4

Sampling

A Reader's Guide to Chapter 4

What Is a Sample?

A *sample* is a portion or subset of a larger group called a *population*. The evaluator's *target population* consists of the institutions, persons, problems, and systems to which or to whom the evaluation's findings are to be applied or generalized. For example, suppose 500 students who are 14 to 16 years of age are selected to be in an evaluation of a program to improve computer programming skills. If the program is effective, the school system's curriculum planners will offer it to all 14- to 16-year-olds in the system. The target population is students between the ages of 14 to 16. The system may include thousands of 14- to 16-year-old students, but only a sample of 500 are included in the evaluation.

Consider the two target populations and samples in Example 4.1.

Example 4.1 Two Target Populations and Two Samples

Evaluation 1

Target population: All homeless veterans throughout the United States

Program: Outreach, provision of single-room occupancy housing, medical and financial assistance, and job training: the REACH-OUT Program

Sample: 500 homeless veterans in four of the fifty U.S. states who received outpatient medical care between April 1 and June 30

Evaluation 2

Target population: All newly diagnosed breast cancer patients

Program: Education in Options for Treatment

Sample: Five hospitals in three of fifty U.S. states; within each hospital, 15 physicians; for each physician, 20 patients seen between January 1 and July 31 who are newly diagnosed with breast cancer

The REACH-OUT Program in Evaluation 1 is designed for all homeless veterans. The evaluator plans to select a sample of 500 homeless veterans in four states between April 1 and June 30. The findings are to be applied to all homeless veterans in all 50 states. Women newly diagnosed with breast cancer are the targets of an educational program in Evaluation 2. The evaluators will select five hospitals and, within them, 20 patients seen from January through July by each of 15 doctors. The findings are to be applied to all patients newly diagnosed with breast cancer.

Why Sample?

Evaluators sample because it is efficient to do so and because it contributes to the precision of their research. Samples can be studied more quickly than entire target populations, and they are also less expensive to assemble. In some cases, it would be impossible to recruit a complete population for an evaluation, even if the necessary time and financial resources were available. For example, it would be futile for an evaluator to attempt to enroll all homeless veterans in an investigation of a program that targets them (see Example 4.1, above). Sampling is also efficient in that it allows the evaluator to invest resources in important evaluation activities, such as monitoring the quality of data collection and standardizing the implementation of the program rather than in the collection of data on an unnecessarily large number of individuals or groups.

Sampling enables the evaluator to focus precisely on the characteristics of interest. For example, suppose an evaluator wants to compare older and younger veterans with differing health and functional statuses. A stratified sampling strategy can give the evaluator just what he or she needs. A sample of the population with precise characteristics is actually more suitable for many evaluations than the entire population.

Inclusion and Exclusion Criteria or Eligibility

A sample is a part of a larger population where the evaluation's findings from the sample will be applied to make inferences or extrapolations about the larger population. For instance, if an evaluation is intended to investigate the impact of an educational program on women's knowledge of their options for surgical treatment for cancer, and not all women with cancer are to be included in the program, the evaluator has to decide on the characteristics of the women who will be the focus of the study. Will the evaluation concentrate on women of a specific age? Women with a particular kind of cancer? Example 4.2 presents hypothetical inclusion and exclusion criteria for an evaluation of such a program.

Example 4.2 Inclusion and Exclusion Criteria for a Sample of Women to Be Included in an Evaluation of a Program for Surgical Cancer Patients

Inclusion: Using the U.S. Medicare claims database, of all patients hospitalized during 2013, those with diagnosis or procedure codes related to breast cancer; for patients with multiple admissions, only the admission with the most invasive surgery

Exclusion: Women under the age of 65 (because women under 65 who receive Medicare in the United States are generally disabled or have renal disease), men, women with only a diagnostic biopsy or no breast surgery, women undergoing bilateral mastectomy, women without a code for primary breast cancer at the time of the most invasive surgery, women with a diagnosis of carcinoma in situ, and women with metastases to regions other than the axillary lymph nodes

The evaluator of this program has set criteria for the sample of women who will be included in the evaluation and for which its conclusions will be appropriate. The sample will include women over 65 years of age who have had surgery for breast cancer. The findings of the evaluation will not be applicable to women under age 65 with other types of cancer who had only a diagnostic biopsy, have not had surgery, or had a bilateral mastectomy.

The independent variables are the evaluator's guide to determining where to set inclusion and exclusion criteria. For example, suppose that in an evaluation of the effects on teens of a preventive health care program, one of the evaluation questions asks whether boys and girls benefit equally from program participation. In this question, the independent variable is sex and the dependent (outcome) variable is benefit. If the evaluator plans to sample boys and girls, he or she must set inclusion and exclusion criteria. Inclusion criteria for this evaluation might include boys and girls under 18 years of age who are likely to attend all of the educational activities for the duration of the evaluation and who are able to read at levels appropriate for their ages. Teens might be excluded if they already participate in another preventive health care program, if they do not speak English, or if their parents object to their participation. If these hypothetical inclusion and exclusion criteria are used to guide sampling eligibility, then the evaluation's findings can be generalized only to English-speaking boys and girls under age 18 who read appropriately for their age and tend to be compliant with school attendance requirements. The evaluation is not designed to enable the findings to be applicable to teens who have difficulty reading or speaking English and who are unlikely to be able to complete all program activities.

Sampling Methods

Sampling methods are usually divided into two types: probability sampling and convenience sampling. Probability sampling is considered the best way to ensure the validity of any inferences made about a program's effectiveness and its generalizability. In probability sampling, every member of the target population has a known probability of being included in the sample.

In convenience sampling, participants are selected because they are available. Thus, in this kind of sampling some members of the target population have a chance of being chosen whereas others do not, even if they meet inclusion criteria. As a result, the data that are collected from a convenience sample may not be applicable at all to the target group as a whole. For example, suppose an evaluator who is studying the quality of care provided by student health services decides to interview all students who come for care during the week of December 26. Suppose also that 100 students come and all agree to be interviewed: a perfect response rate. The problem is that in some parts of the world, the end of December is associated with increased rates of illness from respiratory viruses as well as with high numbers of skiing accidents; moreover, many universities are closed during that week, and most students are not around. Thus, the data collected by the happy evaluator with the perfect response rate could very well be biased, because the evaluation excluded many students simply because they were not on campus (and, if they were ill, may have received care elsewhere).

Simple Random Sampling

In simple random sampling, every subject or unit has an equal chance of being selected. Because of this equality of opportunity, random samples are considered relatively unbiased. Typical ways of selecting a simple random sample include applying a table of random numbers (available free online) or a computer-generated list of random numbers to lists of prospective participants.

Suppose an evaluation team wants to select 10 social workers at random from a list (or sampling frame) of 20 names. Using a table of random numbers (Figure 4.1), the evaluators first identify a row of numbers at random and then a column. Where the two intersect, they begin to identify their sample.

1. *How they randomly identify the row:* A member of the evaluation team tosses a die twice. On the first toss, the die's number is 3; the number on the second toss is 5. Starting with the first column in the table, these numbers correspond to the third block and the fifth row of that block, or the row containing numbers 1 4 5 7 5 in Figure 4.1.
2. *How they randomly identify the column:* A member of the team tosses the die twice and gets 2 and 1. These numbers identify the second block of columns and the first column beginning with 1. The starting point for this sample is where the row 1 4 5 7 5 and the column beginning with 1 intersect, at 3 5 4 9 0.
3. *How they select the sample:* The evaluators need 10 social workers from their list of 20—or two-digit numbers. Moving down from the second column, the 10 numbers between 01 and 20 that appear (starting with the numbers below 3 5 4 9 0, and beginning with the first number, 7) are 12, 20, 09, 01, 02, 13, 18, 03, 04, and 16. These are the social workers selected for the random sample.

Random Selection and Random Assignment

In any given evaluation, the process of random selection may be different from the process of random assignment, as is illustrated in Example 4.3.

Figure 4.1 A Portion Table of Random Numbers

18283	19704[2]	45387	23476	12323	34865
46453	21547	39246	93198	98005	65988
19076	23453	32760	27166	75032	99945
36743	89563	**12**378	98223	23465	25408
22125	19786	23498	76575	76435	63442
76009	77099	43788	36659	74399	**03**432
09878	76549	88877	26587	44633	77659
34534	44475	56632	34350	**01**768	29027
83109	75899	34877	21357	24300	00869
89063	43555	32700	76497	36099	97956
			94656	34689	
09887	67770	69975	54465	**13**896	**04**645
23280	34572	99443	98765	34978	42880
93856	23090	22257	67400	23580	24376
21256	50863	56934	70993	34765	30996
14575[1]	35490[3]	23645	22179	35788	37600
23276	7[4]0870	**20**087	66665	78876	58007
87530	45738	**09**998	45397	47500	34875
00791	32164	97665	27589	90087	**16**004
99003	32567	**02**878	38602	**18**700	23455
14367	64999	78453	40078	53727	28759

Note: Numbers that are bold and underlined = sample of ten, consisting of numbers between 01 and 20.

1 = Two rolls of a die yield 1 (column) and 5 (block).

2 = Two rolls of a die yield 2 (column) and 1 (row).

3 = Intersection of 1 and 2.

4 = Start here to get the sample.

Example 4.3 Random Selection and Random Assignment: Two Illustrations

Evaluation A

Evaluation A had six months to identify a sample of teens to participate in an evaluation of an innovative school-based preventive health care program. At the end of the six months, all eligible teens were assigned to the innovative program or the traditional (control) program. Assignment was based on how close to a participating school each teen lived. The evaluators used this method because participation meant that students would be required to attend several after-school activities, and no funding was available for transportation.

Evaluation B

Evaluation B had six months to identify a sample of teens to participate in an evaluation of an innovative school-based preventive health care program. At the end of the six months, a sample was randomly selected from the population of all teens who were eligible. Half the sample was randomly assigned to the innovative program, and half was assigned to the traditional (control) program.

In the example of Evaluation A, the evaluators selected all eligible teens and then assigned them to the experimental and control groups according to how close each one lived to a participating school. In Evaluation B, the evaluators selected a random sample of all those who were eligible and then randomly assigned the participants either to the experimental group or the control group. Of these two methods, the latter is usually considered preferable to the former. *Random selection* means that every eligible person has an equal chance of becoming part of the sample; if all are included because they just happen to appear during the time allocated for sampling, biases may be introduced. Random assignment can also guard against bias.

Systematic Sampling

Suppose an evaluator is to select a sample of 500 psychologists from a list with the names of 3,000 psychologists. In systematic sampling, the evaluator divides 3,000 by 500 to yield 6, and then selects every sixth name on the list. An alternative is to select a number at random—say, by tossing a die. If the toss comes up with the number 5, for example, the evaluator selects the fifth name on the list, then the tenth, the fifteenth, and so on, until he or she has the necessary 500 names.

Systematic sampling is not an appropriate method to use if repetition is a natural component of the sampling frame. For example, if the frame is a list of names, those beginning with certain letters of the alphabet (in English: Q or X) might get excluded because they appear infrequently.

Stratified Sampling

A *stratified random sample* is one in which the population is divided into subgroups, or *strata,* and a random sample is then selected from each group. For example, in a program to teach women about options for the treatment of breast cancer, the evaluator might choose to sample women of differing general health status (as indicated by scores on a 32-item test), age, and income (high = +, medium = 0, and low = –). In this case, health status, age, and income are the strata. This sampling blueprint can be depicted as shown in Figure 4.2.

The evaluator chooses the strata, or subgroups, based on his or her knowledge of their relationship to the dependent variable or outcome measure—in this case, the options chosen by women with breast cancer. That is, the evaluator in this example has evidence to support the assumption that general health status, age, and income influence women's choices of treatment. The evaluator must be able to justify his or her selection of the strata with evidence from the literature or other sources of information (such as historical data or expert opinion).

Figure 4.2 Sampling Blueprint for a Program to Educate Women in Options for Breast Cancer Treatment

Health Status Scores and Income	Age (years)				
25–32 points High income	< 35	35–55	56–65	66–75	> 75
Average					
Low					
17–24 points High income					
Average					
Low					
9–16 points High income					
Average					
Low					
1–8 points High income					
Average					
Low					

If the evaluator neglects to use stratification in the choice of a sample, the final results may be confounded. For example, if the evaluation neglects to distinguish among women with different characteristics, good and poor performance may be averaged among them, and the program will seem to have no effect even if women in one or more groups benefited. In fact, the program actually might have been very successful with certain women, such as those over age 75 who have moderate incomes and General Health Status scores between 25 and 32.

When evaluators do not use stratification, they may apply statistical techniques (such as analysis of covariance and regression) retrospectively (after the data have already been collected) to correct for confounders or covariates on the dependent variables or outcomes. Evaluators generally agree, however, that it is better to anticipate confounding variables by sampling prospectively than to correct for them retrospectively, through analysis. Statistical corrections require very strict assumptions about the nature of the data, and the sampling plan may not have been designed for these assumptions. With few exceptions, using statistical corrections afterward results in a loss of power, or the ability to detect true differences between groups (such as the experimental and control groups).

The strata in stratified sampling are subsets of the independent variables. If the independent variables are sex, health status, and education, the strata are how each of these is defined. For example, sex is defined as male and female. A variable like health status can be defined in many ways, depending on the measures available to collect data and the needs of the evaluation. For example, health status may be defined as a numerical score on some measure or may be rated as excellent, very good, good, fair, or poor.

Cluster Sampling

Cluster sampling is used in large evaluations—those that involve many settings, such as universities, hospitals, cities, states, and so on. In cluster sampling, the population is divided into batches that are then randomly selected and assigned, and their constituents can be randomly selected and assigned. For example, suppose that California's counties are trying out a new program to improve emergency care for critically ill and injured children, and the control program is the traditional emergency medical system. If you want to use random cluster sampling to evaluate this program, you can consider each county to be a cluster and select and assign counties at random to the new program or the traditional program. Alternatively, you can randomly select children's hospitals and other facilities treating critically ill children within counties and randomly assign them to the experimental system or the traditional system (assuming this is considered ethical). Example 4.4 gives an illustration of the use of cluster sampling in a survey of Italian parents' attitudes toward AIDS education in their children's schools.

Example 4.4 Cluster Sampling in a Study of Attitudes of Italian Parents Toward AIDS Education

Epidemiologists from 14 of Italy's 21 regions surveyed parents of 725 students from 30 schools chosen by a cluster sampling technique from among the 292 classical, scientific, and technical high schools in Rome. Staff visited each school and selected students using a list of random numbers based on the school's size. Each of the students selected for participation was given a letter to deliver to his or her parents explaining the goals of the study and when they would be contacted.

Nonprobability or Convenience Sampling

Convenience samples are samples where the probability of selection is unknown. Evaluators use convenience samples simply because they are easy to obtain. This means that some people have no chance at selection simply because they are not around to be chosen. Convenience samples are considered to be biased, or not representative of the target population, unless proved otherwise.

In some cases, evaluators can perform statistical analyses to demonstrate that convenience samples are actually representative. For example, suppose that during the months of July and August, an evaluator conducts a survey of the needs of county institutions concerned with critically ill and injured children. Because many county employees take their yearly vacations in July and August, the respondents may be different from those who would have answered the survey during other times of the year. If the evaluator wants to demonstrate that those employees who were around to respond and those who were not available in July and August are not different, he or she can compare the two groups on key variables, such as time on the job and experience with critically ill and injured children. If this comparison reveals no differences, the evaluator is in a relatively stronger position to assert that, even though the sample was chosen on the basis of convenience, the characteristics of the participants do not differ on certain key variables (such as length of time on the job) from those of the target population.

The Sampling Unit

A major concern in sampling is the *unit* to be sampled. Example 4.5 illustrates the concept of the sampling unit.

Example 4.5 What Is the Target? Who Is Sampled?

An evaluation of a new program is concerned with measuring the program's effectiveness in altering physicians' practices pertaining to acute pain management for children who have undergone operations. The target population is all physicians who care for children undergoing operations. The evaluation question is "Have physicians improved their pain management practices for children?" The evidence of effectiveness is that physicians in the experimental group show significant improvement over a 1-year period and significantly greater improvement than physicians in a control group. Resources are available for 20 physicians to participate in the evaluation.

The evaluators randomly assign 10 physicians to the experimental group and 10 physicians to the control. They plan to find out about pain management through a review of the medical records of 10 patients of each of the physicians in the experimental and control groups, for a total sample of 200 patients. (This is sometimes called a *nested design*.) A consultant to the evaluation team says that, in actuality, the evaluators are comparing the practices of 10 physicians against those of 10 physicians, and not the care of 100 patients against that of 100 patients. The reason is that characteristics of the care of the patients of any single physician will be very highly related. The consultant advises the evaluators to correct for this lack of "independence" among patients of the same physician by using one of the statistical methods available for correcting for cluster effects. Another consultant, in contrast, advises the evaluators to use a much larger number of patients per physician and suggests a statistical method for selecting the appropriate number. Because the evaluators do not have enough money to enlarge the sample, they decide to "correct" statistically for the dependence among patients.

In this example, the evaluators want to be able to apply the evaluation's findings to all physicians who care for children undergoing surgery, but they have enough resources to include only 20 physicians in the evaluation. In an ideal world, the evaluators would have access to a very large number of physicians, but in the real world, they have only the resources to study 10 patients per physician and access to statistical methods to correct for possible biases. These statistical methods enable evaluators to provide remedies for the possible dependence among the patients of a single physician, among students at a single institution, among health care workers at a single hospital, and so on.

Sample Size

Power Analysis and Alpha and Beta Errors

An evaluation's ability to detect an effect is referred to as its *power*. A power analysis is a statistical method for identifying a sample size that will be large enough to allow the evaluator to detect an effect if one actually exists. A commonly used evaluation research design is one where the evaluator compares two randomly assigned groups to find out whether differences exist between them. Accordingly, a typical evaluation question is "Does Program A differ from Program B in its ability to improve quality of life?" To answer this question accurately, the evaluator must be sure that enough people are in each program group so that if a difference is actually present, it will be uncovered. Conversely, if there is no difference between the two groups, the evaluator does not want to conclude falsely that there is one. To begin the process of making sure that the sample size is adequate to detect any true differences, the evaluator's first step is to reframe the appropriate evaluation questions into *null* hypotheses. Null hypotheses state that no difference exists between groups, as illustrated in Example 4.6.

Example 4.6 The Null Hypothesis in a Program to Improve Quality of Life

Question: Does Experimental Program A improve quality of life?

Evidence: A statistically significant difference is found in quality of life between Experimental Program A's participants and Control Program B's participants, and the difference is in Program A's favor

Data source: The Quality of Life Assessment, a 30-minute self-administered questionnaire with 100 questions. Scores on the Assessment range from 1 to 100, with 100 meaning excellent quality of life

Null hypothesis: No difference in quality of life exists between participants in Program A and participants in Program B. In other words, the average scores on the Quality of Life Assessment obtained by participants in Program A and in Program B are equal

When an evaluator finds that differences exist among programs, but in reality there are no differences, that is called an *alpha error* or *Type I error.* A Type I error is analogous to a false-positive test result; that is, a result indicating that a disease is present when in actuality it is not. When an evaluator finds that no differences exist among programs, but in reality differences exist, that is termed a *beta error* or *Type II error.* A Type II error is analogous to a false-negative test result; that is, a result indicating that a disease is not present when in actuality it is. The relationship between what the evaluator finds and the true situation can be depicted as shown in Figure 4.3.

Figure 4.3 Type I and Type II Errors: Searching for a True Difference

		Truth	
		Differences exist	**No differences exist**
Evaluator's conclusions from hypothesis test	Differences exist (reject null)	Correct	Type I or alpha error
	No differences exist (keep null)	Type II or beta error	Correct

To select sample sizes that will maximize the power of their evaluations, evaluators must rely on formulas whose use requires an understanding of hypothesis testing and a basic knowledge of statistics. Evaluators using these formulas usually must perform the following steps:

- State the null hypothesis.
- Set a level (alpha or ∝) of statistical significance—usually .05 or .01—and decide whether it is to be a one- or two-tailed test. A two-tailed test will use a statistical method to determine if the mean (obtained average score) is significantly greater than x (the predicted mean score) **and** if the mean is significantly less than x (the predicted mean score). A one-tailed test is used to find out if the obtained mean is either statistically greater **or** less than the predicted mean. When using a one-tailed test, you are testing for the possibility of the relationship in one direction and completely disregarding the possibility of a relationship in the other direction.
- Decide on the smallest acceptable meaningful difference (e.g., the difference in average scores between groups must be at least 15 points).
- Set the power (1—β) of the evaluation, or the chance of detecting a difference (usually 80% or 90%).
- Estimate the standard deviation (assuming that the distribution of the data is normal) in the population.

Some researchers have proposed alternative sample size calculations based on confidence intervals. A confidence interval is computed from sample data that have a given probability that the unknown parameter (such as the mean) is contained within the interval. Common confidence intervals are 90%, 95%, and 99%.

Calculating sample size is a technical activity that requires some knowledge of statistics. Several easy-to-use programs for calculating sample size are currently available for free online. To find one, enter the search term "sample size" into any search engine.

The Sampling Report

Evaluators can use the sampling report (SR) form (Figure 4.4) for planning and explaining their evaluations. The SR contains the evaluation questions and evidence, the independent variables and strata, the evaluation design, inclusion and exclusion criteria, the dependent variable, the data source, the criteria for level of acceptable statistical and clinical (or educational or practical) significance, the sampling methods, and the size of the sample.

The form in Figure 4.4 shows the use of the SR for one evaluation question asked in an 18-month program combining diet and exercise to improve health status and quality of life for persons 65 years of age or older. An experimental group of elderly people who still live at home will receive the program and another group will not. To be eligible, participants must be able to live independently. People who are under 65 years of age and those who do not speak English or Spanish are not eligible. Participants will be randomly assigned to the experimental or control groups according to the streets on which they live (that is, participants living on Street A will be randomly assigned, as will participants living on Streets B, C, and so on). The evaluators will be investigating whether the program effectively improves quality of life for men and women equally. A random sample of men and women will be selected from all who are eligible, but no two will live on the same street. Then men and women will be assigned at random to the experimental group or the control group.

Figure 4.4 The Sampling Report Form

The Evaluation	The Report
Evaluation questions	1. To what extent has quality of life improved? 2. Do men and women differ in quality of life after participation?
Evidence of effectiveness	A statistical and clinically meaningful difference between experimental and control group, favoring the experimental group Men and women benefit equally
Evaluation design	An experimental design, with randomly assigned parallel controls
Independent variables	Group participation; sex
Strata	Group participation: experimental and control; sex: male and female

The Evaluation	The Report
Inclusion criteria	1. Must be living at home 2. Must be functionally independent
Exclusion criteria	1. Unable to speak English or Spanish 2. 64 years of age or younger
Dependent variable	Quality of life
Data Source	Quality of Life Questionnaire: a 100-point survey. The Questionnaire's manual states that, when used with persons over 70 years of age, the standard deviation (a measure of how much the score "spread" from the mean or average score) is 10 points
Criterion for clinical meaning	A difference of at least 5 points on the Quality of Life Questionnaire between experimental and control, favoring the experimental group Statistical significance (alpha) = .01; power is 80% (beta = .20)
Sampling method	For group assignment: cluster sampling All streets in the town are eligible All eligible persons in a given street are randomly assigned to the experimental or control group For gender: An equal number of men and women will be randomly selected from all who are eligible; men and women will be randomly assigned to an experimental or a control group No two participants will live on the same street.
Sample size	To compare men and women, a total of 188 people are needed: 94 men and 94 women; 47 of each will be assigned to the experimental or control groups

Summary and Transition to the Next Chapter on Collecting Information

This chapter discusses the advantages and limitations of probability and convenience sampling and how to think about the sample unit and sample size. The next chapter discusses the evaluator's data collection choices.

Exercises

Directions

For each of the following situations, choose the best sampling method from among these choices:

A. Simple random sampling
B. Stratified sampling
C. Cluster sampling
D. Systematic sampling
E. Convenience sampling

Situation 1

The Rehabilitation Center has 40 separate family counseling groups, each with about 30 participants. The director of the Center has noticed a decline in attendance rates and has decided to try out an experimental program to improve them. The program is very expensive, and the Center can afford to finance only a 250-person program at first. If the evaluator randomly selects individuals from among all group members, this will create friction and disturb the integrity of some of the groups. As an alternative, the evaluator has suggested a plan in which five of the groups—150 people—will be randomly selected to take part in the experimental program and five groups will participate in the control.

Situation 2

The Medical Center has developed a new program to teach patients about cardiovascular fitness. An evaluation is being conducted to determine how effective the program is with males and females of different ages. The evaluation design is experimental, with concurrent controls. In this design, the new and traditional cardiovascular programs are compared. About 310 people signed up for the winter seminar. Of these, 140 are between 45 and 60 years old, and 62 of these 140 were men. The remaining 170 are between 61 and 75 years old, and 80 of these are men. The evaluators randomly selected 40 persons from each of the four subgroups and randomly assigned every other person to the new program and the remainder to the old program.

Situation 3

A total of 200 health education teachers signed up for a continuing education program. Only 50 teachers from this group, however, are to participate in an evaluation of the program's impact. The evaluator assigns each participant a number from 001 to 200 and, using a table, selects 50 names by moving down columns of three-digit random numbers and taking the first 50 numbers within the range 001 to 200.

References and Suggested Readings

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dawson, B., & Trapp, R. G. (2005). *Basic and clinical biostatistics* (3rd ed.). New York: Lange Medical/McGraw-Hill.

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., Hearst, N., & Newman, T. B. (2006). *Designing clinical research* (2nd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Suggested Websites

These websites provide definitions and explanations of sampling methods. For other sites, search for statistical sampling; how to sample; sampling methods; sampling strategies.

http://www.cliffsnotes.com/study_guide/Populations-Samples-Parameters-and-Statistics.topicArticleId-267532,articleId-267478.html

http://www.ma.utexas.edu/users/parker/sampling/srs.htm

http://www.census.gov/history/www/innovations/data_collection/developing_sampling_techniques.html

http://www.youtube.com/watch?v=HldwaasDP2A