Practitioner Viewpoint

If a sample is not correctly drawn, the research may produce misleading conclusions. As you will learn, careful attention to the sample plan helps researchers reach their objectives.

Linda Piekarski
Vice President, Research & Database
Survey Sampling, Inc.
In the short span of about 25 years, Survey Sampling, Inc. has grown from a tiny start-up operation to one of the largest specialty marketing research firms in the United States. As its name suggests, Survey Sampling, Inc. deals in samples, and it specializes in household as well as business samples. Currently, Survey Sampling, Inc. operates in 18 countries and it has more than 60 professionals who staff its operations, including Linda Piekarski who provided the preceding Practitioner Viewpoint. The ad lists Survey Sampling, Inc.’s major sample products. Here is a brief description of each one.

- Internet samples (e-samples)—e-mail addresses of individuals suitable for e-mail–delivered surveys. Buyers can select from any of approximately 300 different categories such as type of pet ownership, food products purchases, or sports participation.
- SSI-SNAP—downloadable software that permits users to identify population types by geography, demographic characteristics, lifestyle market segment type, sample selection method, and other parameters. The sample is delivered via e-mail attachment to the user.
- Business-to-business samples—business samples that can be identified by Standard Industrial Classification (SIC) code number. A user can use up to eight SIC code digits, meaning that the industry category can be quite precise, such as 791104 (hula instruction companies). The sample can be targeted or limited to companies with a certain number of employees or sales revenues.
- Global samples—household and business samples are available from selected countries.
- Targeted LITE samples—households that are “low incidence” or only have a small presence in the population. The database of 48 million numbers is compiled and updated based on warranty cards, online surveys, and other ways that individuals can identify their characteristics, hobbies, product usage, and so on.

Survey Sampling, Inc.’s products are used by marketing researchers to obtain samples that represent the target markets, business competitors, potential customers, or other groups that they wish to survey. Survey Sampling, Inc. does not provide custom or syndicated market research services. It has found a unique niche in the marketing research industry, and it is highly regarded by marketing research professionals as a company with high-quality samples that are delivered very quickly at a reasonable price.

International markets are measured in hundreds of millions of people, national markets comprise millions of individuals, and even local markets may constitute hundreds of thousands of households. To obtain information from every single person in a market is usually impossible and obviously impractical. For these reasons, marketing researchers make use of a sample. This chapter describes how researchers go about taking samples. We begin with definitions of basic concepts such as population, sample, and census. Then we discuss the reasons for taking samples. From here, we distinguish the four types of probability sampling methods from the four types of nonprobability sampling methods. Because online surveys are becoming popular, we discuss sampling aspects of these surveys. Last, we present a step-by-step procedure for taking a sample, regardless of the sampling method used.
BASIC CONCEPTS IN SAMPLES AND SAMPLING

To begin, we acquaint you with some basic terminology used in sampling. The terms we discuss here are population, sample, sample unit, census, sampling error, sample frame, and sample frame error.

Population

A population is defined as the entire group under study as specified by the objectives of the research project. Managers tend to have a less specific definition of the population than do researchers. This is because the researcher must use the description of the population very precisely, whereas the manager uses it in a more general way.

For instance, let us examine this difference for a research project performed for Terminix Pest Control. If Terminix were interested in determining how prospective customers were combating roaches, ants, spiders, and other insects in their homes, the Terminix manager would probably define the population as “everybody who might use our services.” However, the researcher in charge of sample design would use a definition such as “heads of households in those metropolitan areas served by Terminix who are responsible for insect pest control.” Notice that the researcher has converted “everybody” to “households” and has indicated more precisely who the respondents will be in the form of “heads of households.” The definition is also made more specific by the requirement that the household be in a metropolitan Terminix service area. Just as problem definition error can be devastating to a survey so can population definition error because a survey’s findings are applicable only to the population from which the survey sample is drawn. For example, if the Terminix population is “everybody who might use our services,” it would include industrial, institutional, and business users as well as households. If a large national chain such as Hilton Hotels or Olive Garden Restaurants were included in the survey, then the findings could not be representative of households alone.

Sample and Sample Unit

A sample is a subset of the population that should represent that entire group. Once again, there is a difference in how the manager uses this term versus how it is used by the researcher. The manager will often overlook the “should” aspect of this definition and assume that any sample is a representative sample. However, the
Basic Concepts in Samples and Sampling

A census requires information from everyone in the population. Whenever a sample is taken, the survey will reflect sampling error. A sample frame is a master list of the entire population.

As you would expect, a **sample unit** is the basic level of investigation. That is, in the Terminix example, the unit is a household. For a Weight Watchers survey, the unit would be one person, but for a survey of hospital purchases of laser surgery equipment, the sample unit would be the hospital purchasing agents because hospital purchases are being researched.

**Census**

Although a sample is a subset of a group, a **census** is defined as an accounting of the complete population. Perhaps the best example of a census is the U.S. census taken every 10 years by the U.S. Census Bureau ([www.census.gov](http://www.census.gov)). The target population in the case of the U.S. census is all households in the United States. In truth, this definition of the population constitutes an “ideal” census, for it is virtually impossible to obtain information from every single household in the United States. At best, the Census Bureau can reach only a certain percentage of households, obtaining a census that provides information within the time period of the census-taking activity. Even with a public-awareness promotional campaign budget of several hundred thousand dollars that covered all of the major advertising media forms such as television, newspaper, and radio, and an elaborate follow-up procedure method, the Census Bureau admits that its numbers are inaccurate.²

The difficulties encountered by U.S. census takers are identical to those encountered in marketing research. For example, there are instances of individuals who are in transition between residences, without places of residence, illiterate, incapacitated, illegally residing in the United States, or unwilling to participate. Marketing researchers undertaking survey research face all of these problems and a host of others. In fact, researchers long ago realized the impracticality and outright impossibility of taking a census of a population. Consequently, they turned to the use of subsets, or samples, which were chosen to represent the target population.

**Sampling Error**

**Sampling error** is any error in a survey that occurs because a sample is used. Sampling error is caused by two factors: (1) the method of sample selection and (2) the size of the sample. You will learn in this chapter that some sampling methods minimize this error, whereas others do not control it well at all. Also, in the next chapter, we show you the relationship between sample size and sampling error.

**Sample Frame and Sample Frame Error**

To select a sample, you will need a **sample frame**, which is some master list of all the sample units in the population. For instance, if a researcher had defined a population to be all shoe repair stores in the state of Montana, he or she would need a master listing of these stores as a frame from which to sample. Similarly, if the population being researched were certified public accountants (CPAs), a sample frame for this group would be needed. In the case of shoe repair stores, a list service such as American Business Lists of Omaha, Nebraska, which has compiled its list of shoe repair stores from Yellow Pages listings, might be used. For CPAs, the researcher could use the list of members of the American Institute of Certified Public Accountants, located in New York City, which contains a listing of all accountants who have passed the CPA exam. Sometimes the researcher is hampered by the lack of a physical list, and the sample frame becomes a matter of whatever access to the population the researcher can conceive of, such as “all shoppers who purchase at least $25 worth of merchandise at a Radio Shack store during the second week of
Chapter 12  Determining the Sample Plan

A listing of the population may be inaccurate and, thus, contain sample frame error.

Taking a sample is less expensive than taking a census.

March.” Here, because some shoppers pay by credit card, some by check, and some with cash, there is no physical master list of qualified shoppers, but there is a stream of shoppers that can be sampled.

A sample frame invariably contains sample frame error, which is the degree to which it fails to account for all of the population. A way to envision sample frame error is by matching the list with the population and seeing to what degree the list adequately matches the targeted population. What do you think is the sample frame in our shoe repair store sample? The primary error involved lies in using only Yellow Pages listings. Not all shops are listed in the Yellow Pages, as some have gone out of business, some have come into being since the publication of the Yellow Pages, and some may not be listed at all. The same type of error exists for CPAs, and the researcher would have to determine how current the list is that he or she is using.3

Whenever a sample is drawn, the amount of potential sample frame error should be judged by the researcher.4 Sometimes the only available sample frame contains much potential sample frame error, but it is used due to the lack of any other sample frame. It is a researcher’s responsibility to seek out a sample frame with the least amount of error at a reasonable cost. The researcher should also apprise the client of the degree of sample frame error involved.

REASONS FOR TAKING A SAMPLE

There are two general reasons a sample is almost always more desirable than a census. First, there are practical considerations such as cost and population size that make a sample more desirable than a census. Taking a census is expensive as consumer populations may number in the millions. If the population is restricted to a medium-sized metropolitan area, hundreds of thousands of individuals can be involved. Even when using a mail survey, accessing the members of a large population is cost prohibitive.

Second, typical research firms or the typical researcher cannot analyze the huge amounts of data generated by a census. Although computer statistical programs can handle thousands of observations with ease, they slow down appreciably with tens of thousands, and most are unable to accommodate hundreds of thousands of observations. In fact, even before a researcher considers the size of the computer or tabulation equipment to be used, he or she must consider the various data preparation procedures involved in just handling the questionnaires or responses and transferring these responses into computer files. The sheer physical volume places limitations on the researcher’s staff and equipment.

Approaching sufficiency of information from a different tack, we can turn to an informal cost–benefit analysis to defend the use of samples. If the project director of our Terminix household survey had chosen a sample of 500 households at a cost of $10,000 and had determined that 20 percent of those surveyed “would consider” switching to Terminix from their current pest control provider, what would be the result if a completely different sample of the same size were selected in identical fashion to determine the same characteristic? For example, suppose the second sample resulted in an estimate of 22 percent. The project would cost $10,000 more, but what has been gained with the second sample? Common sense suggests that very little in the form of additional information has been gained, for if the project director combined the two samples he or she would come up with an estimate of 21 percent. In effect, $10,000 more has been spent to gain 1 percent more of information. It is extremely doubtful that this additional precision offsets the additional cost.

You may be wondering why we are dwelling on populations, samples, and sample selection when you know that there are companies with Internet-based panels that guarantee findings representative of specific population segments. These panel
companies offer custom-made or omnibus surveys that obviate the need for a client to worry about sample selection. In truth, panel companies are prevalent, growing, and popular, but there is a “dark side” to the sample representativeness of Internet panels that has recently become known through the investigations of J. Michael Dennis, vice president and managing director at Knowledge Networks, Inc. We have provided Marketing Research Insight 12.1 so you can learn of the two dangers of using Internet-based panels as samples.

An Online Application

Is There a “Dark Side” to the Sample Representativeness of Internet Panels?

As a result of rising fears and suspicions about telemarketers, a good deal of market research has gravitated to the use of consumer panels comprised of individuals who agree to participate in surveys on an ongoing basis. Internet-based panels afford significant benefits to researchers, including (1) fast data collection, (2) minimal nonresponse, (3) stored database records on respondents, and, perhaps most important, (4) good representativeness of target populations.

Presently, Internet-based samples are evolving, and although the experience base of marketing researchers with Internet-based samples is growing, knowledge about certain dangers that may be inherent in the use of panel samples is low.

There are two conceivable sample representativeness distortions or biases that may lurk in the findings of Internet-based panels. Each one is described next.

1. “Professional” respondents. Individuals who are approached and agree to participate on Internet-based panels may feel the need to be more prepared to take surveys than is the typical respondent. For example, if a panel participant thinks a future survey may involve DVD players, the panelist may be more vigilant to advertising, pay more attention to friends’ casual comments, or even do background research on this topic. On a more subtle level, panelists may actually become aware and learn about products, services, and brands through exposure to these concepts in early surveys. They are then more attuned and knowledgeable than the typical consumer when asked about these topics in future surveys.

2. Sample selection bias. It is impossible to select an Internet-based sample that perfectly mirrors the general population because some individuals do not own computers, and there is uneven access to the Internet among owners. Even those panels whose participants can use WebTV harbor some degree of selection bias as not everyone has a television, nor does everyone own a hardwired telephone (required to connect WebTV to the Internet), and some individuals use cellular phones exclusively.

To investigate these two dangers, Knowledge Networks, Inc., a consumer panel research company that does utilize WebTV, investigated the presence of professional respondents and sample selection bias.

Professional respondents, or panel conditioning, were examined in a number of ways with Knowledge Networks 5,700 panel participants across several topics including non-alcoholic beverages, distilled beverages, attitudes toward new products, personal finances, sensitive questions, and political opinions. Systematically comparing new panelists to those who had been on the panel for various numbers of weeks up to 12 months, the findings uncovered only slight evidence of a “dark side” of respondent bias based on length of participation on the panel.

Sample selection bias was analyzed by inspecting several key sample demographics such as gender and age to see if the patterns shift over the duration of the panel’s existence. That is, with attrition of panel members, is there a systematic decrease or increase in any demographic group’s representation in the sample? In this investigation, the Knowledge Networks panel did not exhibit any worrisome evidence of sample selection bias.

Can these findings be generalized to all Internet-based panels? Knowledge Networks’ panel is designed as follows:

- **We developed the only Web-enabled panel founded on the science of probability sampling.** While most Web-based surveys are limited to interviewing users who already have Internet access and happen to participate, Knowledge Networks uses a random sample of all Americans, including non-Internet households. We provide all Panelist households with a custom-designed Web device and Internet access so that anyone with a telephone can be represented in our Panel. No training or Web experience is necessary to take part in our surveys.

Because the Knowledge Networks panel is a sample that represents this unique system of panel recruitment, replacement, incentives, and questionnaire administration, the findings of this study pertain only to its Internet-based panel.
Chapter 12  Determining the Sample Plan

**TWO BASIC SAMPLING METHODS: PROBABILITY VERSUS NONPROBABILITY**

In the final analysis, all sample designs fall into one of two categories: probability or nonprobability. **Probability samples** are ones in which members of the population have a known chance (probability) of being selected into the sample. **Nonprobability samples**, on the other hand, are instances in which the chances (probability) of selecting members from the population into the sample are unknown. Unfortunately, the terms “known” and “unknown” are misleading for, in order to calculate a precise probability, one would need to know the exact size of the population, and it is impossible to know the exact size of the population in most marketing research studies. If we were targeting, for example, readers of the magazine *People*, the exact size of the population changes from week to week as a result of new subscriptions, old ones running out, and fluctuations in counter sales as a function of whose picture is on the cover. You would be hard-pressed, in fact, to think of cases in which the population size is known and stable enough to be associated with an exact number.

The essence of a “known” probability rests in the sampling method rather than in knowing the exact size of the population. Probability sampling methods are those that ensure that, if the exact size of the population were known for the moment in time that sampling took place, the exact probability of any member of the population being selected into the sample could be calculated. In other words, this probability value is really never calculated in actuality, but we are assured by the sample method that the chances of any one population member being selected into the sample could be computed.

With nonprobability methods there is no way to determine the probability even if the population size is known because the selection technique is subjective. As one author has described the difference, nonprobability sampling uses human intervention whereas probability sampling does not. So it is the sampling method rather than the knowledge of the size of the sample or the size of the population that determines probability or nonprobability sampling.

**Probability Sampling Methods**

There are four probability sampling methods: simple random sampling, systematic sampling, cluster sampling, and stratified sampling (Table 12.1. A discussion of each method follows.

**Simple Random Sampling** With simple random sampling, the probability of being selected into the sample is “known” and equal for all members of the population. This sampling technique is expressed by the following formula:

\[
\text{Probability of selection} = \frac{\text{sample size}}{\text{population size}}
\]

So, with simple random sampling, if the researcher was surveying a population of 100,000 recent DVD player buyers with a sample size of 1,000 respondents, the probability of selection on any single population member into this sample would be 1,000 divided by 100,000, or 1 out of 100, calculated to be 1 percent.

There are a number of examples of simple random sampling, including the “blind draw” method and the table of random numbers method.

**The Blind Draw Method.** The blind draw method involves blindly choosing participants by their names or some other unique designation. For example, suppose that you wanted to determine the attitudes of students in your marketing research class...
Two Basic Sampling Methods: Probability Versus Nonprobability

Table 12.1 Four Different Probability Sampling Methods

Simple Random Sampling
The researcher uses a table of random numbers, random digit dialing, or some other random
selection procedure that guarantees each member of the population has an identical chance of
being selected into the sample.

Systematic Sampling
Using a list of the members of the population, the researcher selects a random starting point for the
first sample member. A constant “skip interval” is then used to select every other sample member.
A skip interval must be used such that the entire list is covered, regardless of the starting point.
This procedure accomplishes the same end as simple random sampling, and it is more efficient.

Cluster Sampling
The population is divided into groups called clusters, each of which must be considered to be very
similar to the others. The researcher can then randomly select a few clusters and perform a census
of each one. Alternatively, the researcher can randomly select more clusters and take samples
from each one. This method is desirable when highly similar clusters can be easily identified.

Stratified Sampling
If the population is believed to have a skewed distribution for one or more of its distinguishing
factors (e.g., income or product ownership), the researcher identifies subpopulations called
strata. A simple random sample is then taken of each stratum. Weighting procedures may be
applied to estimate population values such as the mean. This approach is better suited than other
probability sampling methods for populations that are not distributed in a bell-shaped pattern.

toward a career in marketing research. Assume that the particular class that you
have chosen as your population has 30 students enrolled. To do a blind draw, you
first write the name of every student on a 3-by-5 index card, then take all of these
cards and put them inside a container of some sort. Next, you place a top on the
container and shake it very vigorously. This procedure ensures that the names are
thoroughly mixed. You then ask some person to draw the sample. This individual is
blindfolded so that he or she cannot see inside the container. You would instruct him
or her to take out 10 cards as the sample. (For now, let us just concentrate on sam-
ple selection methods. We cover sample size determination in the next chapter.) In
this sample, every student in the class has a known and equal probability of being
selected with a probability of 10/30 or 33 percent. In other words, each student has
a 1 out of 3 chance of being selected into that sample. Of course, you could use ID
numbers or some other designation for each population member as long as there
were no duplicates.

Random Numbers Method. A more sophisticated application of simple random
sampling is to use a computer-generated number based on the concept of a table of
random numbers, which is a listing of numbers whose random order is assured.
Before computer-generated random numbers were widespread, researchers used
physical tables that had numbers with no discernible relationship to each other. If
you looked at a table of random numbers, you would not be able to see any system-
atic sequence of the numbers regardless of where on the table you begin and
whether you went up, down, left, right, or diagonally across the entries.

To use the random number method to draw the sample in your careers in mar-
keting research study, assign each student in the class a number, say 1 through 30.
Granted, we will select only 10 students, but every member of our population must
be uniquely identified before we begin the selection process. Or you might use social
Determining the Sample Plan

Using random numbers to draw a simple random sample requires a complete accounting of the population.

**MARKETING RESEARCH INSIGHT 12.2**

**Additional Insights**

**How to Use Random Numbers to Select a Simple Random Sample**

**STEP 1:** Assign all members of the population a unique number.

<table>
<thead>
<tr>
<th>Name</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams, Bob</td>
<td>1</td>
</tr>
<tr>
<td>Baker, Carol</td>
<td>2</td>
</tr>
<tr>
<td>Brown, Fred</td>
<td>3</td>
</tr>
<tr>
<td>Chester, Harold</td>
<td>4</td>
</tr>
<tr>
<td>Downs, Jane</td>
<td>5</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Zimwitz, Roland</td>
<td>30</td>
</tr>
</tbody>
</table>

**STEP 2:** Generate random numbers in the range of 1 to \( N \) (30 in this case) by using the random number function in a spreadsheet program such as Microsoft Excel. Typically, such random number functions generate numbers from 0.0 to 1.0, so if you multiply the random number by \( N \) and format the result as an integer, you will have random numbers in the range of 1 to \( N \). The following set of random numbers was generated this way.

23 12 8 4 22 17 6 23 14 2 13

Select the first random number and find the corresponding population member. In the following example, number 23 has been chosen as the starting point.

**STEP 3:** Select the person corresponding to that number into the sample.

#23—Stepford, Ann

**STEP 4:** Continue to the next random number and select that person into the sample.

#12—Fitzwilliam, Roland

**STEP 5:** Continue on in the same manner until the full sample is selected. If you encounter a number selected earlier such as the 23 that occurs at the eighth random number, simply skip over it because you have already selected that population member into the sample. (This explains why eleven numbers were drawn.)

Security numbers because these are unique to each person. If each student is given a number from 1 to 30, it is a simple matter to use a table of random numbers to draw the sample. Of course, using a computer is much more convenient.

Marketing Research Insight 12.2 shows the steps involved in using random numbers generated by a spreadsheet program to select students from this 30-member population. Beginning with the first generated random number, you would progress through the set of random numbers to select members of the population into the sample. If you encounter the same number twice within the same sample draw, the number is skipped over, because it is improper to collect information twice from the same person. Such an occurrence would constitute overrepresentation of that person in the sample and violate our “and equal” requirement.

**Advantages and Disadvantages of Simple Random Sampling.** Simple random sampling is an appealing sampling method simply because it embodies the requirements necessary to obtain a probability sample and, therefore, to derive unbiased estimates of the population’s characteristics. This sampling method guarantees that every member of the population has a known and equal chance of being selected into the sample; therefore, the resulting sample, no matter what the size, will be a valid representation of the population.

However, there are some disadvantages associated with simple random sampling. To use either the blind draw or the random numbers approach, it is necessary to predesignate each population member. In the blind draw example, each student’s
name was written on an index card, whereas in the random numbers example, each student was assigned a specific number. In essence, simple random sampling necessarily begins with a complete listing of the population, and current and complete listings are sometimes difficult to obtain. It is also very burdensome to manually provide unique designations for each population member. Numbering from 1 through an unknown total population size (N) is tedious and invites administrative errors.

**Simple Random Sampling Used in Practice.** There are three practical applications in which simple random sample designs are employed quite successfully: small populations, random digit dialing, and computerized databases.

Obviously, one of the most troublesome aspects of simple random sampling is a listing of the population. Consequently, in those marketing research studies in which “small” and stable populations are involved, it is wise to use simple random sampling. For instance, the regional distributor for Anheuser-Busch desired that a customer satisfaction survey be administered to its many accounts. The distributor wanted to know how its delivery system compared to that of its key competitors: Coors Beer and Miller Beer. The Anheuser-Busch accounts were divided into “on-premises” locations where the beer products were consumed at the location. These locations included restaurants, bars, taverns, and sports arenas. The “off-premises” locations were where buyers purchase beer and carry it away. These locations included supermarkets, convenience stores, and package stores. The beer distributor had detailed computer lists of approximately 200 on-premises clients and about 600 off-premises clients. It was a simple matter to generate two separate lists, number each account, and to use the random selection feature of Microsoft Excel to draw a sample of each type of buyer.

Other situations in which simple random sampling is commonly used rely on a computer to generate the numbers and select the sample. One instance in which simple random sampling is employed quite successfully is through the use of random digit dialing. **Random digit dialing (RDD)** is used in telephone surveys to overcome the problems of unlisted and new telephone numbers. Unlisted numbers are a growing concern not only for researchers in the United States but in all industrialized countries such as those in Europe as well.

In random digit dialing, telephone numbers are generated randomly with the aid of a computer. Telephone interviewers call these numbers and administer the survey to the respondent once the person has been qualified. However, random digit dialing may result in a large number of calls to nonexistent telephone numbers. A popular variation of random digit dialing that reduces this problem is the **plus-one dialing procedure** in which numbers are selected from a telephone directory and a digit, such as a “1,” is added to each number to determine which telephone number is then dialed.

Finally, there is the possibility of selecting respondents from computer lists, company files, or commercial listing services, which have been converted into databases. Practically every database software program has a random number selection feature, so simple random sampling is very easy to achieve if the researcher has a computerized database of the population. The database programs can work with random numbers of as many digits as are necessary, so even social security numbers with nine digits are no problem. Companies with credit files, subscription lists, or marketing information systems have the greatest opportunity to use this approach, or a research company may turn to a specialized sampling company such as Survey Sampling to have it draw a random sample of households or businesses in a certain geographic area using its extensive databases.
Without computer assistance, simple random sampling has practical implementation problems. Systematic sampling is more efficient than simple random sampling.

Researchers would not actually count every listing in a telephone directory. Rather, they estimate the population list size by determining the average listings per page.

These three cases—small populations, random digit dialing, and sampling from databases—constitute the bulk of the use of simple random sampling in marketing research. In the special situation of a large population list that is not in the form of a computer database, such as a telephone book, the time and expense required to use simple random sampling are daunting. Fortunately, there is an economical probability sampling method called systematic sampling that can be used effectively.

**Systematic Sampling** At one time, systematic sampling, which is a way to select a random sample from a directory or list that is much more efficient (uses less effort) than with simple random sampling, was the most prevalent type of sampling technique used. However, its popularity has fallen as computerized databases and generated random number features have become widely available. However, in the special case of a physical listing of the population, such as a membership directory or a telephone book, systematic sampling is often chosen over simple random sampling based primarily on the economic efficiency that it represents. In this instance, systematic sampling can be applied with less difficulty and accomplished in a shorter time period than can simple random sampling. Furthermore, in many instances, systematic sampling has the potential to create a sample that is almost identical in quality to samples created from simple random sampling.

To use systematic sampling, it is necessary to obtain a hard-copy listing of the population, just as in the case of simple random sampling. As noted earlier, the most common listing is a directory of some sort. The researcher decides on a skip interval, which is calculated by dividing the number of names on the list by the sample size. Names are selected based on this skip interval, which is computed very simply through the use of the following formula:

**Formula for skip interval**  \[ \text{Skip interval} = \frac{\text{population list size}}{\text{sample size}} \]

For example, if one were using the local telephone directory and calculated a skip interval of 250, every 250th name would be selected into the sample. The use of this skip interval formula ensures that the entire list will be covered. Marketing Research Insight 12.3 shows how to take a systematic sample.
**Additional Insights**

How to Take a Systematic Sample

**STEP 1:** Identify a listing of the population that contains an acceptable level of sample frame error.

*Example:* the telephone book for your city

**STEP 2:** Compute the skip interval by dividing the number of names on the list by the sample size.

*Example:* 25,000 names in the phone book, sample size of 500, so skip interval = every 50th name

**STEP 3:** Using random number(s), determine a starting position for sampling the list.

*Example:* *Select:* random number for page number
  *Select:* random number for the column on that page

**STEP 4:** Apply the skip interval to determine which names on the list will be in the sample.

*Example:* Jones, William P. (skip 50 names)

**STEP 5:** Treat the list as “circular.” That is, the first name on the list is now the initial name you selected, and the last name is now the name just prior to the initially selected one.

*Example:* when you come to the end of the phone book names (Zs), just continue on through the beginning (As)

---

**Why Systematic Sampling Is Efficient.** Systematic sampling is probability sampling because it employs a random starting point, which ensures there is sufficient randomness in the systematic sample to approximate a known and equal probability of any member of the population being selected into the sample. In essence, systematic sampling envisions the list as a set comprising mutually exclusive samples, each one of which is equally representative of the listed population.

How does the random starting point take place? One option would be to count or estimate the number of population members on the list, and to generate a random number between 1 and N (the population size). Then you would have to count the list until you came to that member’s location. But a more efficient approach would be to first generate a random number between 1 and the number of pages to determine the page on which you will start. Suppose page 53 is drawn. Another random number would be drawn between 1 and the number of columns on a page to decide the column on that page. Assume the 3rd column is drawn. A final random number between 1 and the number of names in a column would be used to determine the actual starting position in that column. Let us say the 17th name is selected. From that beginning point, the skip interval would be employed. The skip interval would ensure that the entire list would be covered, and the final name selected would be approximately one skip interval before the starting point. It is convenient to think of the listing as a circular file, like an old-fashioned Rolodex file, such that A actually follows Z if the list were alphabetized, and the random starting point determines where the list “begins.”

The essential difference between systematic sampling and simple random sampling is apparent in the use of the words “systematic” and “random.” The system used in systematic sampling is the skip interval, whereas the randomness in simple random sampling is determined through the use of successive random draws. Systematic sampling works its way through the entire population from beginning to end, whereas random sampling guarantees that the complete population will be covered but without a systematic pattern. The efficiency in systematic sampling is gained by two features: (1) the skip interval aspect and (2) the need to use random number(s) only at the beginning.
Determining the Sample Plan

With systematic sampling the small loss in sampling precision is counterbalanced by its economic savings.

A cluster sampling method divides the population into groups, any one of which can be considered a representative sample.

Area sampling employs either a one-step or two-step approach.

Disadvantages of Systematic Sampling. Although systematic sampling is simpler, less time consuming, and less expensive to employ than simple random sampling, it is less representative in the final analysis than simple random sampling because it arbitrarily places population members into groups before the sample is selected. Nonetheless, the small loss in sample precision is more than counterbalanced by the economic savings, so systematic sampling is often chosen when simple random sampling is impractical or too expensive. The greatest danger in the use of systematic sampling lies in the listing of the population (sample frame). Sample frame error is a major concern for telephone directories because of unlisted numbers. It is also a concern for lists that are not current. In both instances, the sample frame will not include certain population members, and these members have no chance of being selected into the sample because of this fact.

Cluster Sampling. Another form of probability sampling is known as cluster sampling, in which the population is divided into subgroups, called “clusters,” each of which represents the entire population. Note that the basic concept behind cluster sampling is very similar to the one described for systematic sampling, but the implementation differs. The procedure identifies identical clusters. Any one cluster, therefore, will be a satisfactory representation of the population. Cluster sampling is advantageous when there is no electronic database of the population. It is easy to administer, and cluster sampling goes a step further in striving to gain economic efficiency over simple random sampling by simplifying the sampling procedure used. We illustrate cluster sampling by describing a type of cluster sample known as area sampling.

Area Sampling as a Form of Cluster Sampling. In area sampling, the researcher subdivides the population to be surveyed into geographic areas, such as census tracts, cities, neighborhoods, or any other convenient and identifiable geographic designation. The researcher has two options at this point: a one-step approach or a two-step approach. In the one-step area sample approach, the researcher may believe the various geographic areas to be sufficiently identical to permit him or her to concentrate his or her attention on just one area and then generalize the results to the full population. But the researcher would need to select that one area randomly and perform a census of its members. Alternatively, he or she may employ a two-step area sample approach to the sampling process. That is, for the first step, the researcher could select a random sample of areas, and then for the second step, he or she could decide on a probability method to sample individuals within the chosen areas. The two-step area sample approach is preferable to the one-step approach because there is always the possibility that a single cluster may be less representative than the researcher believes. But the two-step method is more costly because more areas and time are involved. Marketing Research Insight 12.4 illustrates how to take an area sample using subdivisions as the clusters.

Area grid sampling is a variation of the area sampling method. To use it, the researcher imposes a grid over a map of the area to be surveyed. Each cell within the grid then becomes a cluster. The difference between area grid sampling and area sampling lies primarily in the use of a grid framework, which cuts across natural or artificial boundaries such as streets, rivers, city limits, or other separations normally used in area sampling. Geodemography has been used to describe the demographic profiles of the various clusters. Regardless of how the population is sliced up, the researcher has the option of a one-step or a two-step approach.

Disadvantage of Cluster (Area) Sampling. The greatest danger in cluster sampling is cluster specification error that occurs when the clusters are not homogeneous. For
Two Basic Sampling Methods: Probability Versus Nonprobability

Additional Insights

How to Take an Area Sampling Using Subdivisions

**STEP 1:** Determine the geographic area to be surveyed and identify its subdivisions. Each subdivision cluster should be highly similar to all others.

*Example:* 10 subdivisions within 5 miles of the proposed site for our new restaurant; assign each a number

**STEP 2:** Decide on the use of one-step or two-step cluster sampling.

*Example:* use two-step cluster sampling

**STEP 3:** (assuming two-step): Using random numbers, select the subdivisions to be sampled.

*Example:* select four subdivisions randomly, say numbers 3, 5, 2, and 9

**STEP 4:** Using some probability method of sample selection, select the members of each chosen subdivision to be included in the sample.

*Example:* identify a random starting point; instruct field-workers to drop off the survey at every fifth house (systematic sampling)

Example, if a subdivision association used area sampling to survey its members using its streets as cluster identifiers, and one street circumnavigated a small lake in the back of the subdivision, the “Lake Street” homes might be more expensive and luxurious than most of the other homes in the subdivision. If by chance, Lake Street was selected as a cluster in the survey, it would most likely bias the results toward the opinions of the relatively few wealthy subdivision residents. In the case of one-step area sampling, this bias could be severe.

**Stratified Sampling** All of the sampling methods we have described thus far implicitly assume that the population has a normal or bell-shaped distribution for its key properties. That is, there is the assumption that every potential sample unit is a fairly good representation of the population, and any who are extreme in one way are perfectly counterbalanced by opposite extreme potential sample units. Unfortunately, it is common to work with populations in marketing research that contain unique subgroupings; you might encounter a population that is not distributed symmetrically across a normal curve. With this situation, unless you make adjustments in your sample design, you will end up with a sample described as “statistically inefficient” or, in other words, inaccurate. One solution is **stratified sampling**, which separates the population into different subgroups and then samples all of these subgroups.

**Working with Skewed Populations.** A skewed population deviates quite a bit from what is assumed to be the “normal” distribution case in the use of simple random, systematic, or cluster sampling. Because of this abnormal distribution, there exists the potential for an inaccurate sample. For example, let’s take the case of a college that is attempting to assess how its students perceive the quality of its educational programs.

A researcher has formulated the question, “To what extent do you value your college degree?” The response options are along a 5-point scale where 1 equals “not valued at all” and 5 equals “very highly valued.” The population of students is defined by year: freshman, sophomore, junior, and senior. It is believed that the means will differ by the respondent’s year status because seniors probably value a degree more than do juniors who value a degree more than do sophomores, and so on. At the same time, it is expected that seniors would be more in agreement (have less variability) than would underclass-persons. This belief is due to the fact that
freshmen are students who are trying out college, some of whom are not serious about completing it and do not value it highly, but some of whom are intending to become doctors, lawyers, or professionals whose training will include graduate degree work as well as their present college work. The serious freshmen students would value a college degree highly, whereas the less serious ones would not. So, we would expect much variability in the freshmen students, less in sophomores, still less in juniors, and the least with college seniors. The situation might be something similar to the distributions illustrated in Figure 12.1.

What would happen if we used a simple random sample of equal size for each of our college groups? Because sample accuracy is determined by the variability in the population, we would be least accurate with freshmen and most accurate with seniors. To state this situation differently, we would be statistically overefficient with seniors and statistically underefficient with freshmen because we would be oversampling the seniors and undersampling the freshmen. To gain overall statistical efficiency, we should draw a larger sample of freshmen and a smaller one of seniors. We might do this by allocating the sample proportionately based on the total number of the freshmen, sophomores, juniors, and seniors, each taken as a percentage of the whole college population. (Normally, there are fewer seniors than juniors than sophomores than freshmen in a college.) Thus, we would be drawing the smallest sample from the seniors group, who have the least variability in their assessments of the value of their college education, and the largest sample from the freshmen, who have the most variability in their assessments. We discuss this more fully in the next chapter.

With stratified random sampling, one takes a skewed population and identifies the subgroups or strata contained within it. Simple random sampling, systematic sampling, or some other type of probability sampling procedure is then applied to draw a sample from each stratum. The stratum sample sizes can differ based on knowledge of the variability in each population stratum and with the aim of achieving the greatest statistical efficiency.
Two Basic Sampling Methods: Probability Versus Nonprobability

Accuracy of Stratified Sampling. How does stratified sampling result in a more accurate overall sample? Actually, there are two ways this accuracy is achieved. First, stratified sampling allows for explicit analysis of each stratum. The college degree example illustrates why a researcher would want to know about the distinguishing differences between the strata in order to assess the true picture. Each stratum represents a different response profile, and by allocating sample size based on the variability in the strata profiles, a more efficient sample design is achieved.

Second, there is a procedure that allows the estimation of the overall sample mean by use of a weighted mean, whose formula takes into consideration the sizes of the strata relative to the total population size and applies those proportions to the strata’s means. The population mean is calculated by multiplying each stratum by its proportion and summing the weighted stratum means. This formula results in an estimate that is consistent with the true distribution of the population when the sample sizes used in the strata are not proportionate to their shares of the population. Here is the formula that is used for two strata:

\[
\text{Formula for weighted mean} \\
\text{Mean}_{\text{population}} = (\text{mean}_A)(\text{proportion}_A) + (\text{mean}_B)(\text{proportion}_B)
\]

where \(A\) signifies stratum \(A\), and \(B\) signifies stratum \(B\).

Here is an example. A researcher separated a population of households that rent videos on a regular basis into two strata. Stratum \(A\) was families without young children, and stratum \(B\) was families with young children. When asked to use a scale of 1 = “poor” and 5 = “excellent” to rate their video rental store on its video selection, the means were computed to be 2.0 and 4.0, respectively, for the samples. The researcher knew from census information that families without young children accounted for 70 percent of the population, whereas families with young children accounted for the remaining 30 percent. The weighted mean rating for video selection was then computed as \((.7)(2.0) + (.3)(4.0) = 2.6\).

How to Apply Stratified Sampling. There are a number of instances in which stratified sampling is used in marketing research because skewed populations are often encountered. Prior knowledge of populations under study, augmented by research objectives sensitive to subgroupings, sometimes reveals that the population is not normally distributed. Under these circumstances, it is advantageous to
To apply stratified sampling, researchers often use an easily identified surrogate as the stratification mechanism.

Researchers should select a basis for stratification that reveals different responses across the strata.

The weighted mean formula is not needed with a proportionate sample.

apply stratified sampling to preserve the diversity of the various subgroups. Usually, a surrogate measure, which is some observable or easily determined characteristic of each population member, is used to help partition or separate the population members into their various subgroups. For example, in the instance of the college, the year classification of each student is a handy surrogate. With its internal records, the college could easily identify students in each stratum, and this determination would be the stratification method. Of course, there is the opportunity for the researcher to divide the population into as many relevant strata as necessary to capture different subpopulations. For instance, the college might want to further stratify on college of study, gender, or grade point average (GPA) ranges. Perhaps professional school students value their degrees more than do liberal arts students, women differently from men students, and high GPA students more than average GPA or failing students. The key issue is that the researcher should use some basis for dividing the population into strata that results in different responses across strata. There is no need to stratify if all strata respond alike.

If the strata sample sizes are faithful to their relative sizes in the population, you have what is called a proportionate stratified sample design. Here you do not use the weighted formula because each stratum’s weight is automatically accounted for by its sample size. But with disproportionate stratified sampling, the weighted formula needs to be used because the strata sizes do not reflect their relative proportions in the population. We have provided a step-by-step description of stratified sampling in Marketing Research Insight 12.5.

**Nonprobability Sampling Methods**

All of the sampling methods we have described thus far embody probability sampling assumptions. In each case, the probability of any unit being selected from the population into the sample is known, even though it cannot be calculated precisely.

---

**MARKETING RESEARCH INSIGHT 12.5**

**Additional Insights**

**How to Take a Stratified Sample**

**STEP 1:** Be certain that the population’s distribution for some key factor is not bell-shaped and that separate subpopulations exist.

Example: cellular telephone owners differ from nonowners in their use of long-distance calls.

**STEP 2:** Use this factor or some surrogate variable to divide the population into strata consistent with the separate subpopulations identified.

Example: use a screening question on ownership/nonownership of a cellular telephone. This may require a screening survey using random digit dialing to identify respondent pools for each stratum.

**STEP 3:** Select a probability sample from each stratum.

Example: use a computer to select simple random samples for each stratum.

**STEP 4:** Examine each stratum for managerially relevant differences.

Example: do owners differ from nonowners in their long-distance calls? Answer: owners average 35 minutes per month; nonowners average 20 minutes per month.

**STEP 5:** If stratum sample sizes are not proportionate to the stratum sizes in the population, use the weighted mean formula to estimate the population value(s).

Example: if owners are 20 percent and nonowners are 80 percent of the population, the estimate is $35(0.20) + (20)(0.80) = 23$ minutes per month.
The critical difference between probability and nonprobability sampling methods is the mechanics used in the sample design. With a nonprobability sampling method, selection is not based on probability. That is, you cannot calculate the probability of any one person in the population being selected into the sample. Still, each nonprobability sampling method claims to draw a representative sample. There are four nonprobability sampling methods: convenience samples, judgment samples, referral samples, and quota samples (Table 12.2). A discussion of each method follows.

**Convenience Samples**  
Convenience samples are samples drawn at the convenience of the interviewer. Accordingly, the most convenient areas to a researcher in terms of time and effort turn out to be high-traffic areas such as shopping malls or busy pedestrian intersections. The selection of the place and, consequently, prospective respondents is subjective rather than objective. Certain members of the population are automatically eliminated from the sampling process. For instance, there are those people who may be infrequent or even nonvisitors of the particular high-traffic area being used. On the other hand, in the absence of strict selection procedures, there are members of the population who may be omitted because of their physical appearance, general demeanor, or by the fact that they are in a group rather than alone. One author states, “Convenience samples . . . can be seriously misleading.”

Mall-intercept companies often use a convenience sampling method to recruit respondents. For example, shoppers are encountered at large shopping malls and quickly qualified with screening questions. For those satisfying the desired population characteristics, a questionnaire may be administered or a taste test performed. Alternatively, the respondent may be given a test product and asked if he or she would use it at home. A follow-up telephone call some days later solicits his or her reaction to the product’s performance. In this case, the convenience extends beyond easy access of respondents into considerations of setup for taste tests, storage of products to be distributed, and control of the interviewer workforce. Additionally, the researcher or interviewer uses a high-traffic location such as a busy pedestrian area or a shopping mall to intercept potential respondents. Error occurs in the form of members of the population who are infrequent or nonusers of that location.

**Table 12.2  Four Different Types of Nonprobability Sampling Methods**

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Convenience Sampling</strong></td>
<td>The researcher or interviewer uses a high-traffic location such as a busy pedestrian area or a shopping mall to intercept potential respondents. Error occurs in the form of members of the population who are infrequent or nonusers of that location.</td>
</tr>
<tr>
<td><strong>Judgment Sampling</strong></td>
<td>The researcher uses his or her judgment or that of some other knowledgeable person to identify who will be in the sample. Subjectivity enters in here, and certain members of the population will have a smaller chance of selection than will others.</td>
</tr>
<tr>
<td><strong>Referral Sampling</strong></td>
<td>Respondents are asked for the names or identities of others like themselves who might qualify to take part in the survey. Members of the population who are less well known, disliked, or whose opinions conflict with the respondent have a low probability of being selected.</td>
</tr>
<tr>
<td><strong>Quota Sampling</strong></td>
<td>The researcher identifies quota characteristics such as demographic or product use factors and uses these to set up quotas for each class of respondent. The sizes of the quotas are determined by the researcher’s belief about the relative size of each class of respondent in the population. Often quota sampling is used as a means of ensuring convenience samples will have the desired proportion of different respondent classes.</td>
</tr>
</tbody>
</table>
large numbers of respondents can be recruited in a matter of days. The screening questions and geographic dispersion of malls may appear to reduce the subjectivity inherent in the sample design, but in fact the vast majority of the population was not there and could not be approached to take part. Yet, there are ways of controlling convenience sample selection error using a quota system, which we discuss shortly.

**Judgment Samples**

Judgment samples are somewhat different from convenience samples in concept because they require a judgment or an “educated guess” as to who should represent the population. Often the researcher or some individual helping the researcher who has considerable knowledge about the population will choose those individuals that he or she feels constitute the sample. It should be apparent that judgment samples are highly subjective and, therefore, prone to much error.

Focus group studies often use judgment sampling rather than probability sampling. In a recent focus group concerning the need for a low-calorie, low-fat microwave oven cookbook, 12 women were selected as representative of the present and prospective market. Six of these women had owned a microwave oven for ten or more years; three of the women had owned the oven for less than ten years; and three of the women were in the market for a microwave oven. In the judgment of the researcher, these 12 women represented the population adequately for the purposes of the focus group. It must be quickly pointed out, however, that the intent of this focus group was far different from the intent of a survey. Consequently, the use of a judgment sample was considered satisfactory for this particular phase in the research process for the cookbook. The focus group findings served as the foundation for a large-scale regional survey conducted two months later that relied on a probability sampling method.

**Referral Samples**

Referral samples, sometimes called “snowball samples,” require respondents to provide the names of additional respondents. Such lists begin when the researcher compiles a short list of sample units that is smaller than the total sample he or she desires for the study. After each respondent is interviewed, he or she is queried about the names of other possible respondents. In this manner, additional respondents are referred by previous respondents. Or, as the other name implies, the sample grows just as a snowball grows when it is rolled downhill.

Referral samples are most appropriate when there is a limited and disappointingly short sample frame and when respondents can provide the names of others who would qualify for the survey. The nonprobability aspects of referral sampling come from the selectivity used throughout. The initial list may also be special in some way, and the primary means of adding people to the sample is by tapping the memories of those on the original list. Referral samples are often useful in industrial marketing research situations.

**Quota Samples**

The quota sample establishes a specific quota for various types of individuals to be interviewed. It is a form of nonprobability sampling used prevalently by marketing researchers. The quotas are determined through application of the research objectives and are defined by key characteristics used to identify the population. In the application of quota sampling, a field-worker is provided with screening criteria that will classify the potential respondent into a particular quota cell. For example, if the interviewer is assigned to obtain a sample quota of 50 each for black females, black males, white females, and white males, the qualifying characteristics would be race and gender. Assuming our field-workers were working mall intercepts, each would determine through visual inspection where the prospective respondent falls and work toward filling the quota in each of the four cells. So a quota system overcomes much of the nonrepresentativeness danger inherent in convenience samples.
Quota samples are often used by companies that have a firm grasp on the features characterizing the individuals they wish to study in a particular marketing research project. A large bank, for instance, might stipulate that the final sample be one-half adult males and one-half adult females because in the bank’s understanding of its market, the customer base is equally divided between males and females.

Sample quotas are sometimes used to facilitate field-worker control. Often there is considerable opportunity for field-workers to apply their own judgment in the selection of respondents. However, if the field-worker is given guidelines establishing quotas for various types of individuals to be included in the sample, the researcher has an assurance that the final sample composition will be faithful to his or her prior specifications. In no way does this approach eliminate the subjectivity or the nonprobability aspects of the sampling method; however, it does gain the necessary control to ensure that the final sample will include people within the marketing researcher’s definition of the population. Or it may guarantee that he or she has sufficient subsample sizes for meaningful subgroup analysis.

When done conscientiously and with a firm understanding of the quota characteristics, quota sampling can rival probability sampling in the minds of researchers. One researcher has commented, “Probability sampling is the recommended method, but in the ‘real world,’ statistical inferences are often based on quota samples and other non-random sampling methods. Strangely, these heretical uses of statistical theory, in my experience, seem to work just as well as they do for ‘purist’ random samples.”

**PICTURES SAY A THOUSAND WORDS: A GRAPHICAL PRESENTATION OF SAMPLE METHODS**

Figure 12.2 contains representations of each of the eight types of sample methods. The population in every instance is comprised of 25 consumers. One-fifth (20%) of the consumers are unsatisfied, two-fifths (40%) are satisfied, and two-fifths (40%) are indifferent or neutral about our brand. With each probability sample method
(simple random sample, systematic sample, cluster sample, and stratified random sample), the resulting sample’s satisfaction profile is consistent with the population. That is, each probability sample has five consumers, with one unsatisfied, two satisfied, and two indifferent.

However, with each of the nonprobability sample methods (convenience sample, judgment sample, referral sample, and quota sample), the resulting sample’s sat-

---

### Simple Random Sampling

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population identified uniquely by number</td>
<td>Selection by random number</td>
<td>Every member of the population has an equal chance of being selected into the sample</td>
</tr>
</tbody>
</table>

### Systematic Sampling

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directory of the population (sample frame)</td>
<td>Selection via skip interval with a random starting point</td>
<td>Every member of the sample frame has an equal chance of being selected into the sample</td>
</tr>
</tbody>
</table>

### Cluster Sampling

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population in groups (clusters)</td>
<td>Random selection of 2 clusters with random selection of members of these clusters (2-stage)</td>
<td>Every cluster (A, B, C, D, or E) in the population has an equal chance of being selected into the sample, and every cluster member has an equal chance of being selected from that cluster</td>
</tr>
</tbody>
</table>

---

Figure 12.2A  Graphical Presentations for the Various Sampling Methods
Pictures Say a Thousand Words: A Graphical Presentation of Sample Methods

**Stratified Random Sampling**

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population is separated into (e.g.) two subgroups (strata)</td>
<td>Random selection of a proportional number of stratum members from each stratum</td>
<td><img src="image1" alt="Image" /></td>
</tr>
<tr>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
<td>Every member of each stratum (for II) in the population has an equal chance of being selected into the sample (proportional sampling)</td>
</tr>
</tbody>
</table>

**Convenience Sampling**

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population</td>
<td>Selection of those who “pass by” some high traffic location</td>
<td>Only those who pass by the location have a chance of being selected into the sample resulting in error</td>
</tr>
<tr>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
</tbody>
</table>

**Judgment Sampling**

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population</td>
<td>Selection of those who are “typical” and convenient</td>
<td>Only those who are judged to be typical and convenient have a chance of being selected into the sample resulting in error</td>
</tr>
<tr>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 12.2B is satisfaction profile is not a good representation of the population. If you look at the X’s in the middle frame that depicts the sample’s selection specifics, the nonprobability sample method X’s are concentrated in specific areas of the population, whereas they are scattered across the entire population with probability samples. The concentration aspect of nonprobability sampling means that some members of the population have a disproportionate chance of being selected into the sample, resulting in sample selection error.
### Determining the Sample Plan

#### Referral Sampling

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population</td>
<td>Selection based on the referrals of respondents who are selected arbitrarily</td>
<td>Only those who are in the friendship network have a chance of being selected into the sample resulting in error</td>
</tr>
</tbody>
</table>

#### Quota Sampling

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Method</th>
<th>Resulting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>The population distribution is classified by demographics and/or some consumer behavior variable(s)</td>
<td>Selection based on a quota system that ensures the population distribution, but from a convenient location like a shopping mall</td>
<td>Only those who pass by the convenient location have a chance of being selected into the sample resulting in error</td>
</tr>
</tbody>
</table>

#### Online Sampling Techniques

Online sampling can be interpreted in the context of traditional sampling techniques.

To be sure, sampling for Internet surveys poses special challenges, but most of these issues can be addressed in the context of our probability and nonprobability sampling concepts. The trick is to understand how the online sampling method in question works and to interpret the sampling procedure correctly with respect to basic sampling concepts.

For purposes of illustration, we will describe four types of online sampling: (1) random online intercept sampling, (2) invitation online sampling, (3) online panel sampling, and (4) other online sampling types.

**Random Online Intercept Sampling**

Random online intercept sampling relies on a random selection of Web site visitors. There are a number of Java-based or other html-imbedded routines that will select Web site visitors on a random basis such as time of day or random selection from the stream of Web site visitors. If the population is defined as Web site visitors, then this is a simple random sample of these visitors within the time frame of the survey. If the sample selection program starts randomly and incorporates a skip interval system, it is a systematic sample, and if the sample program treats the population of Web site visitors like strata, it uses stratified simple random sampling as long as ran-
Online Sampling Techniques

Random selection procedures are used faithfully. However, if the population is other than Web site visitors, and the Web site is used because there are many visitors, the sample is akin to a mall-intercept sample (convenience sample).

**Invitation Online Sampling**

Invitation online sampling is when potential respondents are alerted that they may fill out a questionnaire that is hosted at a specific Web site. For example, a retail store chain may have a notice that is handed to customers with their receipts notifying them that they may go online to fill out the questionnaire. However, to avoid spam, online researchers must have an established relationship with potential respondents who expect to receive an e-mail survey. If the retail store uses a random sampling approach such as systematic sampling, a probability sample will result. Similarly, if the e-mail list is a truly representative group of the population, and the procedures embody random selection, it will constitute a probability sample. However, if in either case there is some aspect of the selection procedure that eliminates population members or otherwise overrepresents elements of the population, the sample will be a nonprobability one. A good example of a sampling system that overcomes this problem is Opinion PlaceSM. It uses a proprietary sampling method where visitors learn about Opinion PlaceSM through promotions placed throughout America Online (AOL) properties, the Internet, and various rewards programs. Additionally, AOL members can access the area quickly through a permanent placement in AOL Member Perks or through AOL Keyword: Opinion Place. Visitors to Opinion PlaceSM proceed through a complex, sophisticated screening process to ensure random representation across all surveys.

To learn more about the unique sample method used at Opinion PlaceSM and to see a comparison to online panels, read Marketing Research Insight 12.6. When you read about Opinion PlaceSM, you will find that it is operated by Digital Marketing Services, a wholly owned subsidiary of America Online. We have also provided a thumbnail sketch of Chuck Miller, executive vice president of DMS, in Marketing Research Insight 12.7.

**An Online Application**

*When Is a Panel Not a Panel? (When It’s Opinion PlaceSM!)*

Research panels are a common data collection tool for many researchers. These panels typically consist of prescreened respondents who have answered a series of profiling questions prior to participating in any actual studies and who have agreed to participate in future research studies conducted by the panel manager. The panel manager uses the information from the profiling questions to help balance the panel to represent a specific market (frequently the U.S. population) or target group (e.g., teens). They may also use the profiling questions to identify respondents who potentially meet the screening requirements for a specific project.

In the past, most of the interaction between the panelist and the manager was conducted via mail. This meant that panel managers sent out paper questionnaires to respondents, who wrote out their answers to those questions. They then mailed the completed survey back to the panel manager.

More recently, the advent of the Internet produced new opportunities for the market research industry, as it did for so many others. For researchers, this meant a new way to contact and survey respondents. Panel managers reacted by quickly moving their panels online. Instead of mailing questionnaires to respondents, now panel managers e-mail invitations to respondents with information on how to complete the survey online. Respondents complete the survey and submit it—all online. Without question, this process can be much faster than mail and more convenient for respondents.

In contrast, one research company looked at the emerging Internet and saw a way to reinvent respondent interaction, while maintaining the benefits of random selection. Rather

*Four online sampling types are (1) random, (2) invitation, (3) panel, and (4) other.*
Digital Marketing Services (DMS), a wholly owned subsidiary of America Online, envisioned an innovative method of data collection. DMS realized that the traffic on the Internet could provide a continual flow of new respondents to a specific online location (i.e., Opinion Placesm) for the sole purpose of participating in research studies. Of course, there had to be a reason for respondents to do this and, thus, DMS also created an online incentive program that rewarded every respondent for participating in research studies.

Not surprisingly, many in the industry questioned this new development since researchers are by their nature very methodical in accepting new research methods and somewhat risk averse. Many parallel tests were conducted, in which data were collected via both a traditional methodology (i.e., phone, mall, or mail panel) and online. In general, the results of these tests showed that online data collection could provide comparable results to traditional methodologies and that these results were valid and reliable.

But many researchers still found it hard to comprehend a completely new method of data collection. They tried to classify Opinion Place™ as just another online panel. In reality, it operates very differently, with many unique features and capabilities. One common analogy used to illustrate the difference describes Opinion Place™ as a free-flowing river, with fresh respondents continuously flowing to surveys. In contrast, an online panel might be compared to a pond of water; occasionally new drops arrive, but the majority of the water stays the same.

Thus, although panels are very useful tools and provide many benefits when used in the online environment, it is important to remember that Opinion Place™ incorporates capabilities and respondent opportunities distinctive from online panels. In other words, a river is not a pond.

The accompanying table details more differences between Opinion Place™ and standard online panels.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Opinion Place (OP)</th>
<th>Online Panel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Base</td>
<td>More than 80% of the online population has direct access to OP.</td>
<td>The largest online panels may have up to 5 million records.</td>
</tr>
<tr>
<td>Respondents</td>
<td>First-time respondents flow to OP from the AOL Properties (AOL, CompuServe, etc.) and the Internet (AAdvantage, Netcenter, etc.), representing a significant majority of the online population.</td>
<td>Panel members are initially recruited from individual Web sites and may represent a small slice of the online universe. Panelists return repeatedly to participate in surveys.</td>
</tr>
<tr>
<td>Frequency of Participation</td>
<td>Respondent participation is limited through the use of access controls, preventing the development of “professional respondents.”</td>
<td>Panel members may receive invitations to participate in multiple surveys within the space of a week or two.</td>
</tr>
<tr>
<td>Recruitment Process</td>
<td>Ongoing promotions alert respondents to OP, who then determine when it is convenient to participate.</td>
<td>Respondents are selected from the panel and mailed an invitation, which may state the content of the survey.</td>
</tr>
<tr>
<td>Screening Process</td>
<td>Demographic and other screening questions are asked each time a respondent enters OP. Qualified participants are then randomly assigned to a survey based on their answers.</td>
<td>Profiling data are gathered during the sign-up process. Questions are not repeated during subsequent surveys. Respondents are preidentified for participation in a specific survey or allowed to self-select which survey they complete.</td>
</tr>
<tr>
<td>Low-Incidence Targets</td>
<td>The flow of traffic through OP allows low-incidence groups to be identified at a rate relative to the degree of incidence.</td>
<td>Profiling questions enable panels to identify qualified respondents prior to mailing an e-mail invitation.</td>
</tr>
<tr>
<td>Incentive</td>
<td>Multiple incentive programs provide value to all segments of the population. Incentives are given to each respondent who completes a survey.</td>
<td>Incentives are typically in the form of sweepstakes, access to results, or charity donations and may have different levels of value to various respondent groups.</td>
</tr>
<tr>
<td>Response Rate</td>
<td>Respondents initiate contact with OP making response rates irrelevant. Among those who begin surveys, 85% or more complete the interview.</td>
<td>Current figures indicate that a significant majority of e-mail invitations to panel surveys are not accepted.</td>
</tr>
</tbody>
</table>
Chuck Miller heads the operations and technology group at Digital Marketing Services. One of the original founders of DMS, much of the current success of the company is largely due to the innovations, experience, and technical leadership of Miller. DMS was the first company to create a system that executed self-administered Web surveys capable of real-time data evaluation and complex quota controlling, moving the industry beyond e-mail and form-based surveys.

Prior to his tenure at AOL, Miller worked at the M/A/R/C Group for 10 years in a number of positions ranging from account management to the development of quality standards. While at M/A/R/C, Miller was widely recognized for his success in developing automated analysis and reporting systems.

Miller is highly regarded for his expertise in transforming more traditional research methodologies to online applications. Miller regularly shares his expertise when speaking at conferences and through his work with many industry organizations.

**Online Panel Sampling**

Online panel sampling refers to consumer or other respondent panels that are set up by marketing research companies for the explicit purpose of conducting online surveys with representative samples. There is a growing number of these companies, and online panels afford fast, convenient, and flexible access to preprofiled samples. Typically, the panel company has several thousand individuals who are representative of a large geographic area, and the market researcher can specify sample parameters such as specific geographic representation, income, education, family characteristics, and so forth. The panel company then uses its database on its panel members to broadcast an e-mail notification to those panelists who qualify according to the sample parameters specified by the market researcher. Although online panel samples are not probability samples, they are used extensively by the marketing research industry. A major exception is the probability sample panel of Knowledge Networks described in Marketing Research Insight 12.1. In some instances, the online panel company creates the questionnaire; at other times, the researcher composes the questionnaire on the panel company’s software, or some other means of questionnaire design might be used, depending on the services of the panel company. One of the greatest pluses of online panels is the high response rate, which ensures that the final sample closely represents the population targeted by the researcher.

**Other Online Sampling Approaches**

Other online sampling approaches are feasible and limited only by the creativity of the sample designers. To identify the underlying sample method, you simply need to analyze the specifics of how potential respondents are selected. For instance, a respondent may be asked to forward the survey site to his or her friends (referral sampling), or there may be a survey page that pops up after every customer makes an online purchase (census). Regardless of the approach, if you analyze it carefully in the context of basic sampling techniques that are described in this chapter, you should be able to determine if it is a probability or a nonprobability sample.

**DEVELOPING A SAMPLE PLAN**

Up to this point, we have discussed various aspects of sampling as though they were discrete and seemingly unrelated decisions. However, they are logically joined together, and there is a definite sequence of steps, called the sample plan, that the researcher goes through in order to draw and ultimately arrive at the final sample. These steps are
Chapter 12  Determining the Sample Plan

Step 1: Define the relevant population

Step 2: Obtain a “listing” of the population

Step 3: Design the sample plan (size, method)

Step 4: Access the population

Step 5: Draw the sample

Step 6: Validate the sample

Step 7: Resample, if necessary

Figure 12.3  Steps in the Sampling Process

illustrated in Figure 12.3. Now that you are acquainted with basic terms, definitions, and concepts involved with sampling, we can describe these steps in detail.

Step 1: Define the Relevant Population

As you know, the very first step to be considered in the sampling process requires a definition of the target population under study. We indicated earlier in the chapter that the target population is identified by the marketing research study objectives; however, typically at the beginning of the sampling phase of a research project, the focus on the relevant population is necessarily sharpened. This sharpening involves the translation of nebulous descriptions of the target population into fairly specific demographic or other characteristics that separate the target population from other populations. The task here is for the researcher to specify the sample unit in the form of a precise description of the type of person to be surveyed.

For example, with a fast-food restaurant survey done for Taco Bell, it might be assumed that an important descriptor helping to define the relevant population is that its members have very limited time for lunch—many of them have only 30 minutes (actually less, if one were to consider driving or walking time). A population description can result from previous studies, or it may be the collective wisdom of marketing decision makers who have catered to this particular population for a number of years and have had the opportunity to observe members’ behaviors and to listen to their comments.

Step 2: Obtain a Listing of the Population

Once the relevant population has been defined, the researcher begins searching for a suitable list to serve as the sample frame. In some studies, candidate lists are readily available in the form of databases of various sorts—company files or records, either public or private, that are made available to the researcher. In other instances, the listing is available at a price from a third party. Unfortunately, it is rare that a listing is perfectly faithful to the target population. Most lists suffer from sample frame error; or, as we noted earlier, the database does not contain a complete enumeration of members of the population. Alternatively, the listing may be a distorted accounting of the population in that some of those listed may not belong to the population.
A good example of working with sample frame error is a survey that might be conducted for the American Automobile Association (AAA) that guarantees 24-hour emergency road service for its subscribers. As a means of winning new customers, AAA might consider marketing an “automobile travelers’ maintenance and emergency kit.” This kit comes in a tough plastic case about 10 inches long, 6 inches wide, and 4 inches deep. Inside are found various car maintenance devices such as a tire pressure gauge, a quart of oil, a bottle of windshield wiper solvent, a road flare, and a waterless cleaner for your hands. Of course, the kit has the AAA roadside emergency benefits and sign-up telephone number emblazoned prominently on the top. A marketing research company is contacted and asked to come up with a proposal for a survey to determine consumer reactions to this kit at various prices. This company claims that the most appropriate list of the population would be voter registration records. What sample frame error is evident here, and how severe is it?

The answer should be apparent once you realize a list of registered voters is incomplete in that it omits new state residents. Also, it would not be a complete listing of all drivers in your state because, prior to the Motor Voter Act of 1993, not everyone who drove bothered to register to vote. Motor vehicle registration records, on the other hand, would be a more accurate listing of this population. Furthermore, most states mandate that a new resident must acquire state license plates within 60 days of acquiring an address in that state. Obviously, the motor vehicle registration list would be more complete and more current, not to mention more relevant. So why did the research company not recommend motor vehicle registrations as its sample frame? Unfortunately, vehicle registration records typically are not available for outside inspection or sampling procedures. Voter registration records, on the other hand, are available. Consequently, the use of a less accurate list is necessary because of the impracticality involved in securing a more complete listing.

The key to assessing sample frame error lies in two factors: (1) judging how different the people listed in the sample frame are from the population and (2) estimating what kinds of people in the population are not listed in the sample frame. With the first factor, screening questions at the beginning of an interview will usually suffice as a means of disqualifying those contacted who are not consistent with the population definition. As we noted in an earlier chapter, the percentage of people on a list who qualify as members of the population is referred to as the \textit{incidence rate}. For instance, a researcher may have a list of automobile owners in a metropolitan area, but if the survey seeks only drivers who have had an accident in the past year, the incidence rate might be 1 percent or one out of every 100 automobile owners. With the second consideration, if the researcher cannot find any reason that those population members who were left off the list would adversely affect the final sample, the degree of frame error is judged tolerable. In the emergency highway kit example using voter registration records, the incidence rate should be high, and unregistered state residents constitute only a tiny portion of drivers in most states, plus they probably do not differ from registered residents in their gasoline purchasing. So voter registration records would serve as an acceptable sample frame for this survey.

If the population is global and has a low-incidence rate, researchers typically turn to compiled lists, and sampling companies such as Survey Sampling, Inc. have developed services to accommodate this especially problematic aspect of global marketing research. We have written Marketing Research Insight 12.8 to describe how the Survey Sampling, Inc. Global SSI-LITE™ services work.

Now that you have learned about one of Survey Sampling, Inc.’s sampling products, you can meet two of the principals of this impressive company. We have asked Terrence Coen and Chris De Angelis to provide thumbnail descriptions of themselves, and you will find these in Marketing Research Insights 12.9 and 12.10.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{frame_error.png}
\caption{Two forms of sample frame error occur when the names of some people on the list are not part of the population, or when members of the population are not on the list.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{high_incidence.png}
\caption{Lists with high-incidence rates are good candidates for use as sample frames.}
\end{figure}
A major challenge for online sampling is to find an adequate sample frame. This challenge is multiplied many times when the population involves global marketing or, at the least, a population of consumers in a single foreign country.

Survey Sampling, Inc. (www.ssisamples.com) has been a leading sample provider in the United States for the past two decades, and in the late 1990s, it greatly expanded its global samples capabilities. At the same time, it launched a special online survey sampling services called Global SSI-LITe™.

Survey Sampling, Inc. compiles master lists with e-mail addresses of Internet users who have agreed to answer online surveys on specific topics such as financial matters and money management, computer game usage, children’s education, health, child-rearing practices, and toys. In addition, these individuals agree to answer e-mail surveys on general topics such as pets, autos, computers, entertainment, health, music, sports, and travel.

Presently, the Global SSI-LITe™ service exists for consumers in virtually every European country, and it is rapidly growing to represent all countries where Internet usage is commonplace. This sampling system is especially valuable for researchers seeking to sample in low-incidence situations, that is, seeking to find consumers who are difficult to find in the general population, such as Siamese cat owners or vinyl record collectors.

How does it work? A researcher designs and posts an online questionnaire. Then he or she contacts Survey Sampling, Inc. to communicate what countries, what types of respondents, and how many are to be drawn into the sample. Survey Sampling, Inc. then uses its extensive databases that comprise its Global SSI-LITe™ system to send out e-mail invitations to the target populations, telling these individuals where the researcher’s online questionnaire is posted.

Meet a Marketing Researcher

Terrence Coen of Survey Sampling, Inc.

Vice President Sales and Marketing, Survey Sampling, Inc.

Terry Coen has been vice president of sales and marketing for Survey Sampling since 1988.

From 1974 until that time he held a variety of executive positions with organizations responsible for processing large establishment–oriented databases. From the mid-1970s to the mid-1980s he was vice president, sales and marketing, for Market Data Retrieval. As a principal of the company he was instrumental in bringing a start-up company to sales of $25 million a year.

In the mid-1980s Coen served as president of National Business Lists and in the late 1980s was vice president for alternate channels of distribution for Dun & Bradstreet.
Step 3: Design the Sample Plan (Size and Method)

Armed with a precise definition of the population and an understanding of the availability and condition of lists of the target population, the researcher progresses directly into the design of the sample itself. At this point, the costs of various data collection method factors come into play. That is, the researcher begins to simultaneously balance sample design, data collection costs, and sample size. We discuss sample size determination in the next chapter, and you will learn that it is a trade-off between the desire for statistical precision and the requirements of efficiency and economy.

Regardless of the size of the sample, the specific sampling method or combination of sampling methods to be employed must be stipulated in detail by the researcher. There is no one “best” sampling method. The sample plan varies according to the objectives of the survey and its constraints.26

The sampling method description includes all of the necessary steps to draw the sample. For instance, if we decided to use systematic sampling, the sampling method would detail the sample frame, the sample size, the skip interval, how the random starting point would be determined, qualifying questions, recontacts, and replacement procedures. That is, all eventualities and contingencies should be foreseen and provisions should be made for each of them. These contingency plans are most apparent in the directions given to interviewers or provided to the data collection company. Obviously, it is vital to the success of the survey that the sampling method be adhered to throughout the entire sampling process.
Chapter 12  Determining the Sample Plan

Step 4: Draw the Sample

Drawing the sample is a two-phase process. First, the sample unit must be selected. Second, information must be gained from that unit. Simply put, you need to choose a person and ask him or her some questions. However, as you realize, not everyone will agree to answer. So there comes the question of substitutions.  

Substitutions occur whenever an individual who was qualified to be in the sample proves to be unavailable, unwilling to respond, or unsuitable. The question here is, “How is the substitution respondent determined?” If the marketing research project director wishes to ensure that a particular sampling method is used faithfully, the question of substitutions must be addressed. There are three substitution methods in practice: drop-downs, oversampling, and resampling.

The drop-down substitution is often used with systematic sampling. Let us say that we are using a telephone directory as our sample frame, and you are the interviewer who is instructed to call every 100th name. On your first call, the person qualifies but refuses to take part in the survey. If the drop-down method of substitution is in effect, your responsibility is to call the name immediately following the one you just called. You will not skip 100 names but just drop down to the next one below the refusal. If that person refuses to take part, you will drop down another name and so on until you find a cooperative respondent. Then you will resume the 100 skip interval, using the original name as your beginning point. Obviously, interviewers must be provided the complete sample frame to use drop-down substitution.

Oversampling is an alternative substitution method, and it takes place as a result of the researcher’s knowledge of incidence rates, nonresponse rates, and unusable responses. For example, if the typical response rate for a mail survey questionnaire hovers around 20 percent, in order to obtain a final sample of 200 respondents, 1,000 potential respondents must be drawn into the mailout sample. Each data collection method constitutes separate oversampling implications, and it is up to the marketing research project director to apply his or her wisdom to determine the appropriate degree of oversampling. Otherwise, resampling will be necessary at a later point in the marketing research study in order to obtain the desired sample size.

Resampling constitutes a third means of respondent substitution. Resampling is a procedure in which the sample frame is tapped for additional names after the initial sample is drawn. Here the response rate may turn out to be much lower than anticipated, and more prospective respondents must be drawn. CATI random digit dialing uses resampling implicitly because the numbers are generated until the desired sample size is reached. Of course, provision must be made that prospective respondents appearing in the original sample are not included in the resample.

Step 5: Assess the Sample

The final activity in the sampling process is the assessment stage. Sample assessment can take a number of forms, one of which is to compare the sample’s demographic profile with a known profile such as the census. With quota sample validation, of course, the researcher must use a demographic characteristic other than those used to set up the quota system. The essence of sample validation is to assure the client that the sample is, in fact, a representative sample of the population about which the decision maker wishes to make decisions. Although not all researchers perform sample validation, it is recommended when prior knowledge exists about the population’s demographic profile. When no such prior information exits, validation is not possible, and the sample selection method bears the full burden of convincing clients that the sample is representative of the population.
Step 6: Resample If Necessary

When a sample fails assessment, it means that it does not adequately represent the population. This problem may arise even when sample substitutions are incorporated. Sometimes when this condition is found, the researcher can use a weighting scheme in the tabulations and analyses to compensate for the misrepresentation. On the other hand, it is sometimes possible to perform resampling by selecting more respondents and adding them to the sample until a satisfactory level of validation is reached.

Here is how our Blue Ribbon Panel members answered our question about how online research is wrestling with issues of representativeness of online samples.

Q: One of the issues in online survey research has been the questioning of online survey respondents’ representativeness. The charges have been made that Internet users cannot be representative of the entire population and that people who respond to online surveys are not even representative of all Internet users. Now, after a few years of experience with online surveys and evaluating the results of those surveys, how would you respond to the question of representativeness of online survey respondents?

Note: To learn about the panel members, see the Preface of this book.

A: I am in total agreement with the statement regarding the problems of representativeness of online samples being used to model the entire population for many types of study. At the very least, a hybrid methodology should be used in which at least 60 percent of the population is not online and/or in which reasonable bias can be hypothesized. For online populations, however, I have no reason to find that nonresponse makes projecting to that population any more difficult than nonresponse in traditional methods makes projecting to off-line populations. I have not encountered any online research findings that are not as good, if not better, at capturing a representative demographic sample of the online population being studied.

A: One of the paradoxes of online research is that it can be a poor method of capturing a representative set of opinions about topics relating to the Internet. People who are able and willing to take online surveys will tend to be heavier users of the Internet than all people with online access. They will also tend to engage in a wider variety of online activities.

Obviously, opinions of people who complete online surveys will also not be representative of those of the entire population on many topics. By the same token, those who are willing to take telephone surveys will not be representative of the entire population on some issues, nor will those who make the effort to complete mail questionnaires. Whether findings from a particular survey method and sample source will be less representative of the views of the target population depends very much on the specifics of the study.

Before an organization dives wholeheartedly into online research, it is wise to use parallel studies to see how both sample population and methodological differences will affect results.

A: DMS has conducted over 300 online comparability studies side by side with traditional methods that claim to have representative sampling. In each of these we found our online methods yielded similar results and, more importantly, the business decisions would be similar. These findings are not
unique to DMS. In January 1999 the Advertising Research Foundation (ARF) held a conference about online validity in which a dozen online companies presented overwhelming evidence of online comparability and validity.

Although many population subsets are underrepresented online, quota sampling can be employed to build appropriate samples. It should be noted that only a small portion of market research studies is designed to represent the total United States. Typically marketing studies are designed to screen for category users with specific ages and gender quotas rather than the general population.

The Internet population is forecasted to grow by 65 percent between 2000 and 2003 and is predicted to reach 1 billion by the year 2005 (source: PriceWaterhouse Coopers). At this point, the Web is largely comprised of users with higher annual incomes as compared with the general U.S. population, but the fastest-growing segment of the Internet user population is households with incomes of less than $25,000 a year (source: MediaMetrix, The Dollar Divide, August 2000). Although Internet usage levels do not match the U.S. census data and there is no true sampling frame, weighting techniques can be applied.

The demographics of the online population continue to shift as Internet penetration grows. Currently, more than two-thirds of all U.S. households are online. The online population is rapidly converging to that of the general U.S. population—increasingly enabling researchers to use online methodologies for many applications and decreasing representativeness concerns.

Although online research may not currently be suitable for all research needs, certainly every marketer has one or more segments of its customers, markets, or target markets that are well represented online. It is simply a matter of time until Internet penetration drives convergence and representativeness for all market segments.

There is no question that consumers volunteering or opting to complete online surveys do not accurately reflect the population of the United States (or even the online population). The biases are weakly correlated to demographics or other characteristics that might enable developing adjustment factors. The lack of a workable sampling frame requires setting up panels selected according to probability theory. Although such panels bring with them issues around conditioning of panelists and attrition, these are much more easily solved.

**SUMMARY**

This chapter described various sampling methods. It began by acquainting you with various terms such as population, census, and sample frame. A sample is taken because it is too costly to perform a census, and there is sufficient information in a sample to allow it to represent the population. We described four probability sampling methods in which there is a known chance of a member of the population being selected into the sample: (1) simple random sampling, (2) systematic sampling, (3) cluster sampling using area sampling as an example, and (4) stratified sampling. We also described four nonprobability sampling methods: (1) convenience sampling, (2) judgment sampling, (3) referral sampling, and (4) quota sampling. The growing trend to draw online samples was noted, and we related (1) random online intercept sampling, (2) invitation online sampling, and (3) other online sampling types. Finally, we described seven steps needed to develop a sample plan: (1) define the relevant population; (2) obtain a listing of the population; (3) design the sample plan (size and methods); (4) access the population; (5) draw the sample; (6) validate the sample; and (7) resample if necessary.
KEY TERMS
Population (p. 334)  
Sample (p. 334)  
Sample unit (p. 335)  
Census (p. 335)  
Sampling error (p. 335)  
Sample frame (p. 335)  
Sample frame error (p. 336)  
Probability samples (p. 338)  
Nonprobability samples (p. 338)  
Simple random sampling (p. 338)  
Blind draw method (p. 338)  
Table of random numbers (p. 339)  
Random digit dialing (RDD) (p. 341)  
Plus-one dialing procedure (p. 341)  
Systematic sampling (p. 342)  
Skip interval (p. 342)  
Cluster sampling (p. 344)  
Area sampling (p. 344)  
One-step area sample (p. 344)  
Two-step area sample (p. 344)  
Stratified sampling (p. 345)  
Skewed population (p. 345)  
Strata (p. 346)  
Weighted mean (p. 347)  
Surrogate measure (p. 348)  
Proportionate stratified sample (p. 348)  
Disproportionate stratified sampling (p. 348)  
Convenience samples (p. 349)  
Judgment samples (p. 350)  
Referral samples (p. 350)  
Quota sample (p. 350)  
Random online intercept sampling (p. 354)  
Invitation online sampling (p. 355)  
Online panel sampling (p. 357)  
Sample plan (p. 357)  
Incidence rate (p. 359)  
Drop-down substitution (p. 362)  
Oversampling (p. 362)  
Resampling (p. 362)  
Sample validation (p. 362)

REVIEW QUESTIONS/APPLICATIONS
1. Distinguish a nonprobability from a probability sampling method. Which one is the preferable method and why? Indicate the pros and cons associated with probability and nonprobability sampling methods.
2. List and describe briefly each of the probability sampling methods described in the chapter.
3. What is meant by the term “random”? Explain how each of the following embodies randomness: (a) table of random numbers, (b) blind draw, (c) use of random digit dialing, and (d) use of a computer.
4. In what ways is a systematic sample more efficient than a simple random sample? In what way is systematic sampling less representative of the population than simple random sampling?
5. Distinguish cluster sampling from simple random sampling. How are systematic sampling and cluster sampling related?
6. Differentiate one-step from two-step area sampling, and indicate when each one is preferred.
7. What is meant by a “skewed” population? Illustrate what you think is a skewed population distribution variable and what it looks like.
8. What are the popular online sampling methods? Describe each one.
9. Briefly describe each of the four nonprobability sampling methods.
10. Why is quota sampling often used with a convenience sampling method such as mall intercepts?
11. Describe each of the three methods of substitution for individuals who are selected into the sample but refuse to participate in the survey or who did not qualify.
12. Provide the marketing researcher’s definitions for each of the following populations:
Chapter 12  Determining the Sample Plan

a. Columbia House, a mail-order house specializing in tapes and compact discs, wants to determine interest in a 10-for-1 offer on jazz CDs.

b. The manager of your student union is interested in determining if students desire a “universal” debit ID card that will be accepted anywhere on campus and in many stores off campus.

c. Joy Manufacturing Company decides to conduct a survey to determine the sales potential of a new type of air compressor used by construction companies.

13. Here are four populations and a potential sample frame for each one. With each pair, identify (1) members of the population who are not in the sample frame and (2) sample frame items that are not part of the population. Also, for each one, would you judge the amount of sample frame error to be acceptable or unacceptable?

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Buyers of Scope mouthwash</td>
<td>Mailing list of Consumer Reports subscribers</td>
</tr>
<tr>
<td>b. Listeners of a particular FM radio classical music station in your city</td>
<td>Telephone directory</td>
</tr>
<tr>
<td>c. Prospective buyers of a new day planner and prospective clients tracking kit</td>
<td>Members of Sales and Marketing Executives</td>
</tr>
<tr>
<td>d. Users of weatherproof decking materials (to build outdoor decks)</td>
<td>Individuals' names registered at a recent home and garden show</td>
</tr>
</tbody>
</table>

14. Taco Bell approaches an official at your university and proposes to locate one of its restaurants on the campus. Because it would be the first commercial interest of this sort on your campus, the administration requires Taco Bell to conduct a survey of full-time students to assess the desirability of this operation. Analyze the practical difficulties encountered with doing a census in this situation, and provide specific examples of each one. For instance, how long might such a census take, how much might it cost, what types of students are more accessible than others, and what capacity constraints might be operating here?

15. A state lottery (weekly lottery in which players pick numbers from 1 to 20) player is curious about the randomness of winning lottery numbers. He has kept track of the winning numbers in the past five weeks and finds that most numbers were selected 25 percent of the time, but the number 6 was one of the winning numbers 50 percent of the time. Will he be more or less likely to win if he picks a 6 in this week’s lottery, or will it not make any difference in his chances? Relate your answer to simple random sampling.

16. Pet Insurers Company markets health and death benefits insurance to pet owners. It specializes in coverage for pedigreed dogs, cats, or expensive and exotic pets such as miniature Vietnamese pot-bellied pigs. The veterinary care costs of these pets can be high, and their deaths represent substantial financial loss to their owners. A researcher working for Pet Insurers finds that a listing company can provide a list of 15,000 names, which includes all current subscribers to Cat Lovers, Pedigreed Dog, and Exotic Pets Monthly. If the final sample size is to be 1,000, what should be the skip interval in a systematic sample for each of the following: (a) a telephone survey using drop-down substitution, (b) a mail survey with an anticipated 30 percent response rate, and (c) resampling to select 250 more prospective respondents? Also, assess the incidence rate for this sample frame.
17. A market researcher is proposing a survey for the Big Tree Country Club, a private country club that is contemplating several changes in its layout to make the golf course more championship caliber. The researcher is considering three different sample designs as a way to draw a representative sample of the club’s golfers. The three alternative designs are:

a. Station an interviewer at the first hole tee on one day chosen at random, with instructions to ask every 10th golfer to fill out a self-administered questionnaire.

b. Put a stack of questionnaires on the counter where golfers check in and pay for their golf carts. There would be a sign above the questionnaires, and there would be an incentive for a “free dinner in the clubhouse” for three players who fill out the questionnaires and whose names are selected by a lottery.

c. Using the city telephone directory, a plus-one dialing procedure would be used. With this procedure a random page in the directory would be selected, and a name on that page would be selected, both using a table of random numbers. The plus-one system would be applied to that name and every name listed after it until 1,000 golfers are identified and interviewed by telephone.

Assess the representativeness and other issues associated with this sample problem. Be sure to identify the sample method being contemplated in each case. Which sample method do you recommend using and why?

18. A financial services company wants to take a survey of Internet users to see if they are interested in using the company's financial-planning and asset-tracking Internet services. Previous studies have shown that Internet usage differs greatly by age, education, and gender. The total sample size will be 1,000. Using actual information about Internet usage or supplying reasonable assumptions, indicate how stratified simple random sampling should be applied for each of the following stratification situations: (1) 18–25, 26–50, and 51–65 age ranges; (2) education levels of high school diploma, some college, college degree, and graduate degree; and (3) gender: males versus females.

INTERACTIVE LEARNING

Visit the Web site at www.prenhall.com/burnsbush. For this chapter, work through the Self-Study Quizzes, and get instant feedback on whether you need additional studying. On the Web site, you can review the chapter outlines and case information for Chapter 12.

CASE 12.1 BYU's Creamery on North

In the fall of 2001, the Brigham Young University (BYU) Foodservice Department launched a bold extension of its traditional food services by opening the “Creamery on North.” The facility is a grocery-and-drugstore located on BYU’s campus and catering specifically to BYU students, faculty, and staff, but primarily to students. The university-run supermarket was described in the following manner by the director of dining services at the grand opening: “It will be a full-line grocery with fresh meats, produce, an in-house bakery, and an expanded line of canned and paper goods.”

Almost 30,000 students attend BYU, and although most do not live on campus, the Creamery on North is located on the eastern edge of the BYU campus close to a large university apartment complex and a residential area where many off-campus students live. Because its closest competitor is approximately 4 miles away, it is very convenient for most BYU students.
The college crowd is one of the most difficult food preference profiles to figure out because college students are in a volatile period when they are transitioning from their parent-controlled, high school–determined, and/or juvenile peer–influenced meals and food fads to entirely self-decided meals. But they are busy with studies, social activities, or work, and most are not experienced in making home-cooked meals. The “freshman 15” myth predicts that first-year college students will gain 15 pounds due to their diets and appetites being uncontrolled. In short, college students’ eating habits and the foods they like to eat are in no way mainstream.

So the BYU Foodservice Department has a formidable challenge in its need to monitor and predict BYU student food choices in order to stock the items that are going to be “hot.” At the same time, it is tasked with the general well-being of BYU students who buy food or eat at any of the several Foodservices facilities on campus.

In an attempt to measure the grocery store purchasing preference of its target market, the director of the Foodservice Department contacted the chairman of the BYU Marketing Department, who agrees to have a student team perform a survey. The team meets with Creamery on North managers, brainstorms, conducts a focus group, and comes up with the following objectives.

- To determine the typical student’s perception of the Creamery on North’s food assortment.
- To ascertain the typical BYU student’s preferences for various grocery items.
- To identify what acceptable “healthy” items should be added to the grocery store’s shelves.

The student team has decided to design a self-administered questionnaire to collect its data in this survey, and it is now debating the sample plan. Somehow the plan must allow for the research team members to access all BYU students who might use the Creamery on North for grocery purchases, hand the questionnaire to each potential respondent, and retrieve the questionnaire once it is completed.

1. Specify a sample plan for each of the following types of probability sampling methods: simple random sampling, systematic sampling, cluster sampling, and stratified sampling.
2. Specify a sample plan for each of the following types of nonprobability sampling methods: convenience sampling, judgment sampling, and quota sampling.

Don’t forget to include all of the necessary mechanics of each plan.

**CASE 12.2 The Cobalt Group: Online Survey for MyCarTools™**

The Cobalt Group ([www.cobaltgroup.com](http://www.cobaltgroup.com)) touts itself as a leading provider of e-business products and services for the automotive industry. It claims almost 9,000 Web site clients and about the same number of parts locator clients. There are several Cobalt Group products, and a number are geared toward the automobile dealer–automobile buyer/owner relationship. One of Cobalt Group’s products is MyCarTools™, which provides auto dealers with the ability to give personal Web space to its auto buyers. This Web space is set up to let the car owner input his or her auto usage, to track service records, or to schedule a service appointment online. The auto dealership can use this Web space service as a direct-marketing tool of its own, for as its customers enter their information on their Web spaces, the dealership’s database of owners grows. This database can be used to identify customer types and trends, to target specific customer segments with advertisements and special promotions, or even to alert customers to product recalls.

There is great potential value in MyCarTools™, for it is a means of maintaining close contact with buyers of autos from the dealership over the life of their automobiles. If satisfying to the customer, this contact may well translate into a strong propensity in that customer to return to the dealership when he or she is thinking about replacing the auto. Of course, the degree of
participation in MyCarTools™ is entirely voluntary, and it depends on a number of factors. First, the auto buyer must agree to be entered into the dealer’s Web space, and, second, the buyer must have computer availability and Internet connectivity. Finally, the buyer must use the service on a regular basis.

In an online chat problem-solving forum, automobile dealers who use Cobalt Group dealership software brainstormed questions about the value of MyCarTools™. The top question was, “Does the use of MyCarTools™ increase the propensity of a car buyer to return to the dealership where he or she bought the auto as compared to those who do not use MyCarTools™?”

The Cobalt Group Special Projects Team was assigned the task of researching this question. The team immediately created an online questionnaire using Websurveyor, but it is having difficulty deciding on the sample plan.

1. What should be the population definition for this marketing research situation?
2. Given this population definition, what should be the sample frame?
3. Should the Special Projects Team use an e-mail invitation or a Web site pop-up invitation sampling method? Why did you recommend this method over the other option?

CASE 12.3

Your Integrated Case

Hobbit’s Choice Restaurant

This is your integrated case described on pages 42–43.

After some deliberation, Cory Rogers has narrowed the data collection method for The Hobbit’s Choice Restaurant down to two choices: to use either a telephone sample or an online panel. In order for his forecasting model to work properly, Cory needs to use a sample design that will result in a sample that represents the entire greater metropolitan area.

1. Should Cory use a systematic sample using the metropolitan area telephone book as the sample frame? What are the advantages and disadvantages of this sample plan?
2. Should Cory use random digit dialing for the sample plan? What are the advantages and disadvantages of this sample plan?
3. Should Cory use a probability online panel such as the one maintained by Knowledge Networks, Inc? With respect to sample design, what are the advantages and disadvantages involved with using this approach? (You may want to review Knowledge Networks, Inc. services by visiting its Web site at www.knowledgenetworks.com.)