nonsampling versus Sampling Errors in Survey Research," *Journal of Marketing Research* (Spring 1982), pp. 114–23.
4. A. Parasuraman, *Marketing Research,* 2nd ed. (Reading, MA: Addison-Wesley, 1991), p. 477.
5. Fred N. Kerlinger, *Foundations of Behavioral Research,* 3rd ed. (New York: Holt, Rinehart & Winston, 1986), p. 72.
6. Amir D. Aczel, *Complete Business Statistics* (Burr Ridge, IL: Irwin, 1996), p. 180.
7. The correction for a finite population is shown in the example below:
   If a finite population of 20,000 is considered, the sample size is 256 for an interval of ± .5 meals and 95 percent confidence.

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{n-1}} = \sqrt{\frac{N-n}{N-1}} = 0.255 = \frac{4.1}{\sqrt{n-1}} \times \sqrt{\frac{20{,}000-n}{20{,}000-1}}$$

   or

$$n = \frac{s^2N + \sigma_{\bar{X}}^2 (N-1)}{s^2 + \sigma_{\bar{s}}^2 (N-1)}$$

$$n = 256$$

   where
   N = Size of the population
   n = Size of the sample
8. All estimates of costs are hypothetical.
9. Leslie Kish, *Survey Sampling* (New York: Wiley, 1965), p. 188.
10. Ibid., pp. 76–77.
11. Typically, stratification is carried out before the actual sampling, but when this is not possible, it is still possible to stratify after the fact. Ibid., p. 90.
12. W. G. Cochran, *Sampling Techniques,* 2nd ed. (New York: Wiley, 1963), p. 134.

13. Ibid., p. 96.
14. Kish, *Survey Sampling,* p. 94.
15. For detailed treatment of these and other cluster sampling methods and problems, see Kish, *Survey Sampling,* pp. 148–247.
16. J. H. Lorie and H. V. Roberts, *Basic Methods of Marketing Research* (New York: McGraw-Hill, 1951), p. 120.
17. Kish, *Survey Sampling,* p. 156.
18. For specifics on these problems and how to solve them, the reader is referred to the many good sampling texts. Two that have been mentioned already are Kish, *Survey Sampling,* chapters 5, 6, and 7; and Cochran, *Sampling Techniques,* chapters 9, 10, and 11.
19. A proportion is the mean of a dichotomous variable when members of a class receive the value of 1, and nonmembers receive a value of 0.

REFERENCES FOR SNAPSHOTS AND CAPTIONS

Terrorism

"Poll: 83 Percent Back Military Action," MSNBC, September 13, 2001 (http://www.msnbc.com/news/628177.asp).

Jeff Jones, "The Impact of the Attacks on America," Gallup Poll News Service, Gallup Organization, September 14, 2001 (http://www.gallup.com/poll/releases/pr010914c.asp).

Gary Langer, "Gauging Support: History Suggests Steady Support for Military Action," ABCNews.com, September 18, 2001 (http://more.abcnews.go.com/sections/politics/dailynews/wtc_poll010918.html).

Gary Langer, "Prayers for Victims, Support for Reprisals: Poll: Public Responds to Terror Attacks," ABCNews.com, September 11, 2001 (http://more.abcnews.go.com/sections/us/dailynews/wtc_abcpoll010911.html).

Gary Langer, "Standing United: American Public Closes Ranks Behind Bush and Anti-Terror Measures," ABCNews.com, September 15, 2001 (http://abcnews.go.com/scetions/DailyNews/wtc_poll010914.html).

Gary Langer, "Support for Bush: 2-1 Backing for War on Terrorism; Inaction Seen as the Greater Peril," ABCNews.com, September 21, 2001 (http://abcnews.go.com/sections/politics/DailyNews/wtc_poll010921.html).

Frank Newport, "Retaliation," Gallup News Service, Gallup Organization, September 14, 2001 (http://www.gallup.com/poll/releases/pr010914b.asp).

The Henry J. Kaiser Family Foundation

"About The Henry J. Kaiser Family Foundation" (http://www.kff.org/about/).

"Sex Education in America: A View from Inside the Nation's Classrooms," Reproductive and Sexual Health: The Henry J. Kaiser Family Foundation (http://www.kff.org/sections.cgi?section=repro&sub_section=re-sexed&disp=10).

"Summary of Findings: Sex Education in America: A Series of National Surveys of Students, Parents, Teachers, and Principals," The Henry J. Kaiser Family Foundation, September 2000 (http://www.kkf.org/content/2000/3048/sexED.pdf).

Diana Jean Schemo, "Survey Finds Parents Favor More Detailed Sex Education," *New York Times on the Web,* October 4, 2000 (http://www.nytimes.com/2000/10/04/national/04SEX.html).

U.S. Census

2000 Census Questionnaire, U.S. Census Bureau, United States Department of Commerce.

John Fetto, "Lust for Statistics," *American Demographics,* March 2001, pp. 68–70.

James Heckman, "Polls Debate, Researchers Wait, Time's Short," *Marketing News,* March 29, 1999.

Joan Raymond, "The Multicultural Report," *American Demographics,* November 2001, p. s6.

The Census, U.S. Census Bureau, United States Department of Commerce (http://www.census.gov/census/).

IRI

Christopher T. Heun, "Information Resources Redesigns Its Market Research Data," Informationweek.com, CPM United Business Media, August 21, 2001.

"IRI Launches Proprietary InfoScan Advantage to Track Wal-Mart Sales Information," press release, Information Resources, Inc. August 21, 2001 (http://www.infores.com/public/global/news/glo_new_082101.htm).

CLASSIC AND CONTEMPORARY READINGS

Deming, W. Edwards. *Sample Design in Business Research.* New York: Wiley, 1990. A classic by the late author, an authority on sampling.

Kalton, Graham. *Introduction to Survey Sampling.* Beverly Hills, CA: Sage Publications, 1983. An overview with particular attention to survey applications.

Kish, Leslie. *Survey Sampling.* New York: Wiley, 1995. A widely read reference on survey sampling, recently updated.

Namias, Jean. *Handbook of Selected Sample Surveys in the Federal Government.* New York: St. John's University Press, 1969. A unique collection of illustrative uses of sampling for surveys carried out by various federal agencies. Of interest both for the sampling designs presented and for the information on the methodology used to develop various government statistical data.

Yates, F. *Sampling Methods for Censuses and Surveys.* 4th ed. New York: Oxford University Press, 1987. A readable text with emphasis on sampling practices.