>chapter 15

Sampling

❝❝In a recent Gallup 'poll on polls,' . . . when asked about the scientific sampling foundation on which polls are based . . . most said that a survey of 1,500–2,000 respondents—a larger than average sample size for national polls—cannot represent the views of all Americans.❞❞

Frank Newport, The Gallup Poll's editor in chief, The Gallup Organization

>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.

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

"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 Sally hammer out the details of the Glacier Symphony sampling plan.

"Sam Champion, marketing director for CityBus," shares Sally, "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 by 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." Sally 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.'"

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

"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 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. While 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.

>snapshot

For the 2000 census, the U.S. Bureau of the Census proposed for the first time that the every-10-year census substitute statistical sampling to help adjust for the undercounting of some population segments. This undercounting includes renters, as well as immigrant and minority groups, who have been less responsive to past census requests for information. In January 1999, by a 5-4 decision, the U.S. Supreme Court banned sampling in census results for the purpose of federal apportionment but allowed it for counting population cohorts for business research purposes and for districting within states.

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

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

www.census.gov

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.

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 Mattress Example

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. We insert a survey response card in a sample of returned laptops. 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.
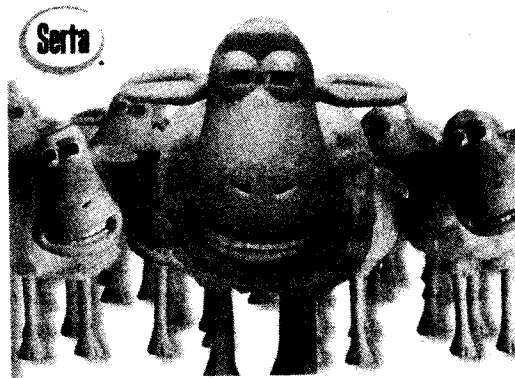
Precision is measured by the standard error of estimate, a type of standard deviation measurement; the smaller the standard error of estimate, the higher is the precision of the sample. The ideal sample design produces a small standard error of estimate. However, not all types of sample design provide 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 15-1. 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.

> **Exhibit 15-1** Sampling Design within the Research Process
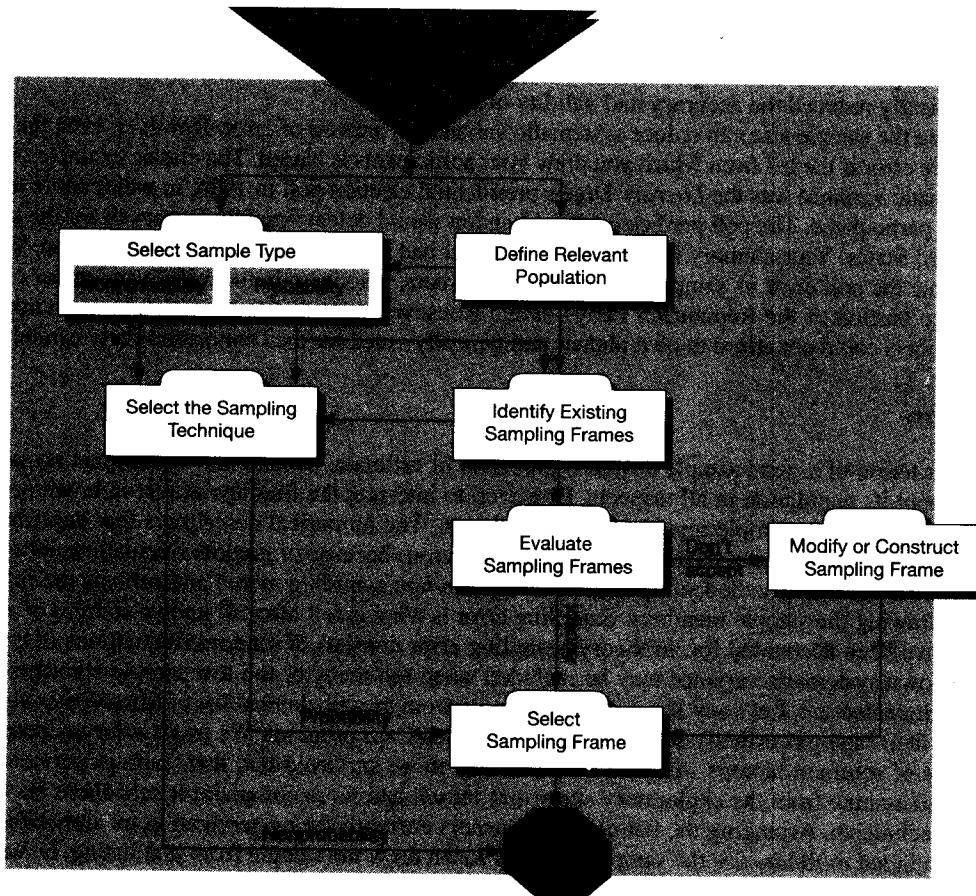
The researchers at Metro U are exploring the feasibility of creating a dining club whose facilities would be available on a membership basis. To launch this venture, they will need to make a substantial investment. Research will allow them to reduce many risks. Thus the research question is, "Would a membership dining club be a viable enterprise?" Some investigative questions that flow from the research question include:

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

We use the last three investigative questions for examples and focus specifically on questions 7 and 8 for assessing the project's risks. First, we will digress with other 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 15-2, classify the different approaches.

Representation

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. While 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, as 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

> **Exhibit 15-2** Types of Sampling Designs

Element Selection	Representation Basis	
	Probability	Nonprobability
Unrestricted	Simple random	Convenience

>snapshot
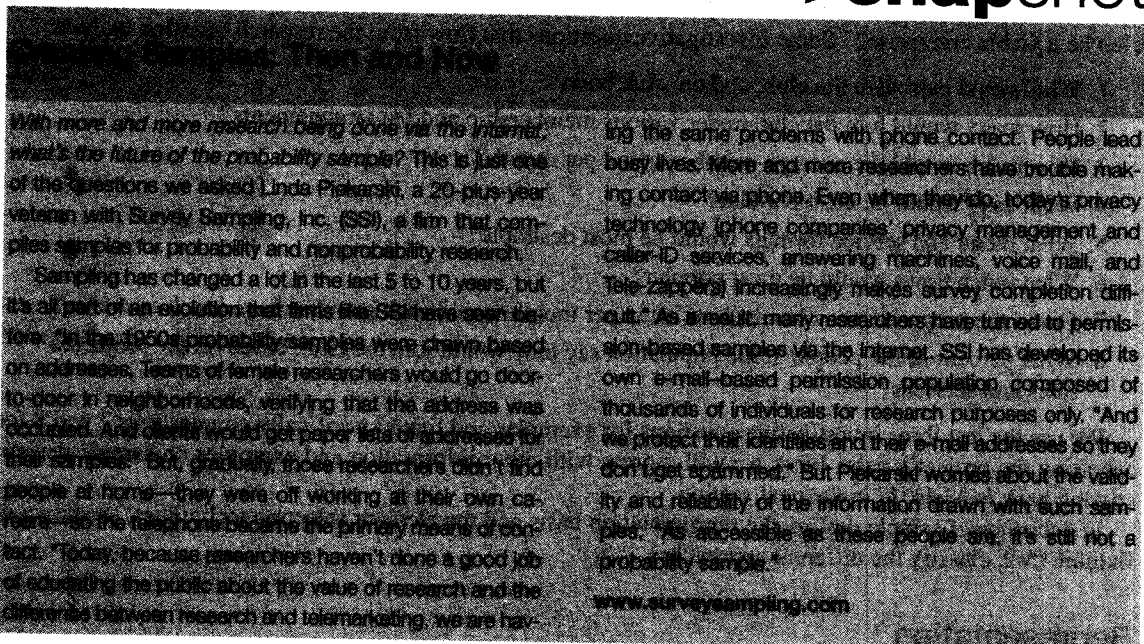
concern, however, is what the Bureau of the Census labels the "great digital divide"—low-income and ethnic subgroups' underrepresentation in their use of the Internet compared to the general population. Additionally, many Internet samples were, and still are, drawn substantially from panels. These are 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. While exploratory research does not necessarily demand this, explanatory, descriptive, and causal studies do.

Element Selection

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.

> Steps in Sampling Design

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.

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

< We discuss data types in greater detail in Chapters 12 and Appendix 15a. >

> **Exhibit 15-3** 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)
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)

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 15-3 indicates population parameters of interest for our three example studies.

> **We discuss proportion estimators in more detail in Appendix 15a.**

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 (for example, 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.

>snapshot

New Product Research Blind Spot

"There has always been a gap in new product research," claims Jim Lane, senior VP, Alliance Research, "between the time a new product is introduced and when it is possible to extract customer information." The reason? While new product marketers can track store sales audit data, locating purchasers to conduct attitudinal studies often lags as much as six months. Alliance Research combines its proprietary techniques for attitude measurement with parent company Catalina Marketing Corp.'s point-of-sale (POS) promotional system to develop a sample frame of purchasers of new products to participate in an interactive-voice-response (IVR) survey. Purchasers receive an invitation printed on the red-trimmed coupons that accompany their register receipts. For participation, Alliance compensates each respondent with up to $5, with current participation rates comparable to those for telephone studies. By inviting participants at the time of first purchase, the six-month wait for attitudinal data is history. Now this critical information is available to brand and marketing managers within weeks of a new product's hitting the shelf.

In what circumstances and why might immediate feedback not be an accurate measure of a new product's reception?

www.allianceresearch.com

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 decade ago, Chinese families with a home telephone were envied. By 2004, China's fixed telephone users totaled 295 million and cellular telephone service rose above 305 million subscribers in a population of 1.3 billion. Business or personal listings are inadequate for developing sampling frames when in a six-month period 68 million households add services.

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

< We discuss screening procedures in Chapter 14.

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. While you may use the general city directory, sampling from this too inclusive frame would be costly and inefficient, as Asians represent only a small fraction of Little Rock's population. The screening task would be monumental. Since ethnic groups frequently cluster in certain neighborhoods, you might identify these areas of concentration and then use a reverse area telephone or city directory, which is organized by street address, to draw the sample. Burbidge had a definite problem, as no sample frame of CityBus riders existed. While some regular riders used monthly passes, infrequent riders usually paid cash for their fares. It might have been possible for Burbidge to 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.

< **You may wish to review the discussions of error in Chapters 10 and 12.**

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.

While 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.

What Size Sample Is Needed?

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 process of determining the sample size is the subject of Appendix 16a.**

- 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.

> Probability Sampling

Simple Random Sampling

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 = .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 15-4 provides an overview of the steps involved in choosing a random sample.

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

> **Exhibit 15-4** 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 C, Exhibit C-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.

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 *k*th line of a listing sheet than it is to use a random numbers table. With systematic sampling, there is no need to number the entries in a large personnel file before drawing a sample. To draw a systematic sample, 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 *k*th 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 under-sample 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.

While systematic sampling has some theoretical problems, from a practical point of view it is usually treated as a simple random sample. When similar population elements are grouped within the sampling frame, systematic sampling is statistically more efficient than a simple random sample. This might occur if the listed elements are ordered chronologically, by size, by class, and so on. Under these conditions, the sample approaches a proportional stratified sample. The effect of this ordering is more pronounced on the results of cluster samples than for element samples and may call for a proportional stratified sampling formula.[12]

Stratified Sampling

Most populations can be segregated into several mutually exclusive subpopulations, or strata. The process by which the sample is constrained to include elements from each of the segments is called **stratified random sampling**. University students can be divided by their class level, school or major, gender, and so forth. After a population is divided into the appropriate strata, a simple random sample can be taken within each stratum. The 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 why a researcher chooses a stratified random sample: (1) to increase a sample's statistical efficiency, (2) to provide adequate data for analyzing the various subpopulations 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.

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.

>snapshot

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.** 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 table below 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.

Cluster Sampling

In a simple random sample, each population element is selected individually. The population can also be divided into groups of elements with some groups randomly selected for study. This is **cluster sampling.** Cluster sampling differs from stratified sampling in several ways, as indicated in Exhibit 15-5.

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.

> **Exhibit 15-5** Comparison of Stratified and Cluster Sampling

| Stratified Sampling	Cluster Sampling
1. We divide the population into a few subgroups.	1. We divide the population into *many* subgroups.
• Each subgroup has *many* elements in it.	• Each subgroup has *few* elements in it.
• Subgroups are selected according to some criterion that is related to the variables under study.	• 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.

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

Area Sampling Much research involves populations that can be identified with some geographic area. When this occurs, it is possible to use **area sampling,** the most important form of cluster sampling. This method overcomes 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?
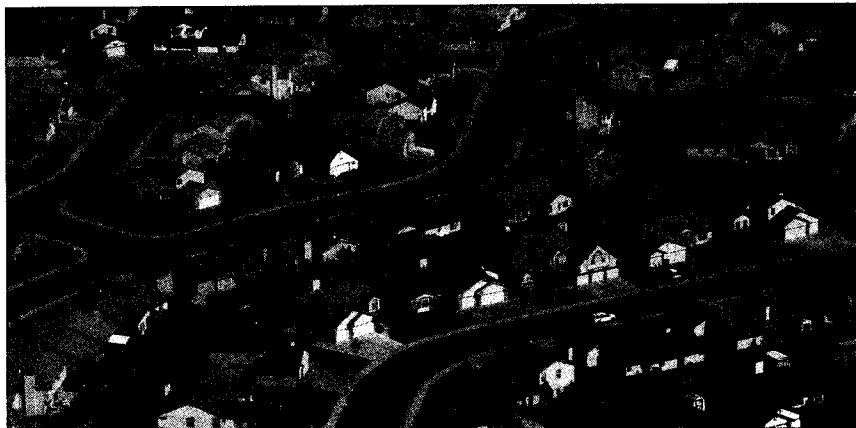
A low-cost, frequently used method, the area cluster sample may use geographic sample units (e.g., city blocks).

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.

Double Sampling

It may be more convenient or economical to collect some information by sample and then use this information as the basis for selecting a subsample for further study. This procedure is called **double sampling,**

sequential sampling, or multiphase sampling. It is usually found with stratified and/or cluster designs. The calculation procedures are described in more advanced texts.

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

Because of the wide range of sampling designs available, it is often difficult to select an approach that meets the needs of the research question and helps to contain the costs of the project. To help with these choices, Exhibit 15-6 may be used to compare the various advantages and disadvantages of probability sampling. Nonprobability sampling techniques are covered in the next section. They are used frequently and offer the researcher the benefit of low cost. However, they are not based on a theoretical framework and do not operate

> **Exhibit 15-6** 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.
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.
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.

>snapshot

The Truth in Florida

In 1997, Florida Department of Health (FDOH) officials decided to do something about an alarming trend: Teen smoking was increasing; some estimates placed the increase at as much as 7 percent in the prior five years. They decided to launch a public service advertising campaign (PSA) aimed at youth as part of a larger control effort to include marketing and public relations initiatives. The $140 million in funding would come from the proceeds of Florida's $11.3 billion legal settlement with the tobacco companies. FDOH turned to Crispin Porter & Bogusky (CP&B) to develop edgy PSAs that would resonate with youth. These ads had names, like "liars" and "junk yard dogs." The creative strategy would be to rebrand smoking or, as CP&B's Heydy Gomez described, "to create a brand as strong and as powerful [as the cool brands the tobacco industry had created] to take the place of tobacco brands." CP&B called the campaign *Truth.*

CP&B involved youths at every step of the process. At a Teen Tobacco Summit, 600 high school students brainstormed approaches that would work with teens. CP&B created a 45-member youth sounding board to serve as marketing advisers. CP&B also involved representatives from Students Working Against Tobacco (SWAT). It also devoted significant resources to determine the impact of the advertising, test advertising effectiveness, and evaluate new concepts as the campaign aged.

After two years, the Florida Truth campaign had achieved unprecedented success:

- Cigarette use dropped 40 percent among middle school students (from 18.8 to 11.1 percent).
- Cigarette use dropped 21 percent among high school students (from 27.4 to 22.6 percent).

- Never-used-tobacco behavior rose 39.8 percent (from 56.4 to 69.3 percent) among middle school students.
- Never-used-tobacco behavior rose 26 percent (from 39.1 to 43.1 percent) among high school students.
- Florida teens' awareness of the "truth" brand stood at 94 percent.

Given such evidence of the success of the Florida Truth campaign, it should be no surprise that the American Legacy Foundation, a public health foundation created as a result of the November 1998 master settlement agreement between state attorneys general and the tobacco companies, signed on to fund the national extension of Truth. But the foundation would want data to substantiate national effectiveness. The Research Triangle Institute fielded the National Youth Tobacco Survey, a self-administered survey taken by 35,828 middle and high school students. A baseline was established in 1999, with the first formal post-Truth data collection in the spring of 2000. Additionally, the Legacy Media Tracking Survey, a longitudinal study conducted with periodic national telephone surveys of 12- to 24-year-olds, monitors youths' exposure to Truth and their tobacco-related beliefs, attitudes, and behaviors. Baseline was administered in fall 1999 and winter 2000 before the national Truth campaign was launched. A second round of random-dialed phone interviews was conducted from April to July 2001 with more than 20,000 youths.

You can see some of the Truth ads at the Crispin Porter & Bogusky Web site and the survey instruments and data sets are available at the American Legacy Web site. What sampling designs were used in this effort?

www.cpbgroup.com; www.americanlegacy.org

from statistical theory; consequently, they produce selection bias and nonrepresentative samples. Despite these weaknesses, their widespread use demands their mention here.

> Nonprobability Sampling

< As we discussed in Chapter 8, qualitative research often uses purposive 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 not only reduce the chance for sampling error but also estimate the range of probable sampling error present.

With a subjective approach like nonprobability sampling, the probability of selecting population elements is unknown. There are a variety of ways to choose persons or cases to include in the sample. Often we allow

the choice of subjects to be made by field workers on the scene. When this occurs, there is greater opportunity for bias to enter the sample selection procedure and to distort the findings of the study. Also, we cannot estimate any range within which to expect the population parameter. Given the technical advantages of probability sampling over nonprobability sampling, why would anyone choose the latter? There are some practical reasons for using the less precise methods.

Practical Considerations

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

Additional reasons for choosing nonprobability over probability sampling are cost and time. Probability sampling clearly calls for more planning and repeated callbacks to ensure that each selected sample member is contacted. These activities are expensive. Carefully controlled nonprobability sampling often seems to give acceptable results, so the investigator may not even consider probability sampling. Burbidge's results from bus 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.

Methods

Convenience

Nonprobability samples that are unrestricted are called **convenience samples.** They are the least reliable design but normally the cheapest and easiest to conduct. Researchers or field workers have the freedom to choose whomever they find, thus the name "convenience." Examples include informal pools of friends and

neighbors, people responding to a newspaper's invitation for readers to state their positions on some public issue, a TV reporter's "person-on-the-street" intercept interviews, or the use of employees to evaluate the taste of a new snack food.

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

Purposive Sampling

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

Judgment sampling occurs when a researcher selects sample members to conform to some criterion. In a study of labor problems, you may want to talk only with those who have experienced on-the-job discrimination. Another example of judgment sampling occurs when election results are predicted from only a few selected precincts that have been chosen because of their predictive record in past elections. Burbidge chose bus 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 undergrad-
uate 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 represen-
tativeness on others is argument by analogy. It gives no assurance that the sample is representative of the vari-
ables 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 look-
ing 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 busi-
ness researchers. Probability sampling is usually much more costly and time-consuming. Advocates of quota
sampling argue that while there is some danger of systematic bias, the risks are usually not that great. Where
predictive validity has been checked (e.g., in election polls), quota sampling has been generally satisfactory.

Snowball

This design has found a niche in recent years in applications where respondents are difficult to identify and
are best located through referral networks. 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 prob-
ability 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" gath-
ers 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 dig-
ital signal processing is being contemplated by one firm. Through its contacts with a select group of record-
ing 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 while some elements in a sample underestimate a population value, others overestimate this value. The result of these tendencies is that a sample statistic such as the arithmetic mean is generally a good estimate of a population mean.

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

3 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 as shown in the accompanying table:

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

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 419	population 402	sample statistics 409
census 402	population element 402	sampling 402
cluster sampling 418	population parameters 409	sampling error 405
convenience sample 423	population proportion of incidence 410	sequential sampling 421
disproportionate stratified sampling 418		simple random sample 414
	probability sampling 408	skip interval 414
double sampling 421	proportionate stratified sampling 417	snowball sampling 425
judgment sampling 424		stratified random sampling 416
multiphase sampling 421	quota sampling 424	systematic sampling 414
nonprobability sampling 407	sample frame 402	systematic variance 405

>discussionquestions

Terms in Review

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?

Making Research Decisions

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 ad-

vance, 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, where 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?

Bringing Research to Life

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?

From Concept to Practice

10 Using Exhibit 15-6 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.

>wwwexercise

Visit the Web site of the Henry J. Kaiser Family Foundation, a nonprofit organization that studies critical issues related to health and health services (**www.kff.org**). Select one of the foundation's studies and evaluate the sampling plan. If you were researching a product or service related to the issue being studied, how would you adjust the sampling plan?

cases*

Calling Up Attendance

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

Can Research Rescue the Red Cross?

Inquiring Minds Want to Know—NOW!

State Farm: Dangerous Intersections

USTA: Come Out Swinging

* All cases appear on the text CD; you will find abstracts of these cases in the Case Abstracts section of this text. Video cases are indicated with a video icon.

>appendix15a

Determining Sample Size

Basic Concepts for Sampling

In the Metro University Dining Club study, we explore probability sampling and the various concepts used to design the sampling process.
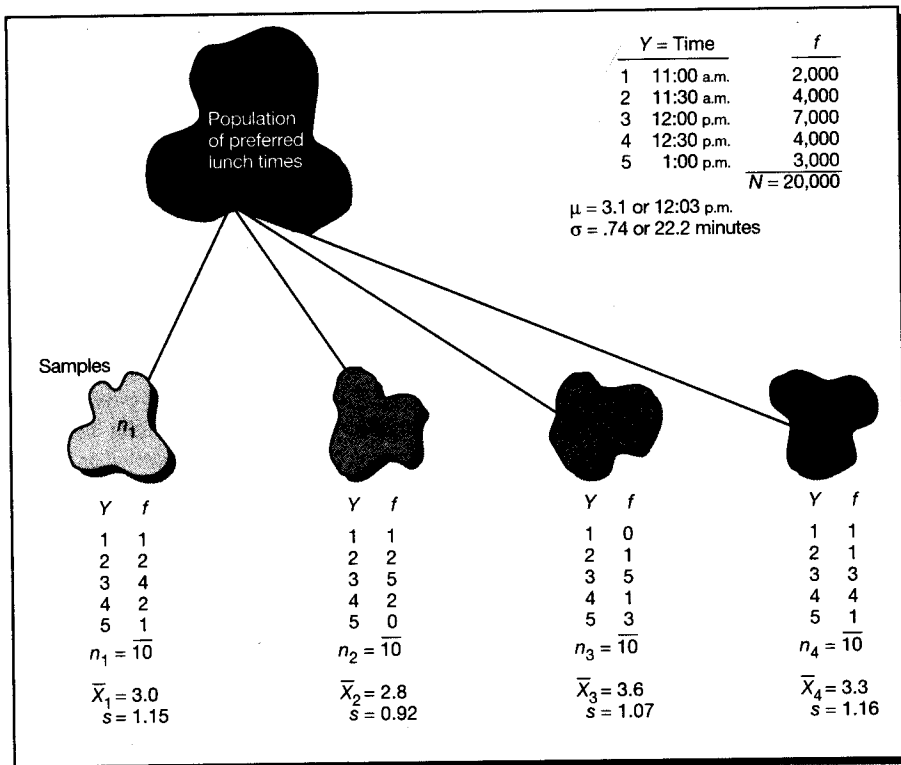
Exhibit 15a-1 shows the Metro U dining club study population ($N = 20,000$) consisting of five subgroups based on their preferred lunch times. The values 1 through 5 represent the preferred lunch times of 11 a.m., 11:30 a.m., 12 noon, 12:30 p.m., and 1 p.m. The frequency of response (f) in the population 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.

Point Estimates

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 15a-1. We also calculate a mean score, $X_1 = 3.0$, for this sample. This mean would place the average preferred lunch time at 12 noon. The mean is a *point estimate* and

> **Exhibit 15a-1** Random Samples of Preferred Lunch Times

Population of preferred lunch times

Y = Time		f
1	11:00 a.m.	2,000
2	11:30 a.m.	4,000
3	12:00 p.m.	7,000
4	12:30 p.m.	4,000
5	1:00 p.m.	3,000
		N = 20,000

$\mu = 3.1$ or 12:03 p.m.
$\sigma = .74$ or 22.2 minutes

Samples

Y	f		Y	f		Y	f		Y	f
1	1		1	1		1	0		1	1
2	2		2	2		2	1		2	1
3	4		3	5		3	5		3	3
4	2		4	2		4	1		4	4
5	1		5	0		5	3		5	1

$n_1 = 10$ $\qquad$ $n_2 = 10$ $\qquad$ $n_3 = 10$ $\qquad$ $n_4 = 10$

$\bar{X}_1 = 3.0$ $\qquad$ $\bar{X}_2 = 2.8$ $\qquad$ $\bar{X}_3 = 3.6$ $\qquad$ $\bar{X}_4 = 3.3$
$s = 1.15$ $\qquad$ $s = 0.92$ $\qquad$ $s = 1.07$ $\qquad$ $s = 1.16$

our best predictor of the unknown population mean, μ (the arithmetic average of the population). Assume further that we return the first sample to the population and draw a second, third, and fourth sample by the same procedure. The frequencies, means, and standard deviations are as shown in the exhibit. As the data suggest, each sample shares some similarities with the population, but none is a perfect duplication because no sample perfectly replicates its population.

Interval Estimates

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 15a-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 15a-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:

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

Estimating the Population Mean

How does this improve our prediction of μ from $\bar{X}$? The standard error creates the interval range that brackets the point estimate. In this example, μ is predicted to be 3.0 or 12 noon (the mean of n_1) $\pm$ .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 .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 $\pm$ 1 Z or 68 percent of the area under the normal curve (see Exhibit 15a-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 .36 by $\pm$ 1.96 (Z), since 1.96 Z covers 95 percent of the area under the curve (see Exhibit 15a-4). Now, with 95 percent confidence, the interval in which we would find the true mean increases to $\pm$ .70 (from 2.3 to 3.7 or from 11:39 a.m. to 12:21 p.m.).

True mean = 3.1

2.64	3.00	3.36
11:49 a.m.	$\bar{X}$	12:11 p.m.
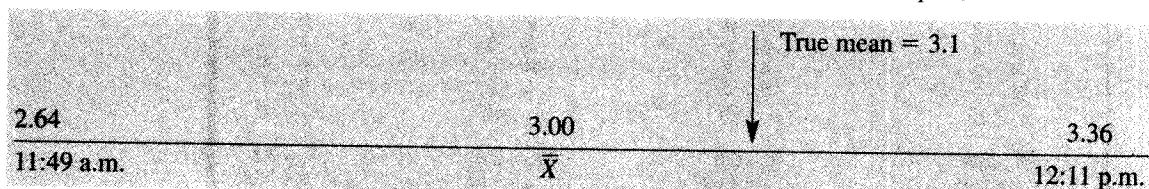
> **Exhibit 15a-2** Effects on Standard Error of Mean of Increasing Precision

	Reducing the Standard Deviation by 50%	Quadrupling the Sample
$\sigma_{\bar{x}} = \dfrac{s}{\sqrt{n}}$	$\sigma_{\bar{x}} = \dfrac{.74}{\sqrt{10}} = .234$	$\sigma_{\bar{x}} = \dfrac{.8}{\sqrt{25}} = .16$
	$\sigma_{\bar{x}} = \dfrac{.37}{\sqrt{10}} = .117$	$\sigma_{\bar{x}} = \dfrac{.8}{\sqrt{100}} = .08$

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

Note: A 400% increase in sample size (from 25 to 100) would yield only a 200% increase in precision (from .16 to .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 15a-3** Confidence Levels and the Normal Curve
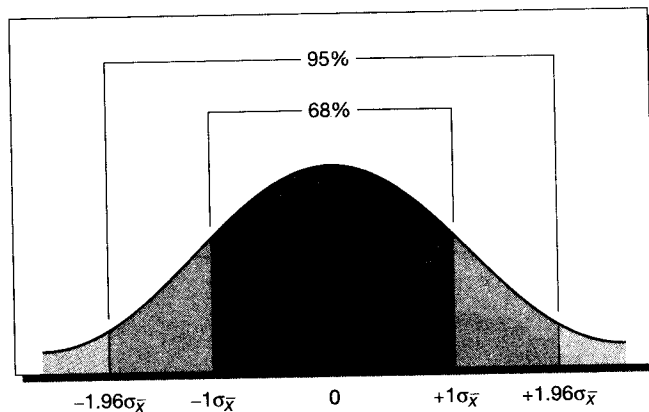
> **Exhibit 15a-4** Standard Errors Associated with Areas under the Normal Curve

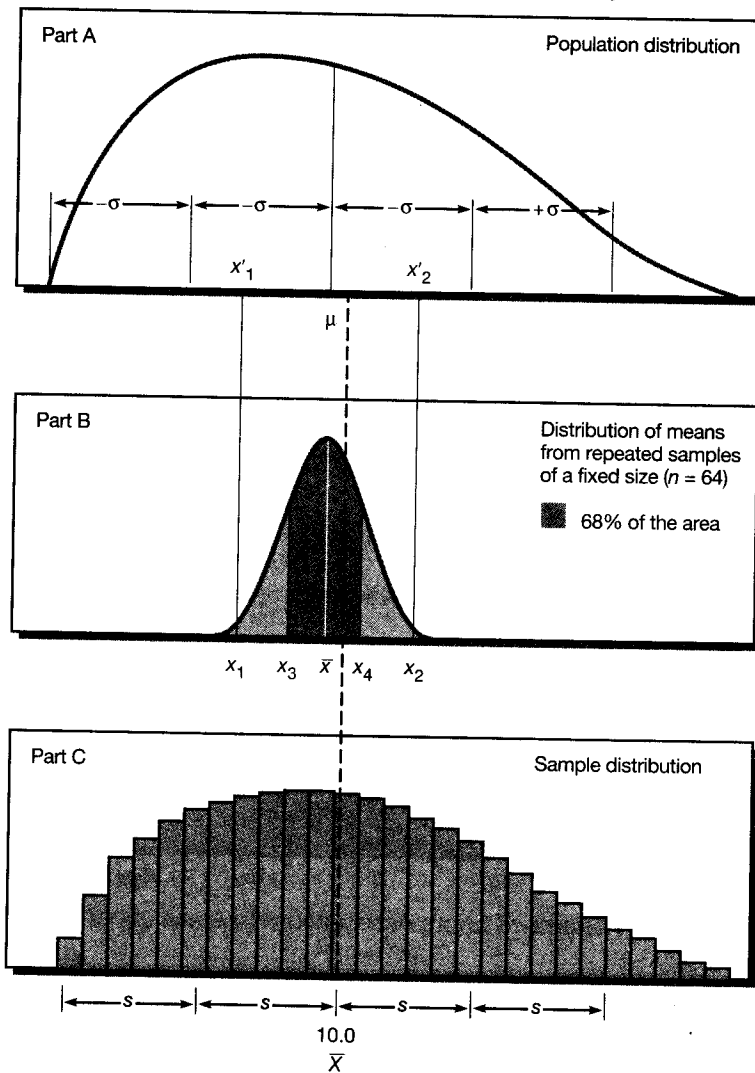| Standard Error (Z)	Percent of Area*	Approximate Degree of Confidence
1.00	68.27	68%
1.65	90.10	90
1.96	95.00	95
3.00	99.73	99

*Includes both tails in a normal distribution.

Parenthetically, if we compute the standard deviation of the distribution of sample means in Exhibit 15a-1, [3.0, 2.8, 3.6, 3.3], we will discover it to be .35. Compare this to the standard error from the original calculation (.36). The result is consistent 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

> **Exhibit 15a-5** A Comparison of Population Distribution, Sample Distribution, and Distribution of Sample Means of Metro U Dining Club Study

Part A Population distribution

$-\sigma$ | $-\sigma$ | $-\sigma$ | $+\sigma$

x'_1 | x'_2

μ

Part B Distribution of means from repeated samples of a fixed size ($n = 64$)

▓ 68% of the area

x_1 | x_3 | $\bar{x}$ | x_4 | x_2

Part C Sample distribution

s | s | s | s

10.0

$\bar{X}$

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

the club?" they would get a distribution something like that shown in part A of Exhibit 15a-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). Turning to part C of Exhibit 15a-5,

three observations about this sample distribution are consistent with our earlier illustration. First, it is shown as a histogram; it represents a frequency distribution of empirical data, while the smooth curve of part A is a theoretical distribution. Second, the sample distribution (part C) is similar in appearance but is not a perfect duplication of the population distribution (part A). Third, the mean of the sample differs from the mean of the population.

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

Estimating the Interval for the Metro U Dining Club Sample

Any sample mean will fall within the range of the distribution extremes shown in part B of Exhibit 15a-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 15a-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:

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

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 14.

Changing Confidence Intervals

The above estimate may not be satisfactory in two ways. First, it may not represent the degree of confidence the researchers want in the interval estimate, considering their financial risk. They might want a higher degree of confidence than the 95 percent level used here. By referring to a table of areas under the normal curve, they can find various other combinations of probability. Exhibit 15a-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, $\bar{X} \pm 3\sigma_X$. Now the population mean lies somewhere between $10.0 \pm 3(0.51)$ 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.

> **Exhibit 15a-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 9.49 and 10.52 visits
90	μ is between 9.14 and 10.86 visits
95	μ is between 8.99 and 11.02 visits
99	μ is between 8.44 and 11.56 visits

> **Exhibit 15a-7** Metro U Sampling Design Decision on "Meal Frequency" and "Joining" Constructs

Sampling Issues	Metro U Decisions	
	"Meal Frequency" (interval, ratio data)	"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)	± .5 meal per month	± .10 (10 percent)
Measure of Dispersion		
• Standard deviation	4.1	
• Measure of sample dispersion		pq = .30(.70) = .21
4. Estimate of standard deviation of population:		
• Standard error of mean	.5/1.96 = 0.255	
• Standard error of the proportion		.10/1.96 = 0.051
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.		

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.

Calculating the Sample Size for Questions Involving Means

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 15a-7.

Precision

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.

Population Dispersion

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:

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

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

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

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

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

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

If the researchers are willing to accept a larger interval range ($\pm$ 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 $\pm$ 0.10 of the sample proportion. The calculation of the sample size proceeds as before:

$\pm 0.10 =$ 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

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

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

$$n = \frac{0.3 \times 0.7}{(.051)^2}$$

$$n = 81$$

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.

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

$$n = \frac{(.50)(.50)}{(0.51)^2}$$

$$n = \frac{.25}{(0.051)^2}$$

$$n = 96$$

where

$pq =$ measure of dispersion

$n =$ sample size

$\sigma_p =$ 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.

>part IV

Analysis and Presentation of Data

>chapter 16

Data Preparation and Description

❝❝In the future, we'll stop moaning about the lack of perfect data and start using the good data with much more advanced analytics and data-matching techniques.❞❞

Kate Lynch, research director, Leo Burnett's Starcom Media Unit

>learningobjectives

After reading this chapter, you should understand . . .

1 The importance of editing the collected raw data to detect errors and omissions.

2 How coding is used to assign numbers and other symbols to answers and to categorize responses.

3 The use of content analysis to interpret and summarize open questions.

4 Problems with and solutions for "don't know" responses and missing data.

5 The options for data entry and manipulation.

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

Dorrie Henry arrives for lunch, joining Jason and Sally, who are wrapping up a MindWriter meeting. The conversation naturally turns to one of Jason's early research projects.

"As a public health doctor," said Dorrie, "I work with statistics—life-threatening ones. I understand Jason's perfectionism and why he carries it to such extremes. You know how 'difficult' he can be when he is under stress."

Sally responds by arching her eyebrows and laughing mirthlessly.

"I can see this turning into a bashing-Jason story," Jason groans.

"Has Jason told you about our near blowup over the bus line survey? No? Jason was retained by a bus line up north to assist in a sample survey of 2,000 passengers. The line has just dismissed its research director, a horrible chap named Burbidge, who could not relate to the blue-collar riders.

"Jason had a 20-question survey of 2,000 randomly selected riders, a total of 40,000 items to be coded, some of them multiple choice, some of them numerical entry, some free response. He estimated this would take 15 seconds an item, on the average, a total of 10,000 minutes, which he rounded up to 12,000 minutes. Call it 200 person-hours, then. He decided to hire 25 temporary personnel for one eight-hour data entry shift."

"That sounds reasonable," says Sally.

"Well, of course it was quite reasonable. Since CityBus receives a county subsidy, the bus line had to follow county purchasing procedures and put the data entry out for competitive bidding.

"As board member of a nonprofit organization that provides temporary employment for former mental patients, I bid on the data entry work. He had never expressed any objections to my finding work for these people before, but when I wanted them involved in his own work, he became prickly, hostile, self-pitying, irritable, and hypercritical."

"But what motivated his attitude was the most intense desire for accuracy and reliability and he ceaselessly worries about threats to validity."

"Which, as his partner, I not only appreciate but demand," supports Sally.

"I won the CityBus contract, which had me bringing my new crew in at 5 p.m. on a Friday evening to bus line HQ, putting each worker in a separate office, using 25 networked computers in employee offices. Each member of the crew, sitting in a separate office, was given 80 surveys to encode, 10 an hour, for eight hours.

"Jason procured a data entry program, which we ran off the system server. The program very cleverly contrived to prevent errors of data entry, so far as possible. For instance, if a survey item—a question—called for choosing between 'Ride the bus less than 8 times,' or 'Ride the bus 8 or more times,' the data clerk would put a cursor over the correct answer and press 'enter.' Making a mistake would be difficult, because the screen would closely resemble the printed page. If the item called for 'Enter your age,' and the clerk entered a three-digit number, the

computer would beep loudly, to indicate that ages less than 10 and greater than 99 were considered suspect.

"Jason continued to face the prospect of data entry night with thinly disguised opposition. He grumbled that my crew would ruin the results of his carefully crafted research. I reminded him that these people were not visually impaired and had no intellectual deficits. They were just folks who were not yet ready to reenter the workplace or were ready but suffered unreasonable discrimination, of exactly the sort Jason was showing.

"Anyway, while it was prudent to warn the HQ staff to secure their offices, Jason went overboard and infected the office staff with his nervousness and negativity about my crew coming in to use their computers. At 4:45 p.m., our bus arrived with me, my crew, and a supervisory social worker, to find office workers departing carrying out their potted plants. Did they think my crew were herbivores? It was a flagrant insult, and I felt awful."

"You are going to tell the part where I redeem myself, aren't you?" asks Jason plaintively.

"Jason, concerned with human error, double-checked the data entry. A 26th crew member ran up and down the halls all night retrieving randomly sampled surveys, which Jason compared with the already entered data in the server. He warned me that if he found more than a certain error rate he intended

to stop the operation and send everyone home. He was working himself into a lather, looking for trouble he was not going to find.

"As he was pulling the forms generated by 25 workers, he had to work at least as fast as any one of them to keep up, slightly faster in fact, and very frankly I had serious doubts that he would not make more errors than my crew and blame the errors on them. He skipped supper at 8 p.m. and worked steadily and crankily and drank too much black coffee. No bathroom breaks or stretch breaks for super sleuth. But at 10 p.m., five hours into our eight-hour shift, he threw down his surveys, smiled broadly, murmured 'darn good.' strolled off to the bathroom, returned, ate two peanut butter and jelly sandwiches, and made a few dumb jokes.

"It seems," continues Dorrie, "the discrepancy rate between his coding and my crew was less than 1 percent for the first five hours."

"In the end, Jason expressed considerable satisfaction with your crew, right?" inquires Sally.

"Of course. After all, when HQ office workers came back Monday, toting their philodendrons, they found their offices clean, orderly, and mysteriously mint-scented. Each of my crew had cleaned the keyboards and monitor screens with little sanitized towelettes. And Jason had error-free data in the bargain."

> Introduction

Once the data begin to flow, a researcher's attention turns to data analysis. **Data preparation** includes editing, coding, and data entry and is the activity that ensures the accuracy of the data and their conversion from raw form to reduced and classified forms that are more appropriate for analysis. Preparing a descriptive statistical summary is another preliminary step leading to an understanding of the collected data. It is during this step that data entry errors may be revealed and corrected. Exhibit 16-1 reflects the steps in this phase of the research process.

> **Exhibit 16-1** Data Preparation in the Research Process
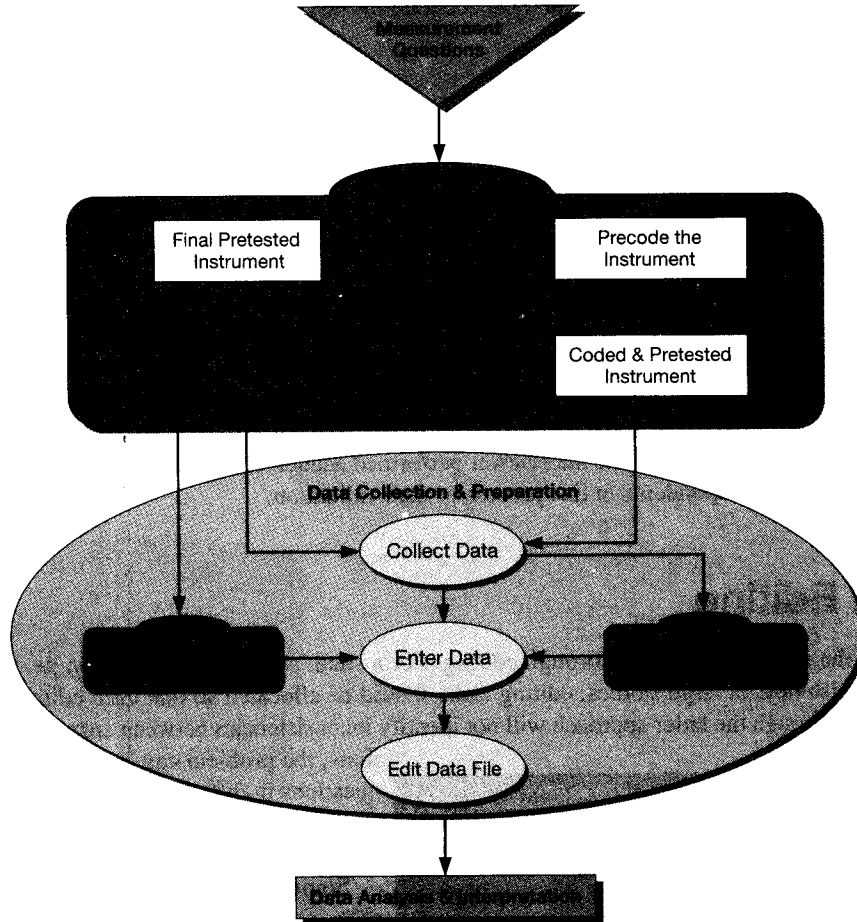
> # Editing

The customary first step in analysis is to edit the raw data. **Editing** detects errors and omissions, corrects them when possible, and certifies that maximum data quality standards are achieved. The editor's purpose is to guarantee that data are:

- Accurate.
- Consistent with the intent of the question and other information in the survey.
- Uniformly entered.
- Complete.
- Arranged to simplify coding and tabulation.

In the following question asked of adults 18 or older, one respondent checked two categories, indicating that he was a retired officer and currently serving on active duty.

Please indicate your current military status:

☒ Active duty	☒ Reserve	☐ Retired
☐ National Guard	☐ Separated	☐ Never served in the military

The editor's responsibility is to decide which of the responses is both consistent with the intent of the question or other information in the survey and most accurate for this individual participant.

Field Editing

In large projects, field editing review is a responsibility of the field supervisor. It, too, should be done soon after the data have been gathered. During the stress of data collection in a personal interview and paper-and-pencil recording in an observation, the researcher often uses ad hoc abbreviations and special symbols. Soon after the interview, experiment, or observation, the investigator should review the reporting forms. It is difficult to complete what was abbreviated or written in shorthand or noted illegibly if the entry is not caught that day. When entry gaps are present from interviews, a callback should be made rather than guessing what the respondent "probably would have said." Self-interviewing has no place in quality research.
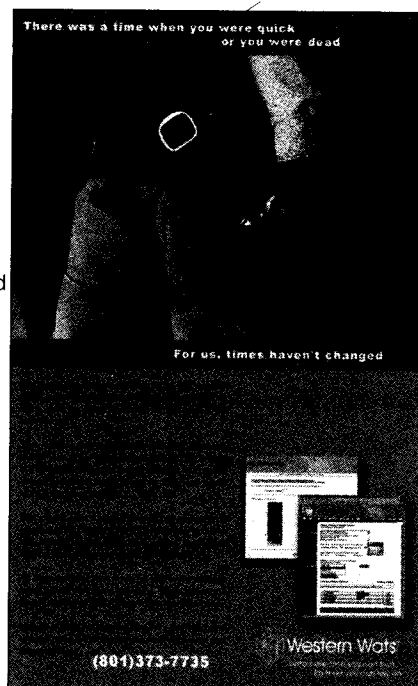
A second important control function of the field supervisor is to validate the field results. This normally means he or she will reinterview some percentage of the respondents, at least on some questions, verifying that they have participated and that the interviewer performed adequately. Many research firms will recontact about 10 percent of the respondents in this process of data validation.

Central Editing

At this point, the data should get a thorough editing. For a small study, the use of a single editor produces maximum consistency. In large studies, editing tasks should be allocated so that each editor deals with one entire section. Although the latter approach will not identify inconsistencies between answers in different sections, the problem can be handled by identifying questions in different sections that might point to possible inconsistency and having one editor check the data generated by these questions.

Western Wats, a data collection specialist, reminds us that speed without accuracy won't help a researcher choose the right direction. "After all, being quick on the draw doesn't do any good if you miss the mark." **www.westernwats. com**

Sometimes it is obvious that an entry is incorrect—for example, when data clearly specify time in days (e.g., 13) when it was requested in weeks (you expect a number of 4 or less)—or is entered in the wrong place. When replies are inappropriate or missing, the editor can sometimes detect the proper answer by reviewing the other information in the data set. This practice, however, should be limited to the few cases where it is obvious what the correct answer is. It may be better to contact the respondent for correct information, if time and budget allow. Another alternative is for the editor to strike out the answer if it is inappropriate. Here an editing entry of "no answer" or "unknown" is called for.

Another problem that editing can detect concerns faking an interview that never took place. This "armchair interviewing" is difficult to spot,

but the editor is in the best position to do so. One approach is to check responses to open-ended questions. These are most difficult to fake. Distinctive response patterns in other questions will often emerge if data falsification is occurring. To uncover this, the editor must analyze as a set the instruments used by each interviewer.

Here are some useful rules to guide editors in their work:

• Be familiar with instructions given to interviewers and coders.

• Do not destroy, erase, or make illegible the original entry by the interviewer; original entries should remain legible.

• Make all editing entries on an instrument in some distinctive color and in a standardized form.

• Initial all answers changed or supplied.

• Place initials and date of editing on each instrument completed.

> Coding

Coding involves assigning numbers or other symbols to answers so that the responses can be grouped into a limited number of categories. In coding, *categories* are the partitions of a data set of a given variable (for example, if the variable is *gender,* the partitions are *male* and *female*). *Categorization* is the process of using rules to partition a body of data. Both closed and free-response questions must be coded.

The categorization of data sacrifices some data detail but is necessary for efficient analysis. Most statistical and banner/table software programs work more efficiently in the *numeric* mode. Instead of entering the word *male* or *female* in response to a question that asks for the identification of one's gender, we would use numeric codes (for example, 0 for male and 1 for female). Numeric coding simplifies the researcher's task in converting a nominal variable, like gender, to a "dummy variable," a topic we discuss in Chapter 20. Statistical software also can use alphanumeric codes, as when we use M and F, or other letters, in combination with numbers and symbols for gender.

QUAL2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Exceptional Quality	12	14.5	15.0	15.0
	Very Good Quality	40	48.2	50.0	65.0
	Good Quality	16	19.3	20.0	85.0
	Average Quality	10	12.0	12.5	97.5
	Poor Quality	2	2.4	2.5	100.0
	Total	80	96.4	100.0	
Missing	System	3	3.6		
Total		83	100.0		

QUAL3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Exceptional Quality	6	7.2	7.5	7.5
	Very Good Quality	34	41.0	42.5	50.0
	Good Quality	27	32.5	33.8	83.8
	Average Quality	13	15.7	16.3	100.0
	Total	80	96.4	100.0	
Missing	System	3	3.6		
Total		83	100.0		

The researcher here requested a frequency printout of all variables when 83 cases had been entered. SPSS presents them sequentially in one document. The left frame indicates all the variables included in this particular output file. Both variables Qual2 and Qual3 indicate 3 missing cases. This would be a cautionary flag to a good researcher. During editing the researcher would want to verify that these are true instances where participants did not rate the quality of both objects, rather than data entry errors. **www.spss.com**

># >**snap**shot

CBS: Some Labs Are Extraordinary

Visitors to Las Vegas have an opportunity to determine the direction of CBS programming by visiting the CBS Television City Research Center in the MGM Grand Hotel and Casino. According to Andrew Wing, president of ACNielsen Entertainment, what makes Las Vegas an ideal location for a research lab is the cross section of the American population and the large number of international citizens among its 36 million visitors each year. In a typical screening with 250 people, individuals represent more than 40 states and every conceivable lifestyle. Participants watch a 30- to 45-minute segment of a new or proposed program, without commercial interruptions, followed by a survey process lasting 15 minutes. Each seat is equipped with a computer touch screen linked to ACNielsen Entertainment's proprietary Nielsen ReelResearch Internet site. Participants share feedback on the show and personal demo-

graphics in real time, while network executives observe participants and their feedback from remote offices around the country. The facility, designed by GES, also provides focus group capabilities, used for follow-up interviews along with other research initiatives. Participants are compensated with a chance to win a home theater system and a $10 coupon that they may redeem on program-logo T-shirts, caps, pins, key chains, even computer software in the CBS Television City store. Arising out of a temporary test conducted in 1991, the research facility today operates 12 hours per day year-round. So on your next visit to the City of Lights, when the slots or big-name entertainers lose appeal, entertain yourself with research.

What are some of the advantages to having touch-screen data entry?

www.acnielsen.com; www.viad.com

Codebook Construction

A **codebook,** or *coding scheme,* contains each variable in the study and specifies the application of coding rules to the variable. It is used by the researcher or research staff to promote more accurate and more efficient data entry. It is also the definitive source for locating the positions of variables in the data file during analysis. In many statistical programs, the coding scheme is integral to the data file. Most codebooks—computerized or not—contain the question number, variable name, location of the variable's code on the input medium (e.g., spreadsheet or SPSS data file), descriptors for the response options, and whether the variable is alphabetic or numeric. An example of a paper-based codebook is shown in Exhibit 16-2. Pilot testing of an instrument provides sufficient information about the variables to prepare a codebook. A preliminary codebook used with pilot data may reveal coding problems that will need to be corrected before the data for the final study are collected and processed.

Coding Closed Questions

The responses to closed questions include scaled items for which answers can be anticipated. Closed questions are favored by researchers over open-ended questions for their efficiency and specificity. They are easier to code, record, and analyze. When codes are established in the instrument design phase of the research process, it is possible to precode the questionnaire during the design stage. With computerized survey design, and computer-assisted, computer-administered, or online collection of data, precoding is necessary as the software tallies data as they are collected. **Precoding** is particularly helpful for manual data entry (for example, from mail or self-administered surveys) because it makes the intermediate step of completing a *data entry coding sheet* unnecessary. With a precoded instrument, the codes for variable categories are accessible directly from the questionnaire. A participant, interviewer, field supervisor, or researcher (depending on the data collection method) is able to assign the appropriate code on the instrument by checking, circling, or printing it in the proper coding location.

> **Exhibit 16-2**   Sample Codebook of Questionnaire Items

Question	Variable Number	Code Description	Variable Name
___	1	Record number	RECNUM
___	2	Respondent number	RESID
1	3	5 digit zip code 99999 = Missing	ZIP
2	4	2 digit birth year 99 = Missing	BIRTH
3	5	Gender 1 = Male 2 = Female 9 = Missing	GENDER
4	6	Marital status 1 = Married 2 = Widow(er) 3 = Divorced 4 = Separated 5 = Never married 9 = Missing	MARITAL
5	7	Own-Rent 1 = Own 2 = Rent 3 = Provided 9 = Missing	HOUSING
6		Reason for purchase 1 = Mentioned 0 = Not mentioned	
	8	Bought home	HOME
	9	Birth of child	BIRTHCHD
	10	Death of relative or friend	DEATH
	11	Promoted	PROMO
	12	Changed job/career	CHGJOB
	13	Paid college expenses	COLLEXP
	14	Acquired assets	ASSETS
	15	Retired	RETIRED
	16	Changed marital status	CHGMAR
	17	Started business	STARTBUS
	18	Expanded business	EXPBUS
	19	Parent's influence	PARENT
	20	Contacted by agent	AGENT
	21	Other	OTHER

Exhibit 16-3 shows questions in the sample codebook. When precoding is used, editing precedes data processing. Note question 4, where the respondent may choose between five categories of marital status and enter the number of the item best representing present status in the coding portion of the questionnaire. This code is later transferred to an input medium for analysis.

Coding Free-Response Questions

One of the primary reasons for using open questions is that insufficient information or lack of a hypothesis may prohibit preparing response categories in advance. Researchers are forced to categorize responses after

> **Exhibit 16-3** Sample Questionnaire Items

the data are collected. Other reasons for using open-ended responses include the need to measure sensitive or disapproved behavior, discover salience or importance, or encourage natural modes of expression.[1] Also, it may be easier and more efficient for the participant to write in a known short answer rather than read through a long list of options. Whatever the reason for their use, analyzing enormous volumes of open-ended questions slows the analysis process and increases the opportunity for error. The variety of answers to a single question can be staggering, hampering postcollection categorization. Even when categories are anticipated and precoded for open-ended questions, once data are collected researchers may find it useful to reassess the predetermined categories. One example is a 7-point scale where the researcher offered the participant three levels of agreement, three levels of disagreement, and one neutral position. Once the data are collected, if these finer nuances of agreement do not materialize, the editor may choose to recategorize the data into three levels: one level of agreement, one level of disagreement, and one neutral position.

Exhibit 16-3, question 6, illustrates the use of an open-ended question for which advance knowledge of response options was not available. The answer to "What prompted you to purchase your most recent life insurance policy?" was to be filled in by the participant as a short-answer essay. After preliminary evaluation, response categories (shown in the codebook, Exhibit 16-2) were created for that item.

Coding Rules

Four rules guide the pre- and postcoding and categorization of a data set. The categories within a single variable should be:

- Appropriate to the research problem and purpose.
- Exhaustive.
- Mutually exclusive.
- Derived from one classification principle.

Researchers address these issues when developing or choosing each specific measurement question. One of the purposes of pilot testing of any measurement instrument is to identify and anticipate categorization issues.

Appropriateness

Appropriateness is determined at two levels: (1) the best partitioning of the data for testing hypotheses and showing relationships and (2) the availability of comparison data. For example, when actual age is obtained (ratio scale), the editor may decide to group data by age ranges to simplify pattern discovery within the data. The number of age groups and breadth of each range, as well as the endpoints in each range, should be determined by comparison data—for example, U.S. census age ranges, a customer database that includes age ranges, or the age data available from Fox TV used for making an advertising media buy.

Exhaustiveness

Researchers often add an "other" option to a measurement question because they know they cannot anticipate all possible answers. A large number of "other" responses, however, suggests the measurement scale the researcher designed did not anticipate the full range of information. The editor must determine if "other" responses appropriately fit into established categories, if new categories must be added, if "other" data will be ignored, or if some combination of these actions will be taken.

While the exhaustiveness requirement for a single variable may be obvious, a second aspect is less apparent. Does one set of categories—often determined before the data are collected—fully capture all the information in the data? For example, responses to an open-ended question about family economic prospects for the next year may originally be categorized only in terms of being "optimistic" or "pessimistic." It may also be enlightening to classify responses in terms of other concepts such as the precise focus of these expectations (income or jobs) and variations in responses between family heads and others in the family.

Mutual Exclusivity

Another important rule when adding categories or realigning categories is that category components should be mutually exclusive. This standard is met when a specific answer can be placed in one and only one cell in a category set. For example, in a survey, assume that you asked participants for their occupation. One editor's categorization scheme might include (1) professional, (2) managerial, (3) sales, (4) clerical, (5) crafts, (6) operatives, and (7) unemployed. As an editor, how would you code a participant's answer that specified "salesperson at Gap and full-time student" or maybe "elementary teacher and tax preparer"? According to census data, it is not unusual for adults in our society to have more than one job. Here, operational definitions of the occupations categorized as "professional," "managerial," and "sales" should help clarify the situation. But the

QSR, the company that provided us with N6, the latest version of NUD*IST, and N-VIVO, introduced a commercial version of the content analysis software in 2004, XSight. XSight was developed for and with the input of researchers. www.qsrinternational.com

editor facing this situation also would need to determine how the second-occupation data are handled. One option would be to add a second-occupation field to the data set; another would be to develop distinct codes for each unique multiple-occupation combination.

Single Dimension

The problem of how to handle an occupation entry like "unemployed salesperson" brings up a fourth rule of category design. The need for a category set to follow a single classificatory principle means every option in the category set is defined in terms of one concept or construct. Returning to the occupation example, the person in the study might be both a salesperson and unemployed. The "salesperson" label expresses the concept *occupation type;* the response "unemployed" is another dimension concerned with *current employment status* without regard to the respondent's normal occupation. When a category set encompasses more than one dimension, the editor may choose to split the dimensions and develop an additional data field; "occupation" now becomes two variables: "occupation type" and "employment status."

Using Content Analysis for Open Questions

Increasingly text-based responses to open-ended measurement questions are analyzed with content analysis software. **Content analysis** measures the semantic content or the *what* aspect of a message. Its breadth makes

it a flexible and wide-ranging tool that may be used as a stand-alone methodology or as a problem-specific technique. Trend-watching organizations like the BrainReserve, the Naisbitt Group, SRI International, and Inferential Focus use variations on content analysis for selected projects, often spotting changes from newspaper or magazine articles before they can be confirmed statistically. The Naisbitt Group's content analysis of 2 million local newspaper articles compiled over a 12-year period resulted in the publication of *Megatrends.*

Types of Content

Content analysis has been described as "a research technique for the objective, systematic, and quantitative description of the manifest content of a communication."[2] Because this definition is sometimes confused with simply counting obvious message aspects such as words or attributes, more recent interpretations have broadened the definition to include latent as well as manifest content, the symbolic meaning of messages, and qualitative analysis. One author states:

> In any single written message, one can count letters, words, or sentences. One can categorize phrases, describe the logical structure of expressions, ascertain associations, connotations, denotations, elocutionary forces, and one can also offer psychiatric, sociological, or political interpretations. All of these may be simultaneously valid. In short, a message may convey a multitude of contents even to a single receiver.[3]

Content analysis follows a systematic process for coding and drawing inferences from texts. It starts by determining which units of data will be analyzed. In written or verbal texts, data units are of four types: syntactical, referential, propositional, or thematic. Each unit type is the basis for coding texts into mutually exclusive categories in our search for meaning.

- *Syntactical* units can be words, phrases, sentences, or paragraphs; words are the smallest and most reliable data units to analyze. While we can certainly count these units, we are more interested in the meaning their use reveals. In content analysis we might determine the words that are most commonly used to describe product A versus its competitor, product B. We ask, "Are these descriptions for product A more likely to lead to favorable opinions and thus to preference and ultimately selection, compared to the descriptions used for product B?"

- *Referential* units are *described* by words, phrases, and sentences; they may be objects, events, persons, and so forth, to which a verbal or textual expression refers. Participants may refer to a product as a "classic," a "power performer," or "ranked first in safety"—each word or phrase may be used to describe different objects, and it is the object that the researcher codes and analyzes in relation to the phrase.

- *Propositional* units are *assertions* about an object, event, person, and so on. For example, a researcher assessing advertising for magazine subscriptions might conclude, "Subscribers who respond to offer A will save $15 over the single issue rate." It is the assertion of savings that is attached to the text of this particular ad claim.

- *Thematic* units are *topics* contained within (and across) texts; they represent higher-level abstractions inferred from the text and its context. The responses to an open-ended question about purchase behavior may reflect a temporal theme: the past ("I never purchased an alternative brand before you changed the package"), the present ("I really like the new packaging"), or the future ("I would buy the product more often if it came in more flavors"). We could also look at the comments as relating to the themes or topics of "packaging" versus a product characteristic, "flavors."

As with all other research methodologies, the analytical use of content analysis is influenced by decisions made prior to data collection. Content analysis guards against selective perception of the content, provides for the rigorous application of reliability and validity criteria, and is amenable to computerization.

What Content Is Analyzed?

Content analysis may be used to analyze written, audio, or video data from experiments, observations, surveys, and secondary data studies. The obvious data to be content-analyzed include transcripts of focus groups, transcripts of interviews, and open-ended survey responses. But researchers also use content analysis on advertisements, promotional brochures, press releases, speeches, Web pages, historical documents, and conference proceedings, as well as magazine and newspaper articles. In competitive intelligence and the marketing of political candidates content analysis is a primary methodology.

Example

Let's look at an informal application of content analysis to a problematic open question. In this example, which we are processing without the use of content analysis software, suppose employees in the sales department of a manufacturing firm are asked, "How might company-customer relations be improved?" A sample of the responses yields the following:

- We should treat the customer with more respect.
- We should stop trying to speed up the sales process when the customer has expressed objections or concerns.
- We should have software that permits real-time tracking of a customer's order.
- Our laptops are outdated. We can't work with the latest software or access information quickly when we are in the field.
- My [the sales department] manager is rude with customers when he gets calls while I'm in the field. He should be transferred or fired.
- Management should stop pressuring us to meet sales quotas when our customers have restricted their open-to-buy status.

> These categories are called *analysis frameworks* in XSight. See the screenshot on page 448.

The first step in analysis requires that the units selected or developed help answer the research question. In our example, the research question is concerned with learning who or what the sales force thinks is a source for improving company–customer relations. The first pass through the data produces a few general categories in one concept dimension: source of responsibility, shown in Exhibit 16-4. These categories are mutually exclusive. The use of "other" makes the category set exhaustive. If, however, many of the sample participants suggested the need for action by other parties—for example, the government or a trade association—then including all those responses in the "other" category would ignore much of the richness of the data. As with coding schemes for numerical responses, category choices are very important.

Since responses to this type of question often suggest specific actions, the second evaluation of the data uses propositional units. If we used only the set of categories in Exhibit 16-4, the analysis would omit a considerable amount of information. The second analysis produces categories for action planning:

> Exhibit 16-4 Open Question Coding Example (before revision)

Question: "How can company–customer relations be improved?"

Locus of Responsibility	Mentioned	Not Mentioned
A. Company		
B. Customer		
C. Joint Company-Customer		
F. Other		