>chapter 14

Sampling

>learningobjectives

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 five questions that must be answered to develop a sampling plan.
4. The two categories of sampling techniques and the variety of sampling techniques within each category.
5. The various sampling techniques and when each is used.

> "We have to hear what's being said in a natural environment, and social media is an obvious place to do this, but we also have to go and discover the opinions that are not being openly shared. Only then can we understand the dichotomy between the public and private persona."
>
> Ben Leet, sales director
> uSamp

>**bringing**research**to**life

Researchers comprise a fairly small professional community. Within a given company, other than those that specialize in research, few trained researchers may be found, making collaboration necessary. Researchers from different companies often share their experiences at professional conferences in an attempt to advance the industry as a whole. As a result, researchers are often privy to each other's successes as well as their failures. They use each other's mistakes to improve their own projects. We join Jason and Sara as they are discussing sampling for a new project with Glacier Symphony.

"The ideal participant is thoughtful, articulate, rational, and, above all, cooperative. Real people, however, are fractious, stubborn, ill-informed, and even perverse. Nevertheless, they are who you have to work with," muses Jason as he and Sara hammer out the details of the Glacier Symphony sampling plan.

"Sam Champion, marketing director for CityBus," shares Sara, "certainly had sampling problems. He allowed a novice researcher—Eric Burbidge—to do the sampling to determine where the company could most effectively promote its new daily route schedule. Its big problem was a small budget and riders from two separate cities, where two different papers had substantial circulations—and just as substantial advertising rates. CityBus was hoping to advertise in only one paper. But the newspapers didn't have circulation figures for specific news vending boxes. Champion told the tale at the last MRA luncheon.

"It seems Burbidge was inexperienced enough to try to answer CityBus's question of which newspaper to use for ads by conducting a survey on one bus that runs between the two cities during evening rush hour. Burbidge boards the bus on route 99 and tells the driver he's from headquarters and there to do an official survey during the evening ride." Sara pauses for effect and lowers her voice to mimic a base frog. "I need to test my hypothesis that readership of newspapers on route 99 is equally divided between the *East City Gazette* and the *West City Tribune*."

Jason, now interested, interrupts, "He said that to a busload of passengers?"

"Well, no, the passengers hadn't yet boarded. The way Champion told the story, Burbidge barged his way to the front of the line and rapped his clipboard on the door to gain entry before any of the passengers could board. He said that to the driver.

"Anyway, Burbidge distributes his questionnaires, and the passengers diligently complete them and bring them forward to where Burbidge sits at the front of the bus. And then they start to play paper-ball hockey in the aisle of the bus."

"Paper-ball hockey?" questions Jason.

"Evidently they wad the newspapers they have been reading while waiting for the bus into balls and bat them through the legs of self-appointed goalies at each end of the bus aisle. Anyway, the driver tells Burbidge that since East City Club plays hockey that night that when he cleans out the bus most of the newspapers will be the *East City Gazette*. The riders evidently like to study the night's pro game in advance, so newsstand sales are brisk in the terminal, but only for the newspaper that does the better job of covering the sport *du jour*. Of course the next night, the riders would be buying the *West City Tribune* because it does a better job of covering pro basketball.

"Burbidge is upset and mumbles something about the survey asking for the paper most recently purchased. The driver tells him not to sweat it. 'They buy the *Gazette* before hockey and the *Trib* before basketball ... but of course in the morning they bring the paper that is dropped on their doorstep.'

"Burbidge is now mumbling that by choosing route 99, and choosing hockey night, he has totally distorted his results.

"The driver, who is Champion's favorite dart opponent, is thoroughly enjoying Burbidge's discomfort because he was acting like such an ass at the beginning. So the driver tells Burbidge, 'I know from reading the CityBus newsletter that by the time you announce the new routes and schedules, we will be finished with hockey and basketball and into the baseball season. And, of course, most of these folks on the 5:15 bus are East City folks, while most on the 5:45 bus are West City folks, so your outcome will naturally be affected by choosing the 5:15 any time of the year.'

"Burbidge, fully exasperated, asks the driver, 'Is there anything else you would care to share with me?'

"The driver evidently couldn't hide his grin when he says, 'The riders on the 5:45 usually don't read the newspaper much at all. They've been watching sports in the terminal bar while waiting for the bus. Most aren't feeling any pain—if you get my meaning—and can't read the small newsprint as I don't turn on the overhead lights.'"

Sara pauses, allowing Jason to ask, "Is there a lesson to this story, Sara?"

"Well, we've been talking about having the student musicians distribute and collect surveys at each Friday evening's performance. I'm wondering if Glacier Symphony has any demographic data from previous surveys that might shed some light on concert attendees. I'd hate to systematically bias our sample, like Burbidge did. Since we won't be present to collect the data—like he was—we might never know."

> 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 ads from a magazine, we usually assume our selection reflects the characteristics of the full set. If some members of our staff favor a promotional strategy, 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 individual participant or object on which the measurement is taken. It is the unit of study. Although an element may be a person, it can just as easily be something else. For example, each staff member questioned about an optimal promotional strategy is a population element, each advertising account analyzed is an element of an account population, and each ad is an element of a population of advertisements. 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. We call the listing of all population elements from which the sample will be drawn the **sample frame.**

For CityBus, the population of interest is all riders of affected routes in the forthcoming route restructuring. In studying customer satisfaction with the CompleteCare service operation for MindWriter, the population of interest 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, while mandated to take a census of the population every 10 years, shows its confidence in sampling by taking sample surveys to check the accuracy of its census. The U.S. Bureau of the Census knows that in a census, segments of the population are seriously undercounted. Only when the population is small, accessible, and highly variable is accuracy likely to be greater with a census than a sample.

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 also discourage prospective customers from defecting to competitive brands due to negative word of mouth.

Availability of Population Elements

Some situations require sampling. Safety is a compelling marketing appeal for most vehicles. Yet we must have evidence to make such a claim. So we crash-test cars to test bumper strength or efficiency of airbags to prevent injury. In testing for such evidence, we destroy the cars we test. A census would mean complete destruction of all cars manufactured. Sampling is also the only process possible if the population is infinite.

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-disc 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.

>snapshot

Ford Reenergizes by Changing Sampling Strategy

In the midst of the financial crisis in the automobile industry, Ford's James Farley decided his research was excluding a very important sample unit: the dealer. With dealers controlling 75 percent of advertising expenditures for the auto giant, Farley thought excluding them as research subjects was suicidal. So he recruited 30 of the most influential dealers to fly to Detroit to provide information and critique the creative proposals of the Ford ad agency, Team Detroit.

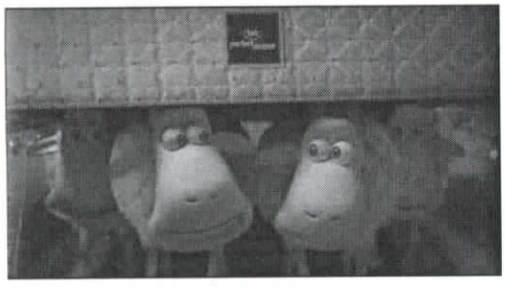
Farmington Hills (MI) full-service research firm Morpace put the dealers through an intensive focus group experience. The dealers were soon challenged with questions. "Which incentives work and which don't?" "What does the Ford brand mean to you?" "What is wrong with Ford's advertising?" In subsequent sessions, the dealers were asked to critique ad slogans and branding strategies, recommending those that best capture the Ford experience. The dealers left the 72-hour marathon session enthusiastic about the direction Ford was taking and with significant buy-in for the next ad campaign. Farley's actions gave voice to its dealers with its altered research sampling strategy.

www.ford.com; www.morpace.com; www.teamdetroit.com

Accuracy

Accuracy is the degree to which bias is absent from the sample. When the sample is drawn properly, the measure of behavior, attitudes, or knowledge (the measurement variables) of *some* sample elements will be *less than* (thus, underestimate) the measure of those same variables drawn from the population. Also, the measure of the behavior, attitudes, or knowledge of *other* sample elements will be *more than* the population values (thus, overestimate them). Variations in these sample values offset each other, resulting in a sample value that is 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 that favors neither overestimation nor underestimation.

For example, assume you were asked to test the level of brand recall of the "counting sheep" creative approach for the Serta mattress company. Hypothetically, you could measure via sample or census. You want a measure of brand recall in combination with message clarity: "Serta mattresses are *so comfortable* you'll feel the difference the minute you lie down." In the census, 52 percent of participants who are TV viewers correctly recalled the brand and message. Using a sample, 70 percent recalled the brand and correctly interpreted the message. With both results for comparison, you would know that your sample was biased, as it significantly overestimated the population value of 52 percent. Unfortunately, in most studies taking a census is not feasible, so we need an estimate of the amount of error.[5]



Serta Counting Sheep

An accurate (unbiased) sample is one in which the underestimators offset the overestimators. **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."[6]

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 only corner homes will cause us to overestimate home values in the area. Burbidge learned that in selecting bus 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.

Increasing the sample size can reduce systematic variance as a cause of error. However, even the large size won't reduce error if the list from which you draw your participants is biased. The classic example of a sample with systematic variance was the *Literary Digest* presidential election poll in 1936, in which more than 2 million people 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, whose members could not afford to own phones and typically voted for the Democratic Party candidate.

Precision

A second criterion of a good sample design is precision of estimate. Researchers accept that no sample will fully represent its population in all respects. However, to interpret the findings of research, we need a measure of how closely the sample represents the population. 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** (or *random sampling error*) and reflects the influence 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 Jason draws a sample from an alphabetical list of MindWriter owners who are having their laptops currently serviced by the CompleteCare program. Assume 80 percent of those surveyed had their laptops serviced by Max Jensen. Also assume from the exploratory study that Jensen had more complaint letters about his work than any other technician. Arranging the list of laptop owners currently being serviced in an alphabetical listing would have failed to *randomize* the sample frame. If Jason drew the sample from that listing, he would actually have increased the 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.

Types of Sample Design

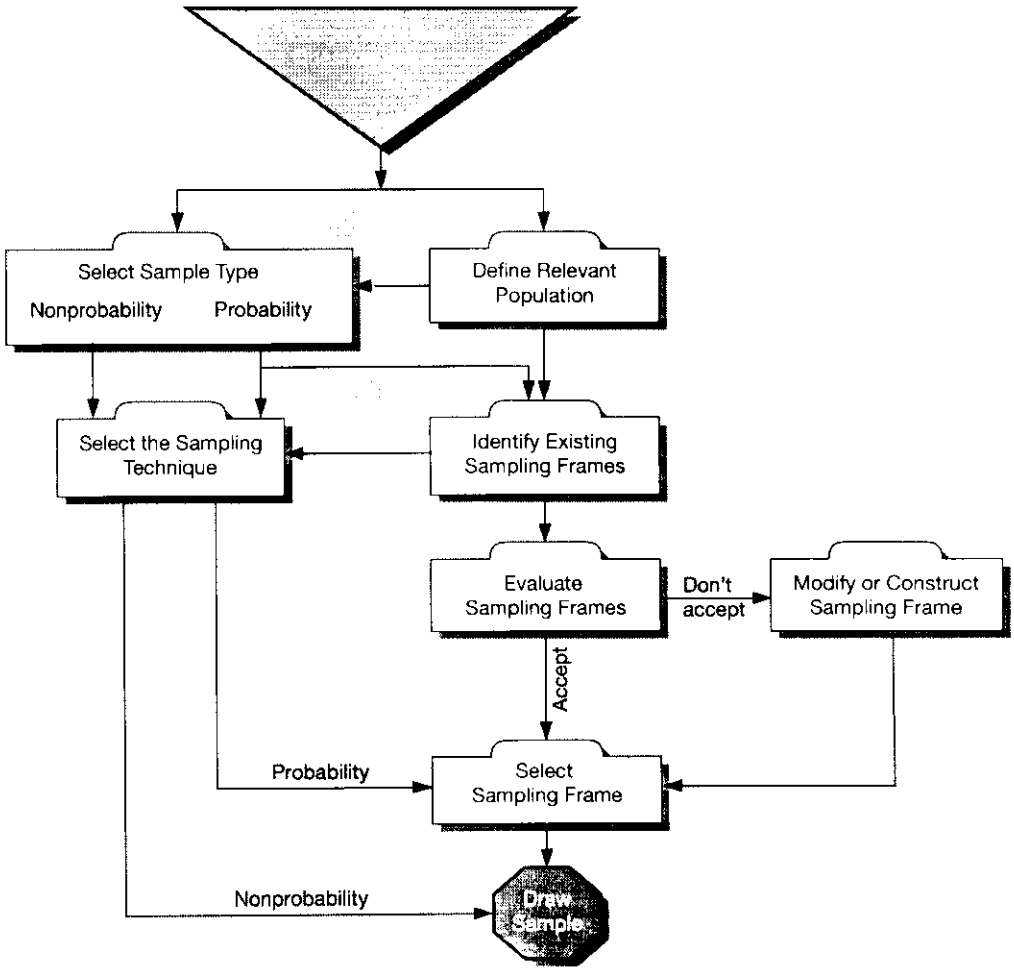
The researcher makes several decisions when designing a sample. These are represented in Exhibit 14-1. The sampling decisions flow from two decisions made in the formation of the management-research question hierarchy: the nature of the management question and the specific investigative questions that evolve from the research question. These decisions are influenced by requirements of the project and its objectives, level of risk the researcher can tolerate, budget, time, available resources, and culture.

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 CompleteCare customer 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

Exhibit 14-1 Sampling Design within the Research Process



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 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 14-2, classify the different approaches.

Exhibit 14-2 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	

The members of a sample are selected using probability or nonprobability procedures.

Nonprobability sampling is arbitrary and subjective; when we choose subjectively, we usually do so with a pattern or scheme in mind (e.g., only talking with young people or only talking with women). Each member of the population does not have a known chance of being included. Allowing interviewers during a mall-intercept study to choose sample elements "at random" (meaning "as they wish" or "wherever they find them") is not random sampling. Although we are not told how Burbidge selected the riders of bus route 99 as his sample, it's clear that he did not use probability sampling techniques.

Early Internet samples had all the drawbacks of nonprobability samples. Those individuals who frequented the Internet were not representative of most target markets or audiences, because far more young, technically savvy men frequented the Internet than did any other demographic group. As Internet use increases and gender discrepancies diminish, many such samples now closely approximate non-Internet samples. Of increasing concern, however, is what the Bureau of the Census labels the "great digital divide"—low-income and ethnic subgroups' underrepresentation in their use of the Internet compared to the general population. Additionally, many Internet samples were, and still are, drawn substantially from panels. These are composed of individuals who have self-selected to become part of a pool of individuals interested in participating in online research. There is much discussion among professional researchers about whether Internet samples should be treated as probability or nonprobability samples. Some admit that any sample drawn from a panel is more appropriately treated as a nonprobability sample; others vehemently disagree, citing the success of such well-known panels as NielsenMedia's People Meter panels for TV audience assessment and IRI's BehaviorScan panel for tracking consumer packaged goods. As you study the differences here, you should draw your own conclusion.

Key to the difference between nonprobability and probability samples is the term *random*. In the dictionary, random is defined as "without pattern" or as "haphazard." In sampling, random means something else entirely. **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. This procedure is never haphazard. Only probability samples provide estimates of precision. When a researcher is making a decision that will influence the expenditure of thousands, if not millions, of dollars, an estimate of precision is critical. Also, only probability samples offer the opportunity to generalize the findings to the population of interest from the sample population. Although exploratory research does not necessarily demand this, explanatory, descriptive, and causal studies do.

Whether the elements are selected individually and directly from the population—viewed as a single pool—or 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.

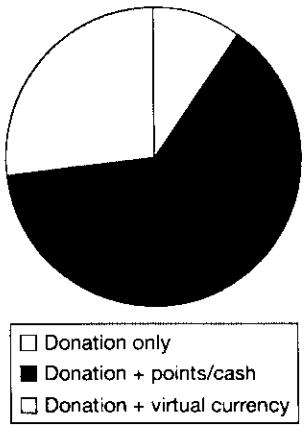
snapshot

If you could feed a hungry child or prevent the euthanasia of a dog by taking a survey, would you?

Researchers have been using online panels to draw samples for Web, email, and mobile surveys for the last decade. Most of these panels are developed using three sources: advertising networks (sample providers advertise via the Web, email, and other media to attract individuals who are willing to participate in surveys), loyalty programs (a sponsor company uses its own list of individuals who are part of its loyalty efforts and recruit them to take surveys) and social media (sample providers use Facebook, Twitter, and numerous other social media to recruit participants). Sampling firms under the current model employing these sources often incentivize their participants with money, Internet currency or points, or prizes. Unfortunately, one of the drawbacks is a small pool of individuals—even though in the millions—from which thousands of firms are drawing their participants.

The founding partners of Research for Good (RFG) were concerned that people receiving a personal incentive to give their opinion tend to only represent a particular segment of the population. Also, there was the ongoing industry concern about the development of "professional respondents"—when volunteer participants are recruited by hundreds of sampling companies in roughly the same three ways to meet an ever-growing demand for survey participants. RFG was also concerned about social responsibility. So, it developed an incentivized sampling model that uses charitable donations to attract uniquely different respondents. Its *SaySo for Good* panel draws on the 90 percent of U.S. and Canadian adults who support at least one charity. When joining the panel, each participant chooses a charity to receive their survey incentive. The RFG database of charities includes every government-registered charity in both the U.S. and Canada. RFG delivers to the specified charity $1.00 or 25 percent of the budgeted cost of a per completed survey for each participant (i.e., if a completed survey is budgeted at $8.00 per participant, then the charity receives $2.00).

Research for Good Sample Composition



RFG has discovered that participants attracted by charity incentives are different—in both behavior and attitude—than those attracted to a research panel by other means. These survey takers tend to be cause-minded, are financial supporters or volunteers to charities, and are not likely to be motivated by typical cash or prize incentives. Their response rates are higher, as well as their completion rates. And they are infrequent survey takers, reducing the concern about "professional respondents." RFG now counts thousands of charity-incentivized members in its panel. "By attracting new participants, we will generate better data quality while we serve the greater good," emphasized cofounder Sean Case. As a result, RFG now composes its samples by including this charity-incentive model.

www.researchforgood.com; www.saysoforgood.com

>

There are several questions to be answered 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.

1. What is the target population?
2. What are the parameters of interest?
3. What is the sampling frame?
4. What is the appropriate sampling method?
5. What size sample is needed?

What Is the Target 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 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 those same 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 level of data. Each different level of data also generates different sample statistics. Thus, choosing the parameters of interest will actually dictate the sample type and its size.

Exhibit 14-3 Example Population Parameters

Study	Population Parameter of Interest	Data Level & Measurement Scale
CityBus	Frequency of ridership within 7 days	Ordinal (more than 10 times, 6 to 10 times, 5 or fewer times)
		Ratio (absolute number of rides)
MindWriter	Perceived quality of service	Interval (scale of 1 to 5, with 5 being "exceeded expectations")
	Proportion by gender of Laptop 9000 customers with problems	Nominal (percent female, male)
Metro U	Frequency of eating on or near campus within the last 30 days	Ratio (actual eating experiences)
	Proportion of students/employees expressing interest in dining club	Nominal (interested, not interested)

Percent of U.S. Households Accessible by Phone Method of Sampling Invitation

Land Line + Cell Phone	Cell Phone Only
55	32
Landline Only	**Neither Cell nor Landline**
11	2

picprofile

Mixed-access sampling means that multiple methods are used to invite participants to a research study—phone, email, mobile/wireless, address-based/mail, etc. Approximately 98 percent of possible participants are reachable by phone, whereas only 80 percent are reachable online. Mixed-access sampling reduces noncoverage error and nonresponse error. Once a participant is recruited, regardless of the means, he or she may complete the study by a different mode (e.g., recruited by phone but take a survey online). Sample recruitment is increasingly done by mixed access.

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 ordinal data estimator. Of course, we could ask the question differently and obtain an absolute count of eating experiences and that would generate ratio data. In MindWriter, the rating of service by CompleteCare on a 5-point scale 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. Exhibit 14-3 indicates population parameters of interest for our three example studies.

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. 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."[7] 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 (e.g., 30 percent; therefore p is equal to 0.3 and q, those not interested, equals 0.7) 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 route 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.

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 (we would want the list organized by these subgroups, or within the list each characteristic of each element would need to be noted), 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; to reach frequent riders CityBus could use on-bus signs or distribute paper schedules rather than using more expensive newspaper ads. 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.

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.

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 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. Although you may use the general city directory, sampling from this too inclusive frame would be costly and inefficient, because 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, because no sample frame of CityBus riders existed. Although some regular riders used monthly passes, infrequent riders usually paid cash for their fares. It might have been possible for Burbidge to anticipate this and to develop 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.

The sampling issues we have discussed so far are fairly universal. It is not until we begin talking about sampling frames and sampling methods that international research starts to deviate. International researchers often face far more difficulty in locating or building sample frames. Countries differ in how each defines its population; this affects census and relevant population counts.[8] Some countries purposefully oversample to facilitate the analysis of issues of particular national interest; this means we need to be cautious in interpreting published aggregate national figures.[9] These distinctions and difficulties may lead the researcher to choose nonprobability techniques or different probability techniques than they would choose if doing such research in the United States or other developed countries. In a study that is fielded in numerous countries at the same time, researchers may use different sampling methodologies, resulting in hybrid studies that will need care to be combined. It is common practice to weight sample data in cross-national studies to develop sample data that are representative.[10] Choice of sampling methods is often dictated by culture as much as by communication and technology infrastructure. Just as all advertising campaigns would not be appropriate in all parts of the world, all sampling techniques would not be appropriate in all subcultures. Our discussion in this text focuses more on domestic than international research. We believe it is easier to learn the principles of research in an environment that you know versus one in which many students can only speculate. Yet we also believe that ethnic and cultural sensitivity should influence every decision of researchers, whether they do research domestically or internationally.

What Is the Appropriate Sampling Method?

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 the 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. Thus, no matter how careful we are in replacing those who refuse or are never located, sampling error is likely to rise.

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 highly 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.

Although the probability cluster sampling technique was available to him, it is obvious that Burbidge chose nonprobability sampling, arbitrarily choosing bus route 99 as a judgment sample and attempting to survey everyone riding the bus during the arbitrary times in which he chose to ride. What drove him to this decision is likely what makes researchers turn to nonprobability sampling in other situations: ease, speed, and cost.

How Large Must a Sample Be?

Much folklore surrounds this question. The most pervasive myths are (1) a sample must be large or it is not representative and (2) a sample should bear some proportional relationship to the size of the population from which it is drawn. With nonprobability samples, researchers confirm these myths

using the number of subgroups, rules of thumb, and budget considerations to settle on a sample size. In probability sampling, 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. Some principles that influence sample size include:

- 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 or smaller the error 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.

Cost considerations influence decisions about the size and type of sample and the data collection methods. Almost all studies have some budgetary constraint, and this may encourage a researcher to use a nonprobability sample. Probability sample surveys incur list costs for sample frames, callback costs, and a variety of other costs that are not necessary when nonprobability samples 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.[11]

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; Jason needs to be highly confident that his recommendations to change CompleteCare procedures and policies are on target and thoroughly supported by the data collected. These considerations justify MindWriter's probability sampling approach.

> [illegible]

The unrestricted, simple random sample is the purest form of probability sampling. Since all probability samples must provide a known nonzero probability 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.

$$\text{Probability of selection} = \frac{\text{Sample size}}{\text{Population size}}$$

The Metro U dining club study has a population of 20,000. If the sample size is 300, the probability of selection is 1.5 percent (300/20,000 = 0.015). In this section, we use the simple random sample to build a foundation for understanding sampling procedures and choosing probability samples. The simple random sample is easy to implement with automatic dialing (random dialing) and with computerized voice response systems. However, it requires a list of population elements, can be time-consuming and expensive, and can require larger sample sizes than other probability methods. Exhibit 14-4 provides an overview of the steps involved in choosing a random sample.

Exhibit 14-4 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 D, Exhibit D-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. Exhibit C-10 in Appendix C is arranged into 10 columns of five-digit strings, but this is solely for readability.

Assume the researchers want a sample of 10 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 Exhibit C-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.

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 kth element in the population is sampled, beginning with a random start of an element in the range of 1 to k. The kth element, or **skip interval,** is determined by dividing the sample size into the population size to obtain the skip pattern applied to the sampling frame. This assumes that the sample frame is an accurate list of the population; if not, the number of elements in the sample frame is substituted for population size.

$$k = \text{Skip interval} = \frac{\text{Population size}}{\text{Sample size}}$$

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 kth 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, do the following:

- Identify, list, and number the elements in the population.
- Identify the skip interval (k).
- 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.

Systematic sampling can introduce subtle biases. A concern with systematic sampling is the possible *periodicity* in the population that parallels the sampling ratio. In sampling restaurant sales of dessert by drawing days of the year, a skip interval of 7 would bias results, no matter which day provides the random start. A less obvious case might involve a survey in an area of apartment buildings where the typical pattern is eight apartments per building. A skip interval of 8 could easily oversample some types of apartments and undersample others.

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 skip interval of 40 beginning with a random start at 16 would exclude the 15 largest banks and give a small-size bias to the findings.

The only protection against these subtle biases is constant vigilance by the researcher. Some ways to avoid such bias include:

Randomize the population before sampling (e.g., order the banks by name rather than size).

Change the random start several times in the sampling process.

Replicate a selection of different samples.

Although 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.[12]

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 results from the study can then be weighted (based on the proportion of the strata to the population) and combined into appropriate population estimates.

There are three reasons 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 or strata, and (3) to enable different research methods and procedures to be used in different strata.[13]

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.

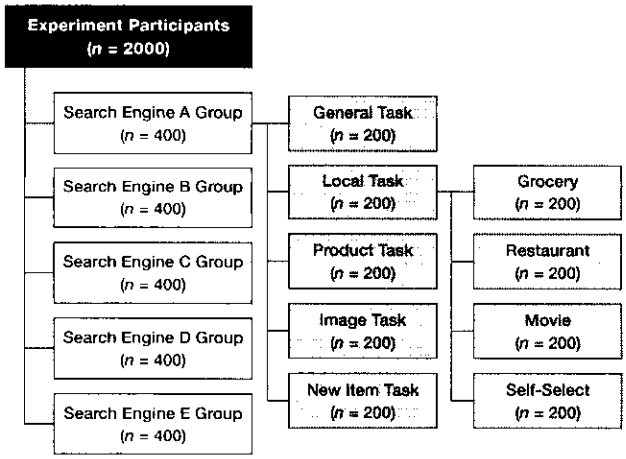
closeup

Twice yearly Keynote Systems evaluates the performance of five search engines, including market leader Google, AOL Search, Yahoo! Search, Ask.com, and MSN Search. Keynote, a "worldwide leader in services that improve online business performance and communications technologies," uses an online panel to perform "interactive Web site tests to assess user experience," profiling not only how people use search engines, but why they search as they do. Keynote allocates participants and experimental treatments as in Exhibit 14-5: 2,000 people are randomly drawn from more than 160,000 panel members and invited to participate via e-mail. They are assigned randomly to five groups of 400; each group is assigned a particular search engine. Whether participants have any experience with that particular engine is not a criterion for assignment. Each group is assigned a series of search tasks, starting with a general task—Think about anything you would like to search for; go and search that—to more specific tasks—find a local establishment, a product, an image, and a news item. Each search engine-allocated group essentially performs the same series of tasks. From their activities, Keynote generates 250,000 metrics (including time involved in the search, whether the search was successful, etc.). It matches these metrics to survey data used to measure satisfaction, perceived difficulty, and specific frustrations. From this combined data it develops several indices.

"One of the things we noted from a series of such tests was that Google repeatedly received rave reviews, even in instances where performance measures told a different story," shares senior research consultant Lance Jones. With almost 60 percent market share, Google has strong recognition and tends to set the bar in search site design. Is its brand that powerful that it can influence attitudes even in the face of conflicting performance experience? If the brand is not a factor, which search engine would produce the most satisfying and useful results, the best sponsored results, and the best presentation and design? Keynote wanted to design an experiment that would show the power of the search engine brand. To do that, they needed to remove brand identity from the search results. Its solution was to design a generic-appearing search engine website and results format page, feeding actual search results into its generic format.

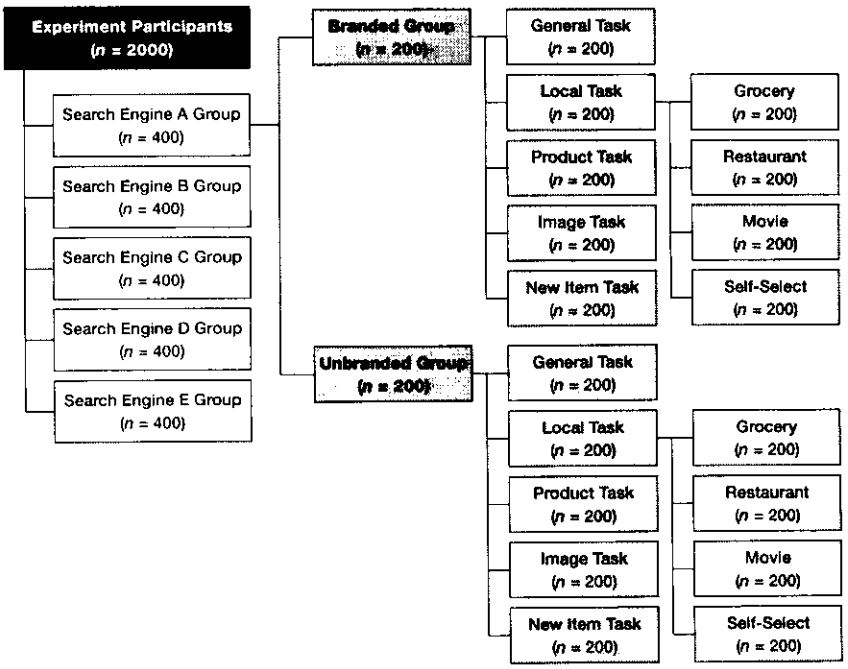
For the brand power test (Exhibit 14-6) 2,000 participants were again divided into five groups and assigned one search engine. This time, however, half the participants were assigned to a branded group (n = 200) and would see the results with a text line "Results brought to you by Yahoo/Google/Ask, etc."; the other half (n = 200) would see the same results but without the brand notation line (n = 200). All five search engines were tested using the tasks performed in the standard twice-annual test, but all the results seen by participants were actually generated using the assigned search engine, then fed into the generic results presentation. "The results pages were delivered live and participants would have perceived no difference in elapsed time, as the results were delivered within milliseconds of what the standard search would have delivered," explained Jones. The test produced 1,600 queries that generated 12 distinct metrics.

Exhibit 14-5 Participant Allocation in Search Engine Test



closeupcont'd

Exhibit 14-6 Participant Allocation in Brand Power Test



Is a brand powerful? Here are some sample results for Google; keep in mind that the branded group and the unbranded group saw the exact same results pages. On the unbranded group, the calculated Google results satisfaction score was 732 (on a 1,000-point scale), while the branded group delivered an 800; Google's sponsored results satisfaction was 763 (unbranded) compared to 809 (branded); full design satisfaction was 753 (unbranded) compared to 806 (branded). Evaluate the design of this sample.

www.keynote.com

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. Similarly, if a restaurant were interested in testing menu changes to attract younger patrons while retaining its older, loyal customers, stratified sampling using age and prior patronage as descriptors would be appropriate. Stratification is also called for when different methods of data collection are applied in different parts of the population, a research design that is becoming increasingly common. This might occur when we survey company employees at the home office with one method but must use a different approach with employees scattered throughout the country.

If data are available on which to base a stratification decision, how shall we go about it?[14] The ideal stratification would be based on the primary variable under study. If the major concern were to learn how often per month patrons would use the Metro U 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 days at work or class schedule as an indication of when a sample element might be near campus at mealtimes.

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.

The more strata used, the closer you come to maximizing interstrata differences (differences between strata) 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. The more strata you have, the higher the cost of the research project due to the cost associated with more detailed sampling. There is little to be gained in estimating population values when the number of strata exceeds six.[15]

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 that the sample size drawn from the stratum 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 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 stratified sampling.** 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 error range estimates for individual strata. The following table shows the relationship between proportionate and disproportionate stratified sampling.

Stratum	Population	Proportionate Sample	Disproportionate Sample
Male	45%	45%	35%
Female	55	55	65

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.[16]

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."[17]

The process for drawing a stratified sample is:

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 from each stratum.

Exhibit 14-7 Comparison of Stratified and Cluster Sampling

Stratified Sampling	Cluster Sampling
1. We divide the population into a few subgroups. • Each subgroup has *many* elements in it. • Subgroups are selected according to some criterion that is related to the variables under study.	1. We divide the population into *many* subgroups. • Each subgroup has *few* elements in it. • Subgroups are selected according to some criterion of ease or availability in data collection.
2. We try to secure *homogeneity* within subgroups.	2. We try to secure *heterogeneity* within subgroups.
3. We try to secure *heterogeneity* between subgroups.	3. We try to secure *homogeneity* between subgroups.
4. We randomly choose *elements* from within each subgroup.	4. We randomly choose several *subgroups* that we then typically study in depth.

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, as indicated in Exhibit 14-7.

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 often don't meet the need for heterogeneity and, instead, are homogeneous. For example, families in the same block (a typical cluster) are often similar in social class, income level, ethnic origin, and so forth. Although 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 the problems of both 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 resulting clusters?
2. Shall we seek equal-size or unequal-size 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. When 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, researchers might have chosen a course as a cluster, choosing to sample all students in that course if it enrolled students of all four class years. Or maybe they could choose a departmental office that had faculty, staff, and administrative positions as well as student workers. In area sampling to increase intracluster variance, researchers could combine into a single cluster adjoining blocks that contain different income groups or social classes.

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 naturally equal, such as households in city blocks. 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:

- 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.[18]

3. There is no *a priori* answer to the ideal cluster size question. Comparing the efficiency of differing cluster sizes requires that we discover the different costs for each size and estimate the different variances of the cluster means. Even with single-stage clusters (where the researchers interview or observe every element within a cluster), it is not clear which size (say, 5, 20, or 50) is superior. Some have found that in studies using single-stage area clusters, the optimal cluster size is no larger than the typical city block.[19]

4. Concerning single-stage or multistage cluster designs, for most large-scale area sampling, the tendency is to use multistage designs. Several situations justify drawing a sample within a cluster, in preference to the direct creation of smaller clusters and taking a census of that cluster using one-stage cluster sampling:[20]

- Natural clusters may exist as convenient sampling units yet, for economic reasons, may be larger than the desired size.
- We can avoid the cost of creating smaller clusters in the entire population and confine subsampling to only those large natural clusters.

The sampling of naturally compact clusters may present practical difficulties. For example, independent interviewing of all members of a household may be impractical.

5. The answer to how many subjects must be interviewed or observed depends heavily on the specific cluster design, and the details can be complicated. Unequal clusters and multistage samples are the chief

complications, and their statistical treatment is beyond the scope of this book.[21] Here we will treat only single-stage sampling with equal-size clusters (called *simple cluster sampling*). It is analogous to simple random 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. Assuming the same specifications for precision and confidence, we should expect that the calculation of a probability sample size would be the same for both clusters.

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 14-8 may be used to compare the various advantages and disadvantages

Exhibit 14-8 Comparison of Probability Sampling Designs

Type	Description	Advantages	Disadvantages
Simple Random *Cost: High* Use: Moderate	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.
Systematic *Cost: Moderate* Use: Moderate	Selects an element of the population at the beginning with a random start, and following the sampling skip interval selects every *k*th element.	Simple to design. Easier to use than the simple random. Easy to determine sampling distribution of mean or proportion.	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 *Cost: High* Use: Moderate	Divides population into subpopulations or strata and uses simple random on each stratum. 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. Especially expensive if strata on the population have to be created.
Cluster *Cost: Moderate* Use: High	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) *Cost: Moderate* Use: Moderate	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.

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 designs, we can estimate an error range within which the population parameter is expected to fall. Thus, we can reduce not only 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 the less precise methods.

Practical Considerations

We may use nonprobability sampling procedures because they satisfactorily meet the sampling objectives. Although 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 in which 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 route 99 would generate questionable data, but he seemed to realize the fallacy of many of his assumptions once he spoke with bus route 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 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. In Internet surveys those who volunteer don't always represent the appropriate cross section—that's why screening questions are used before admitting a participant to the sample. There is, however, some of this self-selection in almost all surveys because every respondent chooses whether to be interviewed.

Method 1

[illegible]

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, a TV reporter's "person-on-the-street" intercept interviews, or the use of employees to evaluate the taste of a new snack food.

Although 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.

[illegible]

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 route 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 route 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 following 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 might 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 business researchers. Probability sampling is usually much more costly and time-consuming. Advocates of quota sampling argue that although 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.

This design has found a niche in recent years in applications where respondents are difficult to identify and are best located through referral networks. It is also especially appropriate for some qualitative studies. 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 refer the researcher to 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.

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 although 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 In developing a sample, five procedural questions need to be answered:

- **a** What is the target population?
- **b** What are the parameters of interest?
- **c** What is the sampling frame?
- **d** What is the appropriate sampling method?
- **e** What size sample is needed?

4 A variety of sampling techniques are available. They may be classified by their representation basis and element selection techniques.

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

Probability sampling is based on random selection—a controlled procedure that ensures that each population element is given a known nonzero chance of selection. 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 contrast, nonprobability selection is "not random." When each sample element is drawn individually from the population at large, this is unrestricted sampling. Restricted sampling covers those forms of sampling in which the selection process follows more complex rules.

5 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 kth 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 may 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.

keyterms

area sampling 356
census 338
cluster sampling 355
convenience samples 359
disproportionate stratified sampling 354
double sampling 357
judgment sampling 359
multiphase sampling 357
nonprobability sampling 343
population 338
population element 338
population parameters 345
population proportion of incidence 346
probability sampling 343
proportionate stratified sampling 354
quota sampling 359
sample frame 338
sample statistics 345
sampling 338
sampling error 341
sequential sampling 357
simple random sample 349
skip interval 350
snowball sampling 360
stratified random sampling 351
systematic sampling 350
systematic variance 340

discussionquestions

1 Distinguish between:
   a Statistic and parameter.
   b Sample frame and population.
   c Restricted and unrestricted sampling.
   d Simple random and complex random sampling.
   e Convenience and purposive sampling.
   f Sample precision and sample accuracy.
   g Systematic and error variance.
   h Variable and attribute parameters.
   i 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 A disproportionate stratified probability sample?
3 You plan to conduct a survey using unrestricted sampling. What subjective decisions must you make?
4 Describe the differences between a probability sample and a nonprobability sample.
5 Why would a researcher use a quota purposive sample?
6 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 by the Glacier Symphony 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?
7 Your large firm is about to change to a customer-centered organization structure, in which employees who have rarely had customer contact will now likely significantly influence customer satisfaction and retention. As part of the transition, your superior wants an accurate evaluation of the morale of the firm's large number of computer technicians. What type of sample would you draw if it was to be an unrestricted sample?

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

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

10 Using Exhibit 14-8 as your guide, for each sampling technique describe the sample frame for a study of employers' skill needs in new hires using the industry in which you are currently working or wish to work.

11 When Nike introduced its glow-in-the-dark Foamposite One Galaxy sneakers, fanatics lined up at distributors around the country. As crowds became restless, jockeying for position at the front of increasingly long lines for the limited-supply shoes, Footlocker cancelled some events. It's been suggested that Nike should sell its limited-release introductions online rather than in stores to avoid putting its customer's safety in jeopardy. What sample group would you suggest Nike use to assess this suggestion?

cases*

Akron Children's Hospital

Calling Up Attendance

Campbell-Ewald Pumps Awareness into the American Heart Association

Campbell-Ewald: R-E-S-P-E-C-T Spells Loyalty

Can Research Rescue the Red Cross?

Goodyear's Aquatred

Inquiring Minds Want to Know—NOW!

Marcus Thomas LLC Tests Hypothesis for Troy-Bilt Creative Development

Ohio Lottery: Innovative Research Design Drives Winning

Pebble Beach Co.

Starbucks, Bank One, and Visa Launch Starbucks Card Duetto Visa

State Farm: Dangerous Intersections

The Catalyst for Women in Financial Services

USTA: Come Out Swinging

Volkswagen's Beetle

* You will find a description of each case in the Case Index section of the textbook. Check the Case Index to determine whether a case provides data, the research instrument, video, or other supplementary material. Written cases are downloadable from the text website (www.mhhe.com/cooper12e). All video material and video cases are available from the Online Learning Center. The film reel icon indicates a video case or video material relevant to the case.

appendix 14a

Probability Sampling

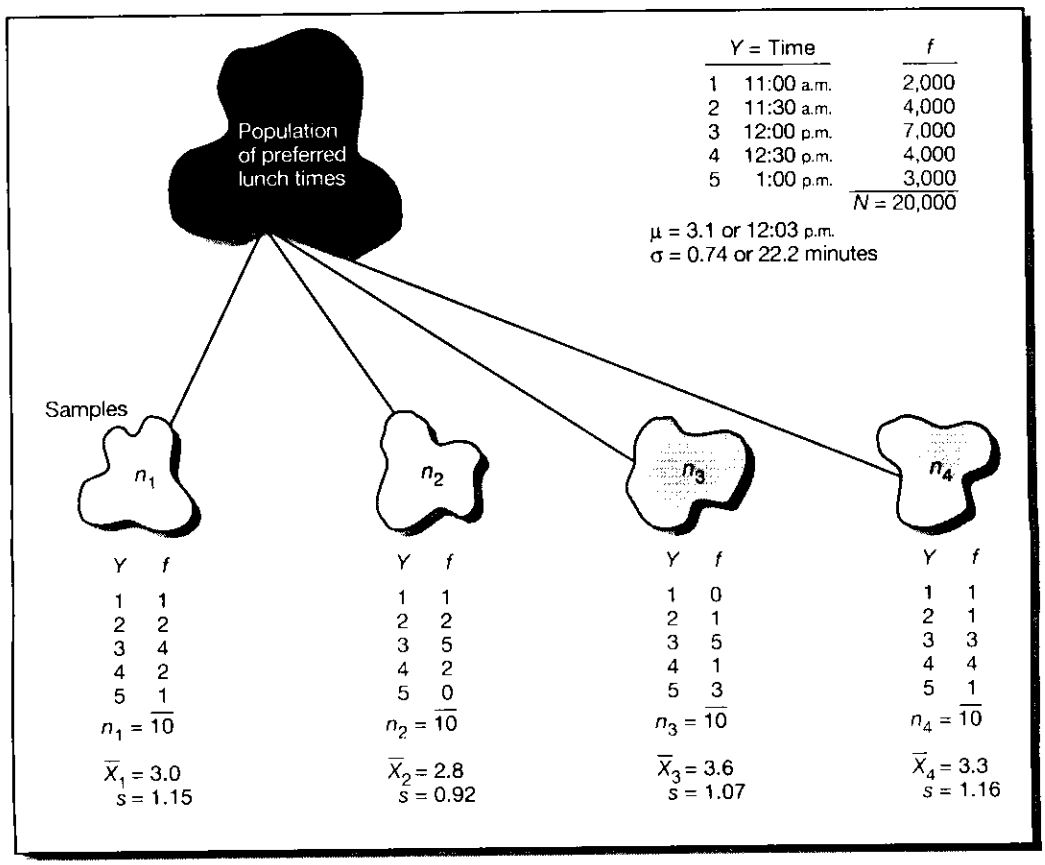
In the Metro University Dining Club study, we explore probability sampling and the various concepts used to design the sampling process.

Exhibit 14a-1 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 a.m., 11:30 a.m., 12 noon, 12:30 p.m., and 1 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.

Sample Data

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 14a-1. We also calculate a mean score, $\bar{X}_1$ = 3.0, for this sample. This mean would place the average preferred lunch time at 12 noon. The mean is a *point estimate* and our best predictor

Exhibit 14a-1 Random Samples of Preferred Lunch Times



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 (see Exhibit 14a-2): 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.

Let's now examine what happens when we apply sample data (n_1) from Exhibit 14a-1 to the formula. The sample standard deviation from sample n 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 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 noon (the mean of n_1) ±0.36. This range may be visualized on a continuum (see diagram at bottom of page).

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 0.64 = (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 14a-3). 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 0.36 by ± 1.96 (Z), since 1.96 Z covers 95 percent of the area under the curve (see Exhibit 14a-4). Now, with 95 percent confidence, the interval in which we would find the true mean increases to ± 0.70 (from 2.3 to 3.7 or from 11:39 a.m. to 12:21 p.m.).

Parenthetically, if we compute the standard deviation of the distribution of sample means in Exhibit 14a-1, [3.0, 2.8, 3.6, 3.3], we will discover it to be 0.35. Compare this to the standard error from the original calculation (0.36). The result is consistent 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 14a-5. 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 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). In part C of Exhibit 14a-5, three observations about this sample distribution are consistent with our earlier illustration. First,

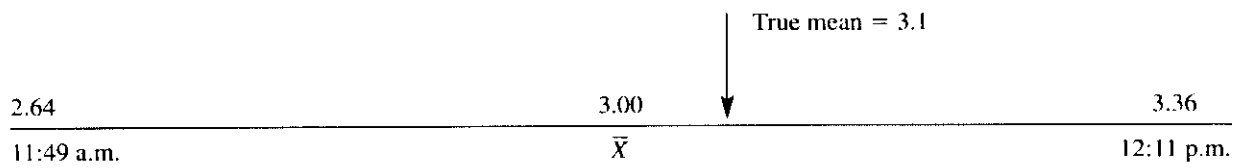


Exhibit 14a-2 Effects on Standard Error of Mean of Increasing Precision

	Reducing the Standard Deviation by 50%	Quadrupling the Sample
$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$	$\sigma_{\bar{x}} = \frac{0.74}{\sqrt{10}} = 0.234$	$\sigma_{\bar{x}} = \frac{0.8}{\sqrt{25}} = 0.16$
	$\sigma_{\bar{x}} = \frac{0.37}{\sqrt{10}} = 0.117$	$\sigma_{\bar{x}} = \frac{0.8}{\sqrt{100}} = 0.08$

where
$\sigma_{\bar{x}}$ = standard error of the mean
$\sigma_{\bar{x}}$ = standard deviation of the sample
n = sample size

Note: A 400 percent increase in sample size (from 25 to 100) would yield only a 200 percent increase in precision (from 0.16 to 0.08). Researchers are often asked to increase precision, but the question should be, at what cost? Each of those additional sample elements adds both time and cost to the study.

Exhibit 14a-3 Confidence Levels and the Normal Curve

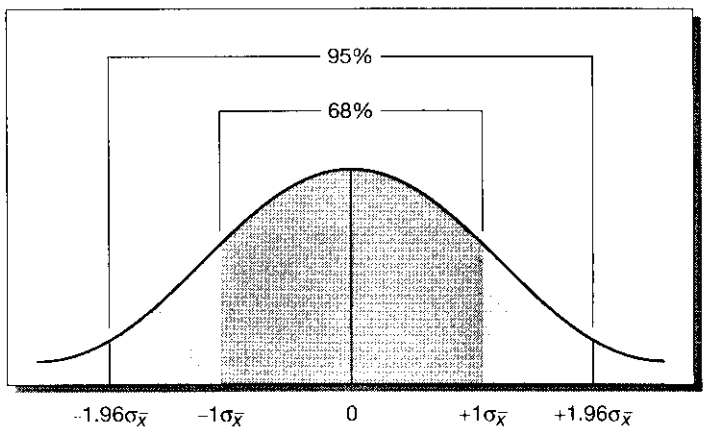


Exhibit 14a-4 Standard Errors Associated with Areas under the Normal Curve

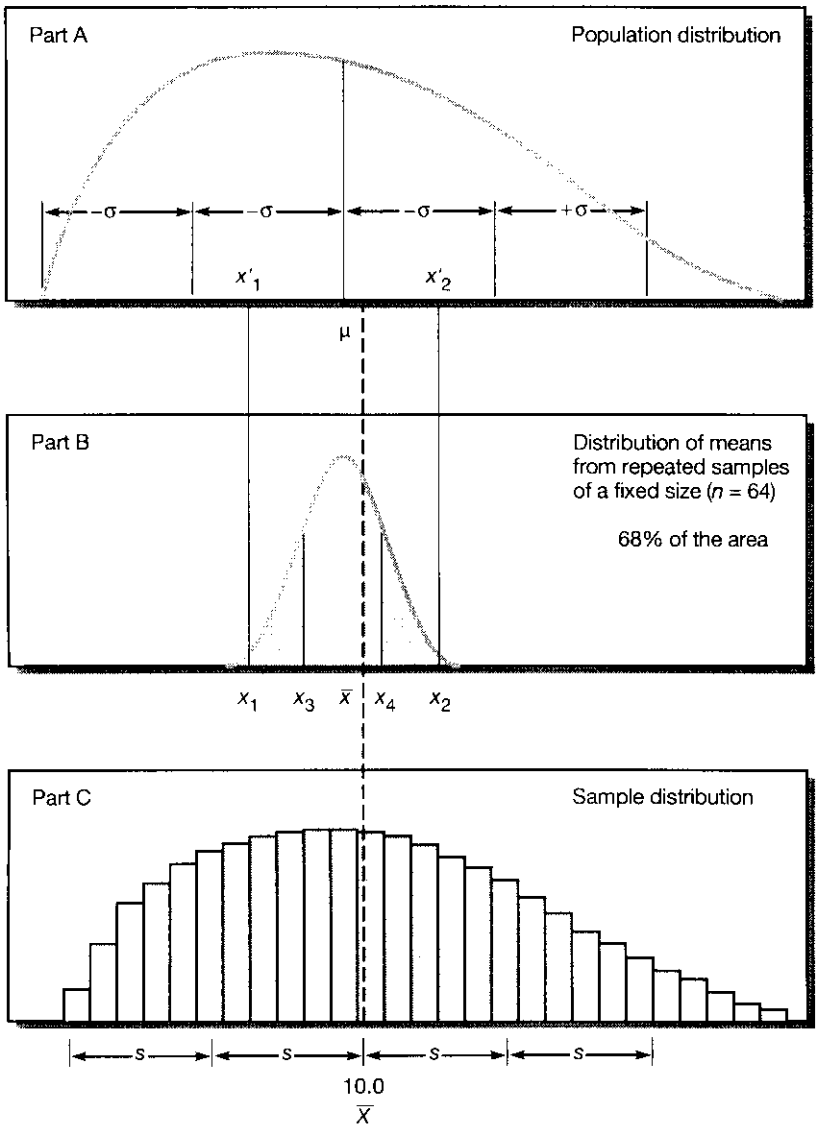
Standard Error (Z)	Percent of Area*	Approximate Degree of Confidence
1.00	68.27	68%
1.65	90.10	90
1.96	95.00	95
3.00	99.73	99

*Includes both tails in a normal distribution.

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.

Exhibit 14a-5 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.

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 14a-5. 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 14a-5) 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:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{4.1}{\sqrt{64}} = 0.51$$

If one standard error of the mean is equal to 0.51 visit, 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 mean visits per month. We discuss pilot tests as part of the pretest phase in Chapter 13.

The preceding 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 14a-6 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 visit, then 0.2 visit would be equal to 0.39 standard error (0.2/0.51 = 0.39). Referring to a table of areas under the normal curve (Appendix C, Exhibit C-1), we find that there is a 30.3 percent chance that the true population mean lies within ± 0.39 standard error of 10.0. With a sample of 64, the sample 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.

Before we compute the desired sample size for the Metro U dining club study, let's review the information we will need:

1. The *precision* desired and how to quantify it:
   a. The *confidence level* we want with our estimate.
   b. The *size of the interval estimate.*
2. The expected *dispersion in the population* for the investigative question used.
3. Whether a finite population adjustment is needed.

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 14a-7.

Exhibit 14a-6 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

Exhibit 14a-7 Metro U Sampling Design Decision on "Meal Frequency" and "Joining" Constructs

Sampling 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 (based on risk)	±0.5 meal per month	±0.10 (10 percent)
2. The expected range in the population for the question used to measure precision:	0 to 30 meals	0 to 100%
Measure of Central Tendency		
• Sample mean	10	
• Sample proportion of population with the given attribute being measured		30%
Measure of Dispersion		
• Standard deviation	4.1	
• Measure of sample dispersion		$pq = 0.30(0.70) = 0.21$
3. Whether a finite population adjustment should be used	No	No
4. Estimate of standard deviation of population:		
• Standard error of mean	0.5/1.96 = 0.255	
• Standard error of the proportion		0.10/1.96 = 0.051
5. Sample size calculation	See formula (p. 396)	See formula (p. 396)
6. Calculated sample size	$n = 259$*	$n = 81$

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

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 5-point measurement scale, one-tenth of a point is a very high degree of precision in comparison to a 1-point interval. Given that a patron could eat up to 30 meals per month at the dining club (30 days times 1 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 5 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.

Population Size

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:

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

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

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

$$n = \frac{(4.1)^2}{(0.255)^2}$$

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

where

$\sigma_{\bar{x}} = 0.255$ (0.51/1.96)

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 Questions Involving Proportions

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)[1]—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 these 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)

n = sample size

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

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

$$n = \frac{0.3 \times 0.7}{(0.051)^2}$$

$$n = 81$$

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 ($0.4 \times 0.6 = 0.24$, 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.

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

$$n = \frac{(0.50)(0.50)}{(0.51)^2}$$

$$n = \frac{0.25}{(0.051)^2}$$

$$n = 96$$

where

pq = measure of dispersion
n = sample size
σ_p = standard error of the proportion

If we use this maximum variance estimate in the dining club example, we find the sample size needs to be 96 persons in order to have an adequate sample for the question about joining the club.

When there are several investigative questions of strong interest, researchers calculate the sample size for each such variable—as we did in the Metro U study for "meal frequency" and "joining." The researcher then chooses the calculation that generates the largest sample. This ensures that all data will be collected with the necessary level of precision.