**CHAPTER 9**

**SELECTING THE SAMPLE**

**LEARNING OBJECTIVES**

In this chapter you will learn:

9-1 Basic concepts involved with samples and sampling

9-2 The reasons for taking a sample

9-3 Differences between probablity and nonprobability sampling

9-4 How to perform each of four different types of probability sampling

9-5 How to perform each of four different types of nonprobability sampling

9-6 About online sampling techniques

9-7 The steps involved with developing a sampling plan

**CHAPTER OUTLINE**

**Basic Concepts in Samples and Sampling**

* **Population**
* **Census**
* **Sample and Sample Unit**
* **Sample Frame and Sample Frame Error**
* **Sampling Error**

**Reasons for Taking a Sample**

**Probability Versus Nonprobability Sampling Methods**

* **Probability Sampling Methods**
	+ Simple random sampling
		- The random device method
		- The random numbers method
		- Advantages and disadvantages of simple random sampling
		- Simple random sampling used in practice
	+ Systematic sampling
		- Why systematic sampling is “fair”
		- Disadvantage of systematic sampling
* Cluster sampling
	+ Area sampling as a form of cluster sampling
	+ Disadvantage of cluster (area) sampling
* Stratified sampling
	+ Working with skewed populations
	+ Accuracy of stratified sampling
	+ How to apply stratified sampling
* **Nonprobability Sampling Methods**
	+ Convenience samples
	+ Purposive samples
	+ Chain referral samples
	+ Quota samples

**Online Sampling Techniques**

* **Online Panel Samples**
* **River Samples**
* **Email List Samples**

**Developing A Sample Plan**

**KEY TERMS**

Population Census

Sample Sample unit

Sampling error Sample frame

Sample frame error Probability samples

Nonprobability samples Simple random sampling

Random device method Blind draw method

Random numbers Random digit dialing

Plus-one dialing procedure

Systematic sampling Skip interval

Cluster sampling Area sampling

One-step area sample Two-step area sample

Stratified sampling Skewed population

Strata Weighted mean

Surrogate measure Proportionate stratified sample

Statistical Efficiency Disproportionate stratified sampling

Convenience samples Purposive samples

Chain referral samples Quota sample

Online panel sample River sample

Email list samples Sample plan

**TEACHING SUGGESTIONS**

1. The equiprobable aspects of a blind draw random sample can be demonstrated a number of different ways. Here are two examples.
* Have students research weekly lottery numbers and determine the percent of times each number appears.
* Use 3x5 cards with students’ names in a hat or a box and have a series of actual blind draw samples. Replace the drawn names to the population pool after each sample. Maintain a record of how often each student’s name is selected.
1. The greater efficiency of systematic sampling over simple random sampling can be demonstrated with a class exercise. Identify two groups of students and give each a copy of the same page from the telephone book. Tell the first group to select 10 household names using a table of random numbers or random numbers generated via a spreadsheet program such as Microsoft Excel (simple random sampling), and tell the second group to select 10 names using systematic sampling. The second group should finish before the first group.
2. When learning about the skip interval used in systematic sampling, students sometimes ask how to determine the population size when a directory or phone book is used. In the absence of a precise number, the size is usually estimated by multiplying the approximate number of names on each page by the number of pages. Some adjustments may need to be made for multiple listings such as children’s phone, discontinued phones, or areas of “white space” in the directory.
3. A disadvantage of systematic sampling noted in the chapter is hidden patterns or “periodicities.” These are extremely rare in most lists used in marketing research. Such lists are typically listed in alphabetic order, which effectively guards against periodicities. An example of a periodicity is in quality control sampling where a machine has a defect that follows a pattern. For example, every tenth unit may contain the defect, and there is a danger that the skip pattern will miss the defects or greatly understate their incidence.
4. Some students may have difficulty understanding the weighted mean calculations in systematic sampling. It may be necessary to illustrate how the mean changes with different stratum configurations. Here are some comparisons than can be used to demonstrate the effects.

Stratum Mean Estimated Population Mean

A B 50/50 40/60 10/90

5 8 6.5 6.8 7.7

1. Students should come to realize that the success of quota sampling is greatly dependent on a priori knowledge of the population’s characteristics. One way to facilitate this understanding is to ask students what quota characteristics should be used in the following two cases.
	* Case one. Kellogg’s wants to know the reactions of parents to a new children’s cereal called “Cheery-O’s”
	* Case two. Proctor and Gamble wants the reactions of potential buyers to its new hair rinse called “Gentle Care.”

With case one, the quota characteristics would be: (1) parents (percent female versus male), (2) marital status (percent married versus separated), and (3) age of youngest child (percent 4, 5, 6, etc.). With case two, however, the target market is not identified well other than it is implicitly made up of women.

1. Students tend to recall little about tables of random numbers, and a worthwhile class exercise is to bring in a table or have one made into a PowerPoint slide. Use the table to show how a simple random sample would be selected as well as how the starting page in a directory (such as the telephone directory) would be selected by use of the table of random numbers. (Note: We have opted to include no statistical tables, so you will need to turn to a statistics textbook for a table of random numbers.)
2. A different random number example is to use Excel or a spreadsheet program and program random numbers into it, say in a block of 10 rows by 10 columns. Multiplying the decimal random number by 100 and rounding it will create random numbers between 0 and 99. In theory, the average of any 10 random numbers (any row or any column) should be approximately equal to the average of any other ten random numbers. The standard deviations should be approximately equal as well.
3. When students work with a familiar population, they are better able to apply sample methods. Ask how the full-time students in your university would be sampled using each sample method described in the chapter. For example, where would they station interviewers for a convenience sample? How would they set up clusters or strata using student characteristics?
4. Here is a table the summarizes the differences between probability sampling methods and nonprobability sampling methods:

|  |  |
| --- | --- |
| **Probability Sampling** | **Nonprobability Sampling** |
| Known chance of selection | Unknown chance of selection |
| Takes more time | Takes less time |
| Higher cost | Lower cost |
| Can compute sample error | Cannot compute sample error |

**ACTIVE LEARNING EXERCISES**

**Are Random Numbers Really Random?**

This exercise demonstrates the generation of random numbers with a computer (Microsoft Excel). If students follow the instructions correctly, they will find that each number from 1 to 100 has one chance out of 100 of being selected, do the random numbers are random.

**Take a Systematic Sample Using Your Telephone Book**

This exercise takes students step-by-step thorough how to select a systematic sample from a telephone book.

*Assume that you expect a 50% response rate. What adjustment to in the skip interval calculation can you make to accommodate the fact every other prospective respondent will refuse to take part in the survey when asked?*

To adjust for a 50 percent response rate, the sample size must be doubled.

**Assess the Representativeness of Various Convenience Samples**

*Suppose the Athletic Department at your university is disappointed at student attendance of its “minor” collegiate sports events such as wrestling, cross country, and softball. It wants to learn why students do not attend them. Listed below are possible locations for a convenience sample. With each one, indicate what types of students would be overrepresented in the sample and what types would be underrepresented in the sample versus the population of students at your university for each case.*

|  |  |  |
| --- | --- | --- |
| **Convenience sample location** | **What students would be overrepresented?** | **What students would be underrepresented?** |
| The University Recreation Center | Students who work out, are active, or into exercise and health | Sedentary students, commuter students, students who study too hard to exercise or recreate |
| The University Commons  | Students who hang out there; those who buy food and snacks; those who participate in the activities at the Commons | Commuters, nontraditional students (such as nighttimers), students who do not buy food there or participate in the activities |
| The Library | Students who use the Library (typically underclasspersons and graduate students) | Students who do not use the Library (may be a huge number if the Library has many electronic services) |
| Physics 401 (Advanced class for physics majors) | Students majoring in physics | All other students |

**ANSWERS TO END-OF-CHAPTER QUESTIONS**

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.*

A probability sample is one 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. A probability sample is preferred to a nonprobability sample because probability samples are inherently more representative of the population.

For its pros probability sampling methods are those that assure 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. On its cons, probability samples are more time-consuming, more detailed, and more expensive than nonprobability samples. With nonprobability sampling, the pros are less time, less detail, and less cost, but the con is that representation of the population is less, and there may even be great misrepresentation.

1. *List and describe briefly each of the probability sampling methods described in the chapter.*

They are described in brief in Table 12.1, repeated below.

**Simple Random Sampling**

The researcher uses random numbers from a computer, random digit dialing, or some other random selection procedure that guarantees each member of the population in the sample frame has an identical chance of being selected into the sample.

**Systematic Sampling**

Using a sample frame that lists members of the population, the researcher selects a random starting point for the first sample member. A constant *skip interval*, calculated by dividing the number of population members in the sample frame by the sample size, is then used to select every other sample member from the sample frame. A skip interval must be used so 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 sample frame is divided into groups called clusters, each of which must be considered to be similar to the others. The researcher can then randomly select a few clusters and perform a census of each one (one stage). Alternatively, the researcher can randomly select more clusters and take samples from each one (two stage). This method is desirable when highly similar clusters can be easily identified, such as subdivisions spread across a wide geographic area.

**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 usage), the researcher identifies subpopulations in the sample frame 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 (i.e., skewed).

1. *What is meant by the term random? Explain how each of the following embodies randomness:*

With random sampling, the probability of being selected into the sample is “known” for all members of the population. The way in which each method embodies randomness is described below.

*a. “Blind draw”*

Assuming the draw is fair such as in the case of ping pong balls in a revolving drum, each member of the population represented by each ping pong ball has an equal probability of being selected.

*b. Random digit dialing*

Telephone numbers are generated with a random device such as a computer program. All numbers have the same probability of being selected.

*c. Computer generated random numbers*

Random number generators in computers supposedly guarantee that any generated number has an equal probability of being generated.

1. *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?*

The systematic sample’s popularity over simple random sampling is based primarily on the “economic efficiency” that it represents, for 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. 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.

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. Another factor in the representativeness of a systematic sample is sample frame error.

1. *Distinguish cluster sampling from simple random sampling. How are systematic sampling and cluster sampling related?*

The basic difference between cluster sampling and simple random is the sample unit. With cluster sampling, theoretically identical clusters are selected, while in simple random sampling, individual population members are selected. The difference is especially apparent when comparing one-stage cluster to simple random sampling.

A systematic sample is really a one-stage cluster sample. The random starting point determines which of all possible clusters in the list is selected. For example, assume a skip interval of 10. If the starting point is 1, the cluster members are 1, 11, 21, 31, and so forth. If the starting point is 2, cluster members are 2, 12, 22, 32, and so on. Continue itemizing cluster members up to a starting point of 10 where cluster members are 10, 20, 30, 40, and so on. The entire list is now accounted for with the 10 clusters.

1. *Differentiate one-step from two-step area sampling, and indicate when each one is preferred.*

With one-step area sampling, a few areas are selected at random, and a census is undertaken for each area selected. With two-step area sampling, more areas are selected at random, and samples are taken within each selected area. In the one-step approach, the researcher may believe the various geographic areas to be sufficiently identical to permit him or her to concentrate his attention on just one and then generalize the results to the full population. The two-step 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 it to be. But the two-step method is more costly although more areas and time are involved.

1. *What is meant by a “skewed” population? Describe a skewed population distribution variable and provide an example.*

A skewed population is one where a key variable’s distribution is not normal or uniform. Figure 9.2 illustrates that a skewed distribution has a long tail at one end. Students’ examples should embody this basic pattern.

1. *What are some alternative online sampling methods? Describe each one.*

The types are described below.

**Random online intercept sampling** relies on a random selection of web site visitors.

**Invitation online sampling** is when potential respondents are alerted that they may fill out a questionnaire that is hosted at a specific web site.

**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.

**Other online sampling approaches** are feasible and limited only by the creativity of the sample designers.

1. *Briefly describe each of the four nonprobability sampling methods.*

Review question. Students will need to find a description of each nonprobability sampling method.

**Convenience Sampling**

The researcher or interviewer uses a high-traffic location, such as a busy pedestrian area or a shopping mall as the sample from which to intercept potential respondents. Sample frame error occurs in the form of members of the population who are infrequent or nonusers of that location. Other error may result from any arbitrary way the interviewer selects respondents from the sample frame.

**Purposive Sampling**

The researcher uses his or her judgment or that of some other knowledgeable person to identify who will be in the sample. Subjectivity and convenience enter in here; consequently, certain members of the population will have a smaller chance of selection than will others.

**Referral Sampling**

Respondents are asked for the names or identities of others like themselves that 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 low probability of being selected.

**Quota Sampling**

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 for 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.

1. *Why is quota sampling often used with a convenience sampling method such as mall intercepts?*

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.

Convenience samples are inherently not representative because they rely on high traffic locations, and some people are more likely to frequent these locations, while others are less likely to frequent them. The imposition of a quota guarantees that the final sample will have the desired demographic (or other relevant characteristics) profile even though it is drawn by convenience.

1. *Provide the marketing researcher’s definitions for each of the following populations:*

Each is an exercise in population definition. The definitions are listed under each case.

a. *Nest Thermostat, a company that sells a home thermostat that runs on the Internet of Things, wants to determine interest in an entry camera that activates with motion anytime someone enters a dwelling via the front door.*

Owners of homes who are interested in technology

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

Students who buy items on campus and in off-campus stores using a checking account

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

Construction companies who buy air compressors for their projects

1. *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?*

Students must make judgments about sample frame error. The evaluation is provided beneath each case.

***POPULATION SAMPLE FRAME***

*a. Buyers of Scope mouthwash Mailing list of* ***Consumer Reports***

 *subscribers*

Omitted are Scope buyers who do not subscribe to *Consumer Reports*, a very considerable number, no doubt. Included are all those who do subscribe but do not buy Scope, a very considerable percentage. The sample frame error is unacceptable because much of the population is not included, and those members included will have low incidence in the list.

*b. Subscribers to Sirius XM State registration records of new*

 *satellite radio automobile buyers*

Omitted are unlisted and new telephone listings, while included are those with listed phones that do not listen to FM music on the radio. The incidence rate will be low, but the sample frame is acceptable unless there is some peculiarity about listeners who have unlisted phones.

*c. Prospective buyers of a new Members of Sales and Marketing*

*client and prospective Executives International (a national*

*client tracking software product organization of sales managers)*

Omitted are sales persons who do not belong to SMEI, probably a very considerable proportion of all sales persons in the area. Included are retired sales persons, sales managers who no longer do direct sales, and members who have moved away if the list has not been purged recently. The frame error is acceptable for a pilot study, but it does not include a large proportion of the target market.

*d. Users of weatherproof decking Individuals’ names registered*

*materials (to build outdoor at a recent home and garden*

*decks) show*

Omitted are all those who did not attend the show plus those who did attend but did not register. Included are people who have decks, people who do not want a deck, and people who cannot have a deck (e.g., may live in rented home). Assuming the show was well attended, the sample frame error is acceptable.

1. *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 of 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 include the following:*

*Assess 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?*

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.

This is a systematic sample. The sample frame is golfers playing on any day as every day has an equal chance of being selected.

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

This is a convenience sample. The representativeness depends on how long the survey is in place. If for only one day, there is no mention of random selection, so it the sample frame would be golfers renting golf carts on that day. If the survey lasted a week, it the sample frame would be golfers golfing that week and renting golf carts, so it would be more representative than the one-day duration. However, a week might be idiosyncratic such as including a holiday or special country club event, making it less representative that a “typical week.”

c. *Using the city telephone directory and a plus-one dialing procedure. With this procedure, a random page in the directory and a name on that page would be selected, both by means of 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.*

This is effectively a random sample of golfers in the city, so the sample frame is all golfers living in the city. Those with unlisted numbers are automatically included with the plus-one system. It would not be a random sample of Big Tree Country Club members, however, unless there was a qualification that the respondents must be members of the country club.

*14. A researcher has the task of estimating how many units of a revolutionary new high-speed office copier machine (it does not require ink cartridges and is guaranteed not to jam) will be purchased by business firms in Cleveland, Ohio for the upcoming annual sales forecast. Her plan is to ask the likelihood that they will purchase the new device, and for those who are “very likely” to purchase, she wants respondents to estimate how many machines their company will buy. She has data to divide the companies into small, medium, and large firms based on number of employees at the Cleveland office.*

*a. What sampling plan should be used?*

Stratified random sample

*b. Why?*

We know the answers to the research question “How many machines will your company use?” will vary according to size of the company: small, medium, and large.

If we have some pilot study data that gives us an indication as to different variances within each group, we can use a *disproportionate* sample in order to obtain an accurate estimate within each stratum (i.e. small, medium, large) without having to increase the total sample size.

*15. Honda USA is interested in learning what its 550 U.S. dealers think about a new service program the carmaker provided to dealers at the beginning of last year. Honda USA wants to know if the dealers are using the program and, if so, their likes and dislikes about it. The carmaker does not want to survey all 550 dealers but hopes to ensure that the results are representative of all dealers.*

*a. What sampling plan should be used?*

Simple random sample

*b. Why?*

Simple Random Samples are preferred when we have either a. a small population (which can easily be numbered) or b. a sample frame that is available in electronic, i.e. spreadsheet, format. In this case, we have a. and very likely have b.

We should use Simple Random Samples when appropriate. The word “simple” is not to be construed as inferior. Large populations and non-electronic sample frames make Simple Random Sampling difficult.

*16. Applebee’s Restaurants has spent several tens of thousands of dollars advertising the restaurant during the last two years. Marketing executives want to measure what effect the advertising has had, and they decide to measure top-of-mind awareness (TOMA). A TOMA score for such a restaurant is the ranking a firm has as a result of asking a representative sample of consumers in the service area to “name a non-fast-food restaurant.” The restaurant that is named by the most persons has the top TOMA score. It is important that Applebee’s management conduct the TOMA survey on a representative sample in the metropolitan area.*

*a. What sampling plan should be used?*

Systematic Random Sample

*b. Why?*

We need a representative survey so it must be a probability sample and secondly, our population is all consumers. The closest sample frame would be all those consumer households having telephones. Since we have a large sample frame (telephone directory) and it is not on an electronic medium, we need an efficient method for drawing the sample. Using systematic sampling with Plus One dialing will allow us to achieve a representative sample of consumers in the metropolitan area.

*17. Belk has a chain of department stores across the South. Top management requires that each store manager collect, maintain, and respond to customer complaints (emails, letters, calls, etc.). Each store manager is supposed to keep a list of complaints that have been received. Top management is considering establishing a more formalized method of monitoring and evaluating the response managers give to the complaints. They want some information that will tell them whether they need to develop such a formalized program or whether they can leave well enough alone and allow managers to use their discretion in handling the complaints. They want to review a sample of these complaints and the responses to them.*

*a. What sampling plan should be used?*

A nonprobability sample, such as a convenience sample or a purposive sample, should be used.

*b. Why?*

First, management is considering a formal method. We need to collect some information quickly and inexpensively that will allow us to evaluate the need for more information. By asking 10 store managers (convenience sample) to send in a copy of their files and having someone go through and evaluate the responses would give management information that would tell them whether or not they need to pursue further action. Top management could use a judgment sample by simply asking one or two stores in each geographic division to send in their complaint files.

**CASE SOLUTIONS**

**Case 9.1 Peaceful Valley Subdivision: Trouble in Suburbia**

**Case Objective**

This case requires students to ponder various sample methods as to representativeness and bias.

**Answers to Case Questions**

1. *There is only one street into and out of the subdivision. The president is thinking of paying his teenage daughter to stand at the stop light at the entrance to Peaceful Valley next week between the hours of 7:00 and 8:30 a.m. to hand out questionnaires to exiting drivers while they wait for the red light to change. The handouts would include addressed, postage-paid envelopes for returns. Identify what sample method the president would be using, list its pros and cons, and indicate how representative a sample would result.*

This is a convenience sample, using the high traffic location where cars stop for the red light. Pros: quick and convenient for the researcher. Cons: some population members will not be sampled as they may not commute at these times, may work at home, might be sick that week, etc. It would only be representative of those who commute to work at those times on those days.

2.*The chairperson of the Suburb Steering Committee thinks the 1,000 homeowners whose houses are on the waterfront properties of Peaceful Lake are the best ones to survey because they paid more for their lots, their houses are bigger, and they tend to have lived in Peaceful Valley longer than other residents. If these 1,000 homeowners are used for the sample, what sample method would be involved, what are its pros and cons, and how representative a sample would result?*

The chairperson is judging these 1,000 homeowners to be representative, so it is a judgment sample. Pros: quick and easy. Cons: leaves out some members of the population. In this case, it leaves out the 3000 “newer” subdivision households.

3.*Assume that the Steering Committee chairperson’s point that the 1000 waterfront owners are not the same as the 5000 other Peaceful Valley Subdivision homeowners is true. How should this fact be utilized to draw a representative sample of the entire subdivision? Identify the probability sampling method that is most appropriate, and indicate, step-by-step, how it should be applied here.*

The 1000 and 5000 households are strata – meaning that they are different groups but in the same population. So, a stratified random sample is called for. Draw a probability sample from each stratum and do the survey with each stratum sample. Then, if necessary, combine the results with a weighted average formula to determine the overall population average opinion.

4.*How would you select a simple random sample of those Peaceful Valley homeowners who paid their subdivision association dues last year? What, if any, sample bias, might result from this approach?*

Uniquely number each of the payees and use a table of random number or some random number generator such as Excel to randomly select the sample. Every homeowner who paid dues last year would have an equal chance of being selected.

The sample frame here is payees, so those who did not pay and new members who did not live there last year would not be included in the sample frame. If there were a lot of nonpayees and/or a lot of new members, there would be appreciable bias. But the description does not indicate a lot of new members nor does it imply that there are many deadbeats living there.

5. *How could a two-step cluster sample be used here? Identify this sample method and describe how it could be used here to select a representative sample of Peaceful Valley households?*

Consider each the 50 streets with 120 houses as a cluster, so we have 50 identical clusters. Draw a random sample of clusters (streets), and then draw a random sample of houses on each street selected.

**Case 9.2 Jet’s Pets**

**Case Objective**

Jet wants to survey the approximately 10,000 families in two ZIP code areas. Of course, he cannot survey all of them, so he must use a sample. For each of the following possible ways of selecting a sample of the families living in several subdivisions in two ZIP code areas: (1) identify the type of sample method; (2) identify the sample frame; (3) indicate what, if any, sample frame error there is; and (4) indicate the degree to which the resulting sample will be representative of all families living in the two ZIP code areas.

**Answers to Case Questions**

1. *Place questionnaires in veterinarian clinics located in the two ZIP code areas for pet owners to fill out while they are waiting for the doctor to examine their pet.*

This is a convenience sample, a “high traffic” convenience sample.

Sample frame – those who visit the vet clinics during the time of the survey.

Sample frame error – those who do not visit vet clinics during the time period.

It will be representative only of those who visit vet clinics for: problems, boarding, check-ups (also probably more representative of dog and cat owners than owners of other pet types).

*2. Select every 100th name in the city telephone book; call and interview only those who live in the two ZIP code areas.*

• This is a systematic sample

• Sample frame – homes with listed phone numbers

• Error – will not include those whose phones are not listed

• Those unlisted are probably wealthier and more highly educated households

*3. Use a random number system to select a single subdivision located somewhere in the two ZIP code areas, and then place questionnaires in the mailboxes of every home in that selected subdivision.*

• This is a one-step cluster sample (aka area sample)

• Sample frame – all houses in the 2 zip code area

• Error – practically none (assuming the subdivisions are identical)

• It will represent all families in the target area (assuming all subdivisions are equal or very similar)

*4. Announce in the local newspaper a “Cutest Dog Contest” with contestants sending in a photo and address information. Use the contestants who live in the two ZIP code areas as the sample.*

• This is a convenience sample

• Sample frame – those who enter the contest

• Error – only “cute” dog owners will enter the contest, so all other pet owners are not included

• It will represent only those who have “cute” dogs and who bother to enter the contest.

*5. Go to the local animal shelter and get the addresses of the past pet adopters who live in the two ZIP code areas. Send a mail survey to the nearest neighbor’s address for each of the addresses obtained from the animal shelter. For example, if the adopter lives at 1 Green Street, send the mail questionnaire to the occupants at 2 Green Street.*

• It is a convenience sample of pet adopters, but modified by use of a “plus one” type of system (similar to the one used to select phone numbers for all phones but based on only those phone numbers listed in the phone book)

• Sample frame - neighbors of the pet adopters

• Error – not much as the neighbors can be just about anyone in the target area

• It will represent the families in the target area fairly well