This book is dedicated to the ones I love:
John C. Beck and Ingvard

HOW TO CONDUCT SURVEYS

5 EDITION

A Step-by-Step Guide

Arlene Fink

University of California at Los Angeles,
The Langley Research Institute

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Acquisitions Editor: Vicki Knight
Assistant Editor: Kalie Koscielak
Production Editor: Eric Garner
Copy Editor: Mark Bast
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Wendy Jo Dymond
Indexer: Sheila Bodell
Cover Designer: Anupama Krishnan
Marketing Manager: Nicole Elliott
Permissions Editor: Karen Ehrmann

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Fink, Arlene.

How to conduct surveys : a step-by-step guide / Arlene Fink.—5th ed.

p. cm.
Includes bibliographical references and index.

ISBN 978-1-4522-0387-4 (pbk.)

1. Social surveys. 2. Educational surveys. I. Title.

HN29.F53 2013
300.72'3—dc23 2012038588

This book is printed on acid-free paper.

Certified Chain of Custody
Promoting Sustainable Forestry
SUSTAINABLE FORESTRY INITIATIVE
www.sfiprogram.org
SFI-01268
SFI label applies to text stock

13 14 15 16 10 9 8 7 6 5 4 3 2

BRIEF CONTENTS

Three professors receive a grant from the National Library of Education to do an online survey of 60 schools to find out about the success of their data exchange system. Each professor teaches at a different college. Two of the professors receive permission from their respective IRBs to gain access to respondents' e-mail addresses in order to send out the survey, follow up with nonrespondents, and perform the statistical analysis. The third professor does not have explicit permission to do any of these things, but he is in charge of sending out the financial reimbursements for participation in the survey. The three professors used the E-Line Online Survey Company to build their survey, so with the proper user name and password, anyone, including the third professor, can log on to the survey site and view the results by individual and by question. The third professor needs to obtain e-mail addresses to ask for mailing addresses, which are not included in the survey. He logs on to the survey and records the e-mail addresses. He does not look at any responses to the survey because he does not have IRB clearance to do so. He e-mails the participants and, immediately upon receiving their mailing addresses, sends out the reimbursement checks.

ARTICLES

Asgeirsdottir, B. B., Sigfusdottir, I. D., Gudjonsson, G. H., & Sigurdsson, J. F. (2011). Associations between sexual abuse and family conflict/violence, self-injurious behavior, and substance use: The mediating role of depressed mood and anger. *Child Abuse & Neglect*, *35*(3), 210–219.

Bauer, D. L., Whitman, L. A., & Kosson, D. S. (2011). Reliability and construct validity of psychopathy checklist. *Criminal Justice and Behavior*, *38*(10), 965–987.

Dunn, K. I., Mohr, P., Wilson, C. J., & Wittert, G. A. (2011). Determinants of fast-food consumption. An application of the theory of planned behaviour. *Appetite*, *57*(2), 349–357. doi: 10.1016/j.appet.2011.06.004.

Flynn, K. E., Smith, M. A., & Freese, J. (2006). When do older adults turn to the Internet for health information? Findings from the Wisconsin Longitudinal Study. *Journal of General Internal Medicine*, *21*(12), 1295–1301.

Graham, J. M. (2011). Measuring love in romantic relationships: A meta-analysis. *Journal of Social and Personal Relationships*, *28*(6), 748–771.

Holman, E. A., Silver, R. C., Poulin, M., Andersen, J., Gil-Rivas, V., & McIntosh, D. N. (2008). Terrorism, acute stress, and cardiovascular health: A 3-year national study following the September 11th attacks. *Archives of General Psychiatry*, *65*(1), 73–80.

O'Connor, T. M., Yang, S.-J., & Nicklas, T. A. (2006). Beverage intake among preschool children and its effect on weight status. *Pediatrics*, *118*(4), e1010–e1018.

Overby, M., Carrell, T., & Bernthal, J. (2007). Teachers' perceptions of students with speech sound disorders: A quantitative and qualitative analysis. *Language, Speech, & Hearing Services in Schools*, *38*(4), 327–341.

Richardson, J., Coid, J., Petruckevitch, A., Chung, W. S., Moorey, S., & Feder, G. (2002). Identifying domestic violence: Cross sectional study in primary care. *British Medical Journal*, *324*(7332), 274.

Skoffer, B. (2007). Low back pain in 15- to 16-year-old children in relation to school furniture and carrying of the school Bag. *Spine*, *32*(24), E713.

Wolfson, M., Champion, H., Rogers, T., Neiberg, R. H., Barker, D. C., Talton, J. W., et al. (2011). Evaluation of free to grow: Head start partnerships to promote substance-free communities. *Evaluation Review*, *35*(2), 153–188.

4 SAMPLING

OVERVIEW

Should you survey everyone or just take a sample? The answer depends on how quickly you need results, if the credibility of the findings will suffer if someone or some group is left out, and your financial and technical resources.

Sampling can be divided into two categories. A random sample is selected by an objective method (such as drawing names at random from a hat), and you can also calculate each person's chances of selection. Nonrandom samples are convenient: You select only those respondents who are nearby and willing and available to complete the survey.

How large should a sample be? Relatively larger samples reduce sampling errors. If you want to evaluate the performance of two (or more) groups, statistical methods can guide you in selecting sample sizes large enough to detect differences between the groups—if differences actually exist. Statistical methods may also tell you how large a difference you can observe (and if you will be able to observe differences) with the sample size you have.

The response rate consists of the number of completed surveys divided by the number of surveys eligible for completion. To improve response rate, make sure that respondents can easily complete and submit the survey, train surveyors to administer surveys and follow up, monitor the quality of survey administration, keep responses confidential, and provide respondents with incentives and rewards when possible and ethical.

Sample Size and Response Rate: Who and How Many?

When you conduct a survey, do you have to include everyone? The answer is probably not. Often, the target population—"everyone"—consists of a large number of people. Surveying them is not practical.

If you decide to sample, you must ask the following: How many people should be included? If your company has 1,000 employees, and you want to survey just some of them, how do you decide how many people to include? Say you want to compare long- and short-term job satisfaction. Statistical methods can help you decide how many persons should be included each time (short versus long term) so that if a difference in satisfaction actually occurs over time, you will be able to detect it. Say you have 3 months to conduct the survey, and you think you can survey about 300 of the company's employees. Statistical methods can help you determine how much of an effect you can detect, given the size of the sample your resources allow you to assemble.

Suppose you want to find out whether your neighbors will support a community vigilance program in which each household takes responsibility for watching at least one other house when the owners are away. Consider also that you define your community as having 1,000 homes. Do you need to include all households? If you do, will the program be more likely to work than otherwise? Here are some questions you should answer.

1. *How quickly are data needed?* Suppose a recent increase in house burglaries is the motivation for the survey and you want to get started immediately. If you wait to survey all 1,000 homes in your neighborhood, you may waste precious time.
2. *What type of survey is planned?* If you are going to use a telephone or online survey, your survey will probably take less time than if you plan to interview people in their homes or wait for them to return their surveys in the mail.
3. *What are your resources?* If they are limited, you have to select a survey method such as telephone interviewing rather home interviewing.
4. *How credible will your findings be?* If all 1,000 homes participate, then you will have no problem arguing that your survey is representative of the neighborhood. If only 10 homes participate, you will run into a credibility problem.
5. *How familiar are you with sampling methods?* Sampling methods can be relatively simple or complex. National polls and large research studies use sophisticated techniques that depend on the skills of statisticians and other trained experts. Other methods may be easier to implement, but the resulting sample may not be an accurate reflection of the larger population.

How do you select a sample? You can use a random or convenience sampling method. A random sample is selected objectively. An example would be the use of special software to randomly select winning lottery numbers. Statistical methods are available to calculate the probability that each person has of being chosen (or winning the lottery). A convenience sample includes people who are available and willing to take the survey. The sample selection process is not considered "objective" because not every eligible person has an equal chance. If you are ill on the day of the survey, you will not be able to participate even if you meet all other requirements.

Consider these two cases.

Example: Random and Convenience Sampling

Case 1. A survey aims to find out how teachers in the Loeb School feel about certain school reforms. All 100 teachers' names are put into a hat, the names are jumbled, and the principal selects 50 for the survey.

Case 2. A survey is conducted to find out Los Hadas School District teachers' views on certain school reforms. Ten teachers are chosen to be interviewed from each of the district's six elementary schools, 15 are selected from its four intermediate schools, and 60 are chosen from its one high school. Participating teachers are volunteers who were recommended by their principal. They meet the following criteria: They have been teaching in the district for 5 years or more, they belong to one of three teachers' associations or unions, and they have participated in at least one meeting during the past year on the district's school reform.

■

In the first survey, a sample of 50 of 100 teachers is chosen from a hat. This type of sampling is consistent with random sampling because the selection method is objective, and you can use a mathematical formula to calculate the probability of each person being selected. Also, you can usually expect that the two groups of teachers will not be systematically different from one another. In each group of 50, you are likely to have similar numbers of people who are smart, healthy, and generous and similar numbers who are not smart, unhealthy, and miserly.

In the second survey example, principals make recommendations, and eligible teachers voluntarily choose to participate. Principals may have their favorites. Teachers who volunteer may systematically differ from those who do not. They may be more enthusiastic about the survey or have more time to complete it, for example. This sampling strategy is not objective. It comes up with a nonrandom sample.

Random Sampling Methods

Three of the most commonly used random sampling methods are the following:

- Simple random sampling
- Stratified random sampling
- Simple random cluster sampling
- Stratified random cluster sampling

A simple random sample is one in which each person has an equal chance of being selected from a population. The population contains everyone who is eligible for the survey.

The following is a simple random sampling.

Example: Simple Random Sampling

You want to select 100 people from philanthropic foundations to survey them about the types of grants they sponsor. A total of 400 people can provide this information. You place their names in any order. Each name is given a number from 001 to 400. Then, using a table of random numbers (found in statistics textbooks), you select the first 100 people whose numbers show up on the table.

■

Each person in this scenario has an equal opportunity for selection. The population consists of 400 people. Chance alone decides which of them is sampled.

The following is not random sampling.

Example: Not Random Sampling

You want to sample 100 people from philanthropic foundations to survey them about the types of grants they sponsor. A total of 400 people can provide this information. You select 25 people in each of four areas of the country.

■

Some people in this scenario have no chance of selection: those who do not live in your four chosen geographic areas.

Here is another example of simple random sampling.

Example: Simple Random Sampling

Two hundred nurses, therapists, and social workers employed by a Midwest city signed up for an adult day care seminar. The city only had enough money to pay for 50 participants. The seminar director used the random assignment feature in her statistics program software to select 50 names.

■

To facilitate simple random sampling for telephone surveys, some surveyors use a technique called random digit dialing. In one of its variations, called the plus-one approach, a digit is added to the telephone number that is actually selected. If the selected telephone number is (311) 459-4231, the number called is (311) 459-4232. This technique helps to make up for the fact that in many areas of the country, particularly in urban areas, people do not list their telephone numbers. These people are not fair shakes for selection for telephone surveys.

Making the Decision

When should you choose simple random sampling? The advantages of simple random sampling are these:

- It is the simplest of all random sampling methods.
- Aids are available to assist you. Most statistics textbooks have easy-to-use tables for drawing a random sample. An alternative is to use the random-number feature found in all statistical software.

A major disadvantage of simple random sampling is that it cannot be used to divide respondents into subgroups or strata (e.g., 60% male and 40% female). To make sure you have the proportions you need in a sample, you need to stratify.

Stratified Random Sampling

In simple random sampling, you choose a subset of respondents at random from a population. In stratified random sampling, you first subdivide the population into subgroups or strata and select a given number or proportion of respondents from each stratum to get a sample.

You can, for example, use stratified random sampling to get an equal representation of males and females. You do this by dividing the entire group into subgroups of males and females and then randomly choosing a given number of respondents from each subgroup. This method of sampling can be more precise than simple random sampling because it homogenizes the groups, but only if you choose the strata properly. That is, do not sample men and women unless you are planning to make comparisons between them. You should plan to make comparisons only if you have some reasons to believe, in advance, that those comparisons might be meaningful. In a survey of voter preference, for example, if you have some evidence that men and women vote differently, then it makes sense to be sure that your survey includes enough males and females to compare them. With random sampling alone, you might find that by chance you have a survey sample that consists mainly of men or mainly of women.

Here is how stratified random sampling works.

Example: Stratified Random Sampling

The University Health Center is considering the adoption of a new program to help young adults lose weight. Before changing programs, the administration commissioned a survey to find out, among other things, how their new program compared with the current one and how male and female students of different ages performed. Previous experience had suggested that older students appeared to do better in weight reduction programs. Therefore, the surveyors planned to get a sample of men and women in two age groups—17 to 22 years and 23 to 28 years—and compare their performance in each of the programs.

About 310 undergraduates signed up for the health center's regular weight reduction program for the winter seminar. Of the 310 participants, 140 were between 17 and 22 years old, and 62 of these were men. Some 170 students were between 23 and 28 years, and 80 of these were men. The surveyors randomly selected 40 persons from each of the four subgroups (male, female, ages 17–22, and ages 23–28) and randomly assigned every other student to the new program. The sample looked like this:

University Health Center's Weight-Loss Program

	Ages 17–22 Years		*Ages 23–28 Years*		
	Male	*Female*	*Male*	*Female*	*Total*
Regular program	20	20	20	20	80
New program	20	20	20	20	80
Totals	40	40	40	40	160

Making the Decision

An advantage of stratified random sampling is that the surveyor can choose a sample that represents the various groups and patterns of characteristics in the desired proportions.

The disadvantages of stratified random sampling are that it requires more effort than simple random sampling, and it often needs a larger sample size than does a random sample to produce statistically meaningful results. Remember, for each stratum or subgroup, you probably need 20 to 30 persons to make meaningful statistical comparisons.

If you have difficulty selecting a stratified random sample, keep in mind that the same increase in precision obtained with stratification can generally be produced by increasing the sample size of a simple random sample. Increasing sample size may be easier than implementing a stratified random sample.

Simple Random Cluster Sampling

Simple random cluster sampling is used primarily for administrative convenience, not to improve sampling precision. Sometimes random selection of individuals simply cannot be used. For instance, it would interrupt every hospital ward to choose just a few patients from each ward for a survey. Sometimes random selection of individuals can be administratively impossible.

One solution to the problem of using individuals as a sampling unit is to use groups or clusters of respondents.

In simple random sampling, you randomly select a subset of respondents from all possible individuals who might take part in a survey. Cluster sampling is analogous to random sampling, except that groups rather than individuals are assigned randomly. This method presupposes that the population is organized into natural or predefined clusters or groups. Here is how it works.

Example: Simple Random Cluster Sampling

The Community Mental Health Center has 40 separate family counseling groups, each with about 30 participants. The center's director noticed a decline in attendance in the last year and decided to try out an experimental program in which each individual would be tested and interviewed separately before beginning therapy. The program was expensive, and the center could only afford to finance a 150-person program at first.

Randomly selecting individuals from all group members would have created friction and disturbed the integrity of some of the groups. Instead, a simple random cluster sampling plan was used, in which 5 of the 30-member groups—150 people all together—were randomly selected to take part in the experimental program. Each group was treated as a cluster. At the end of the 6 months, the progress of the experimental program was compared with that of the traditional one.

Making the Decision

The advantages of simple random cluster sampling are these:

- It can be used when selecting individuals randomly is inconvenient or unethical.
- It simplifies survey administration.

The disadvantage of simple random cluster sampling is that it requires complex statistical methods to reconcile sampling units (the hospital, street, school) and analytic units (patients, homeowners, students). That is, you are sampling by cluster (such as schools), but you are analyzing data from individuals (such as students).

Although in the example you have 150 people in the survey, you really have just 5 units (the 5 groups of 30 persons each) to study. Why can't you study each of the 150 persons individually? When people are in special groups—classes, clubs, organizations, neighborhoods—they tend to share similar characteristics and views. Studying each individual may be redundant because one person may be similar to the next. You need a relatively large number of people for each cluster for this problem to become less noticeable. Just how large a number you need is a statistical issue.

Complex sampling strategies require an understanding of sampling statistics.

Systematic Sampling

In systematic sampling, you pick a number, say 5, and select every 5th name on a list that represents the population. If a list contains 10,000 names and the surveyor wants a sample of 1,000, he or she must select every 10th name for the sample.

Suppose you have a list of 500 names from which you want to select 100 people. You can randomly select a number between 1 and 10. If you chose the number 3, you would begin with the 3rd name on the list and count every 5th name after that. Your sample selection will result in the 3rd name, 8th, 13th, and so on until you had 100 names. If you select the "start" number at random, systematic sampling resembles simple random sampling.

Making the Decision

There is a danger in systematic sampling. Lists of people are sometimes arranged so that certain patterns can be uncovered; if you use one of these lists, your sample will be subject to a bias imposed by the pattern. For instance, relatively few people in the United States have last names beginning with W, X, Y, and Z, and they may be underrepresented in a systematic sample. Here is another example. Suppose you are sampling classrooms so that you can survey students to find out about their attitudes toward school. Say also that the classrooms are arranged in this order:

Floor 1	1a	1b	1c
Floor 2	2a	2b	2c
	.	.	.
	.	.	.
	.	.	.
Floor *N*	*N*a	*N*b	*N*c

Suppose further that you select every third class starting with 1a. The sample will consist of Classrooms 1a, 4a, 7a, and so on, to *N*a. The survey of attitudes toward school can be biased if each "a" corresponds to a location within the school that faces the lawn and is quiet and the "b" and "c" classrooms face the sports arena and are noisy.

In considering the use of systematic sampling, carefully examine the list of potential respondents first. If you suspect bias because of the order or placement of sampling units (people, classrooms), use another sampling method.

Convenience Samples

A convenience sample is one that you get because people who are willing to complete the survey are also available when you need them. Say you want to find out whether the student health service was any good. You plan to interview 50 students. If you stand near the clinic entrance during the day, you can recruit each person walking in. When you have a complete set of 50 interviews, you have a convenience sample. Here are several sources of bias in this sample:

- Students who are willing to be interviewed may be more concerned with the health service than those who refuse.
- Students who use the service at the time of your interview may be going for convenience; sicker students may use the service at night.
- Students who talk to you may have a gripe and want to complain.
- Students who talk to you may be the most satisfied and want to brag.
- Students may want to talk but may have no time at the moment; these may be working students. Perhaps working students are different from other students (older?) in their needs and views.

Making the Decision

Because of bias, convenience samples are unconvincing unless you prove otherwise. Here's how you might improve on the credibility of your convenience sample:

- Ask refusers and participants how concerned they are with their health and compare the responses. You may find no differences; if so, then your convenience sample's findings are supported.
- Visit the clinic at night to find out whether students using the health service then are different from the day students in their health status. You may find no differences. Again, if so, then your sample is supported.
- Ask students if they have a gripe.
- Ask students about their satisfaction.
- Ask students why they refuse to participate. Is it because they presently do not have the time?
- Compare students who participate with those who do not in terms of age, gender, and so on. If no differences are found, your convenience sample's findings are easier to support than otherwise. If differences are found, you need to rethink your survey methodology.

Other Convenience Sampling Methods

Consider these common situations.

Example: Other Convenience Samples

1. A survey of 100 deans of law schools, senior partners in large law firms, and judges is conducted to find out to which lawyers they go to solve their own legal problems. The results are published in *Global News and World Reports* under the title "The World's Best Lawyers."
2. What makes college students liberal or conservative? Family background? Region of the country in which they were born? Current religious practices? Educational attainment? Income? A survey is conducted of members of the Young Conservative Association and the Young Liberal Society. An assessment of the results reveals the reasons for students' views.
3. A popular movie downloading service has substantially raised its monthly fee. How will the increased price affect subscribers? Will most remain with the service or go somewhere else? A total of 80,000 people are contacted to complete an online survey. All potential respondents have participated previously in surveys conducted by SurveyPanels, an Internet survey company.

■

In the first example, the "top" lawyers may provide the services needed by deans, senior law firm partners, and judges. How applicable are the top lawyers' services to typical legal problems? In the second example, students in only two organizations are being surveyed. Can we trust that they represent all students, including those who belong to other groups or choose not to join any? In the third example, the survey is sent to 80,000 people who have participated in other online surveys conducted by SurveyPanels. All respondents are willing to complete an Internet survey. Do they represent the average user of the film downloading service?

You use convenience samples because they are available, and access saves the time and expertise needed to develop a sampling plan and implement it. Convenience samples are useful if the surveyor is certain that the responses apply to the target population. They become problematic if the responses come from people who differ substantially in important ways (e.g., age, gender, Internet savvy) from the target population.

These are standard convenience sampling techniques:

Snowball samples. Previously identified members of a group identify other members. For example, you select CEOs and ask them to nominate others.

Quota samples. The group is divided into subgroups in specific proportions. This is similar to stratified sampling.

Focus groups. Ten to 20 people are brought together to answer specific questions. A trained leader conducts the sessions. A transcriber is usually present to summarize the discussion.

Panels. People are recruited to participate in surveys on an as-needed basis. Many online survey companies recruit panels of respondents from the millions of people who participate in their surveys. Large, reputable companies screen respondents so that they have access to a diverse group of people, say, for instance, savvy smartphone users, working moms, video game players, or a random selection of U.S. adults. Usually, the surveyor designs the questionnaire and specifies who should be included in the survey. The survey company does the rest. Payment is by response. For instance, if the surveyor needs 300 responses, and the survey company charges $3.00 for each response, then the cost of the survey is $900. It usually takes about a week to get a complete set of responses. To encourage responses, survey companies usually provide incentives to the panel members. For example, one popular online company donates $0.50 to a charity chosen by the individual panelist and also gives him or her a chance to win a $100 gift card. The company does not provide direct payment or credits so as to discourage survey takers interested only in collecting rewards.

Finding the Sample: Who Is In? Who Is Out?

How do you find a sample? Before you look, you must decide who should be included (and excluded) from the survey. Suppose you want to evaluate the effectiveness of COMPULEARN, an online program to keep employees up to date on how to market products electronically. You decide to survey a sample of users and set these standards for respondents to be included in the survey:

- Participate in the program for 1 month or more
- Be willing to learn to use FREENet software (about 30 minutes)
- Be 30 years of age or older
- Read English

These eligibility criteria help you narrow your sample. But they do something else as well: They restrict your findings just to other people who also meet the standards. Here's why. After your survey is complete, you will be able to tell about the effectiveness of COMPULEARN only for persons who have been in the program for 1 month or more, who can use the specified software, and who are 30 years of age or older and English speaking. A person who is 29 and meets all other standards is, by definition, ineligible, and you will not be able to apply the survey's findings to 29-year-olds. (After all, you chose 30 years of age as an eligibility criterion because it presumably meant a critical demarcation point.)

Every time you do a survey, you must decide this: Do I want to sample everyone or just people or places with certain characteristics? Usually the answer is "just people or places with certain characteristics (such as schools with 350 or more students)."

How Large Should Your Sample Be?

Some surveys take place with just one group. A poll of a sample of voters is this type of survey sample. The trick is to select a sample that is "representative" of all voters who interest you. Other surveys are of two or more groups. For example, a survey comparing the career plans of students in the JOBS program with students in the CAREER program is this second sample. When comparing students' career plans, you may want representative samples of students in both groups. In addition, you have to think about the number of students you need in each of the two groups so that if a difference exists between them you have enough "power" to detect it.

Consider these examples.

1. *One group, no intervention or program:* You work for the park system and want to find out where young people do their mountain cycling.

   Objective: To survey young mountain bicyclists' preferences

   Population: Mountain bicyclists 21 years and younger

   Sampling question: How many mountain bicyclists should be surveyed to make sure that the sample is a fair representation of mountain bike riders 21 years and younger?

2. *Two or more groups and an intervention:* You have two groups. You want to compare them for differences after one of the two has been part of a new activity.

   Objective: To evaluate a chess program

   Population: Children who are in an experimental chess program and a control group of children who are not

   Sampling question: How many children have to be in each group to detect a positive difference if one occurs?

When you think about sample size, you must also think about the standard error, a statistic used to describe sampling errors. Error exists because when you sample, you select from a larger population, and the sample typically differs from the population. This difference is a random result of sampling. You can control it but probably not eliminate it entirely. If you drew an infinite number of samples, the means or averages would form a distribution that clusters around the true population mean. This distribution, which has a bell shape, is the so-called normal distribution. In general, larger samples are more likely to collect around the true population mean and be a more accurate estimation of the population mean.

The Standard Error

Larger samples tend to reduce sampling errors when the samples are randomly selected. The statistic used to describe sampling error is called the *standard error of the mean.* It is the standard deviation of the distribution of sample estimates of means that could be formed if an infinite number of samples of a given size were drawn.

Try this: In a survey of 300 randomly selected respondents, 70% answer yes to the question, "Is mountain biking one of your favorite sports?" The surveyor reports that the sampling error is 9.2 percentage points. If you add 9.2 to 70%, or subtract 9.2 from 70%, you get an interval—called a *confidence interval*—that ranges between 60.8% and 79.2%. Using standard statistical methods, this means that the surveyor can estimate with 95% confidence (the 95% is a convention) that the "true" proportion of mountain bicyclers who answer yes falls within the interval.

The trick is keeping the confidence interval small. In practice, larger samples usually reduce sampling errors in random samples. But adding to the sample reduces the error a great deal more when the sample is small than when it is large. Also, different sampling methods (such as systematic sampling) may have different error rates from random sampling.

Remember that not all errors come from sampling. Although you want a large enough sample to keep the error low, you do not want sample size pressures to distract you so that other sources of error blight your survey. Other sources of error include ambiguous eligibility criteria, badly designed and administered surveys, and poor returns.

Statistical Methods: Sampling for Two Groups and an Intervention

Suppose you want to compare two groups. First, divide the population in two (say, randomly). Then, use statistical calculations (such as the ones that follow) to find out whether each group's sample size is large enough to pick up a difference, if one is present. If the population is large, you may want to select a sample (say, at random) and then assign persons to the two groups. If you select a sample at random, you have random sampling. If you assign people to groups at random, you have random assignment. If you select all five schools in a city and randomly assign all students in each of the schools to groups, you have nonrandom cluster sampling and random assignment. If, however, you randomly select five schools in a city, assign three to an experiment and two to a control, and put all students in the experimental schools in the experiment, you have random cluster sampling and nonrandom assignment.

Use the following checklist to get or evaluate a sample size when you have two groups and an intervention.

Example: Sample Size Calculations for Sampling Two Groups and an Intervention

Assemble and clarify survey objectives and questions.

Decide the survey's purposes. Consider these:

Survey 1: Quality of Life

Objective: To determine whether younger and older women differ in their quality of life after surgery for breast cancer

Question: Do younger and older women differ in their quality of life after surgery for breast cancer?

Survey 2: Anxiety in School

Objective: To determine the nature and type of anxiety associated with school

Question: Do boys and girls differ in their anxieties? How do younger and older students compare?

Each objective or question contains independent and dependent variables. Independent variables are used to predict or explain the dependent variables. They often consist of the groups (experimental or control, men or women) to which respondents belong or their characteristics (under 50 years old and 51 years of age and older). Take the question, "Do boys and girls differ in their anxieties?" The grouping or independent variable is sex.

The dependent variables are the attitudes, attributes, behaviors, and knowledge the survey is measuring. In statistical terms, they are the variables for which estimates are to be made or inferences drawn. In the question, "Do boys and girls differ in their anxieties?" the dependent variable is anxieties.

Identify subgroups.

The subgroups refer to the groups whose survey results must be obtained in sufficient numbers for accurate conclusions. In the two previous surveys, the subgroups can be identified by looking at the independent variables. Survey 1's subgroups are older and younger women. Survey 2's subgroups are older and younger boys and girls.

Identify survey type and data collection needs.

The dependent variables tell you the content of the survey. For example, Survey 1's specific questions will ask respondents about various aspects of their quality of life. Survey 2's will ask about anxiety. Suppose Survey 1 is a face-to-face interview, and Survey 2 is a self-administered questionnaire.

Check the survey's resources and schedule.

A survey with many subgroups and measures will be more complex and costly than those with few subgroups. Consider this:

Subgroups, Measures, Resources, and Schedule

Survey	*Subgroup*	*Type of Survey*	*Comment*
1: Do younger and older women differ in their quality of life after surgery for breast cancer?	Younger and older women: two subgroups	Face-to-face interview	May need time to hire and train different interviewers for younger and older women
			May have difficulty recruiting sufficient numbers of eligible younger or older women
2: Do boys and girls differ in their anxieties?	Boys and girls, younger and older: four subgroups	Self-administered questionnaire	May need time to translate the questionnaire from English into other languages
How do younger and older students compare?			Must develop methods to ensure confidentiality of responses

The number of subgroups ranges from two to four. Administering, scoring, and interpreting the survey for one group is difficult enough; with more than one, the difficulties mount.

Calculate sample size.

Suppose a survey is concerned with finding out whether a flexible-time work program improves employee satisfaction. Suppose also that one survey objective is to compare the goals and aspirations of employees in the program with other nonparticipating employees. How large should each group of workers be? To answer this question, five other questions must be answered.

Five Questions to Ask When Determining Sample Size

1. What is the null hypothesis?

The null hypothesis (H_o) is a statement that no difference exists between the average or mean scores of two groups. For example, one null hypothesis for the survey of employee satisfaction is this:

H_o = No difference exists between goals and satisfaction (as measured by average survey scores) between employees participating in the program and nonparticipating employees.

2. What is the desired level of significance (α level) related to the null hypothesis involving the mean in the population (μ_o)?

The level of significance, when chosen before the test is performed, is called the *alpha value* (denoted by the Greek letter alpha: α). The alpha gives the probability of rejecting the null hypothesis when it is actually true. Tradition keeps the alpha value small—.05, .01, or .001—to avoid rejecting a null hypothesis when it is true (and no difference exists between group means). The *p* value is the probability that an observed result (or result of a statistical test) is due to chance (rather than to participation in a program). It is calculated *after* the statistical test. If the *p* value is less than alpha, then the null is rejected.

When differences are found to exist between two groups, but in reality there are no differences, that is called an alpha or Type I error. When no differences are found between groups, although in reality there is a difference, that is termed a beta or Type II error.

3. What chance should there be of detecting an actual difference?

Power is the ability to detect a difference of a given size if the difference really exists. It is calculated as $1 - \beta$ (Greek letter beta). It is defined as the probability of rejecting the null hypothesis when it is false or accepting the alternative hypothesis when it is true. You want high power.

4. What differences between means are important? That is, what is a meaningful $\mu_1 - \mu_2$?

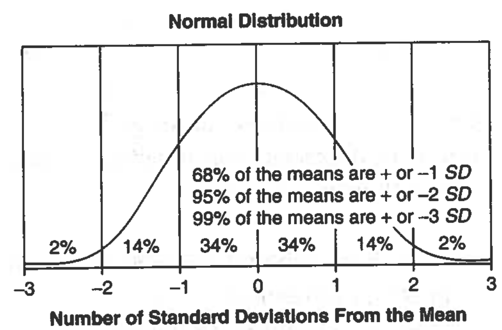
Suppose the survey uses the Goals and Satisfaction Scale (GASS). This hypothetical scale has 50 points. If the scale is valid, you will have access to published scoring rules and data describing what the scores mean. Ask these questions: Are higher scores better? How many "points" make a practical (educational or clinical) difference?

The answers to questions such as these will help you decide how much difference you want to detect in your two groups.

5. What is a good estimate of the standard deviation σ in the population?

The standard deviation (σ, lowercase Greek letter sigma) is a common measure of dispersion or spread of data about the mean.

If the distribution of values or observations is a bell-shaped or normal distribution, then 68% of the observations will fall between the mean ± 1 standard deviation; 95% of the observations between ±2 standard deviations; and 99% of the observations between ±3 standard deviations. Look at this:



Estimates of the standard deviation can come from previously done surveys. Check that the sample used to derive the standard deviation is similar to your own. If it was not, the standard deviation in your group is likely to be different and so is your group.

Another way to estimate the standard deviation in your group is to conduct a small pilot test using about 25 to 50 people. You can also have experts give you estimates on the highest and lowest values or scores as the basis for calculating the standard deviation.

What follows is one formula for calculating sample size for comparing the means from two independent groups (such as two groups of employees). Group 1 is in a program to improve satisfaction, but Group 2 is not. This formula assumes that the two groups' sample sizes and standard deviations are equal:

$$\frac{(z_\alpha - z_\beta)\sigma^2}{\mu_1 - \mu_2}$$

where

$\mu_1 - \mu_2$ is the magnitude of the difference to be detected between the two groups

z_α the upper tail in the normal distribution) and z_β (the lower tail) are defined as

$$z_\alpha \frac{X - \mu_1}{\sigma / \sqrt{n}} \text{ and } z_\beta \frac{X - \mu_2}{\sigma / \sqrt{n}}.$$

Here is an example of how to apply the formula.

Example: Calculating Sample Size in a Survey of Employees in an Experimental and Control Group

Survey situation. Devsoft's employees are participating in a program to improve their job satisfaction. At the conclusion of the 3-year program, participants in the experimental and control groups will be surveyed to find out about their goals and aspirations. The highest possible score on the survey is 100 points. The Type 1 or alpha level is set at .05. The probability of detecting a true difference is set at .80. A panel of experts in satisfaction measures says that the difference in scores between the experimental and control groups should be 10 points or more. Previous employee surveys found a standard deviation of 15 points.

The calculations. For the calculation, assume that a standard normal distribution or *z* distribution is appropriate. The standard normal curve has a mean of 0 and a standard deviation of 1. (For more about the standard normal distribution, one- and two-tailed tests, and *z* values, see Chapter 6. Actual *z* values are obtainable in statistics books and program manuals.)

The two-tailed *z* value related to $\alpha = .05$ is +1.96. For $\alpha = .01$, the two-tailed *z* value is 2.58; for $\alpha = .10$, the *z* value is 1.65; and for $\alpha = .20$, the *z* value is 1.28. The lower one-tailed *z* value related to β is −.84 (the critical value or *z* score separating the lower 20% of the *z* distribution from 80%). Applying the formula, we have the following:

$$(1.96 + 0.84)\,(15)^2 = 2\left(\frac{42}{10}\right)^2$$

$$= 2(17.64) \text{ or about } 36.$$

At least 36 employees are needed in each group to have an 80% chance of detecting a difference in scores of 10 points.

■

Sometimes—for practical reasons—you can assemble only a certain number of persons for your survey. How do you know whether the number is large enough to find differences? Again, statistical methods to the rescue. Look at this.

Example: Power to Detect Differences

The Alcohol Rehabilitation Unit has a program to reduce risks from alcohol use. It sets this standard of success:

> At the end of the program, 20% of the harmful drinkers in the treatment group will reduce their risks, whereas 10% of the harmful drinkers in the control group will do so.

The unit hopes to be able to survey 150 persons in each group regarding their risks. Is this sample large enough? A statistician is called in to answer the question. The statistician produces this table for comparing percentage changes using a two-sided test with $\alpha = .05$.

N or Sample Size	*%1*	*%2*	*Power*
50	20	10	.29
100	20	10	.52
200	20	10	.81
250	20	10	.89
300	20	10	.93

You interpret this table as follows:

> If we have 150 persons (final analytic sample size) in the experimental group and 150 in the control group, we will have 69% power to distinguish a shift of 20% in the experimental group from harmful to less risky from a shift of 10% in the control group.

■

You can use a statistical approach to distinguish different *effect sizes*. The effect size is the difference divided by the standard deviation. Look at this.

Example: 80% Power and Effect

The Alcohol Rehabilitation Unit will have complete survey data on 150 persons in an experimental group and 150 in a control group. These persons will have completed a survey before they participate in the unit and immediately after. How much *power* will this sample size yield? Power is the ability to detect a difference or effect. Put another way, it is the ability of a statistical test to detect an alternative hypothesis of difference (between groups) of a specified size when the alternative is true (and the null is rejected). A statistician provides this table. What does it mean?

N *or Sample Size*	*Effect Size*
50	.56
100	.40
150	.32
200	.28
250	.25
300	.23

You interpret this table as follows:

> If we have 150 (final analytic sample size) in our experimental group and 150 in our control group, we will be able to distinguish a .32 effect size between the difference over time in the experimental group versus the difference over time in the control group.

Because you estimate effect by dividing the mean or average difference in scores by the standard deviation, you can see, for example, whether the experimental group's mean improvement had a standard deviation of .50 and the control group's mean improvement had one of .10. If so, then the effect size would be 50 – 10 = 40 >.32. In this case, you will have at least 80% power to detect the difference of .40 between the treatment and control groups.

The preceding discussion is to help you learn terminology and to aid you in evaluating the usefulness of your sampling methods and outcomes. *Sample sizes*, *effects*, and *power* are statistical terms. Unless you plan to learn statistics, call in an expert to help. Alternatively, consider using the sample-size calculators found in statistical programs or on the web. Go to your favorite search engine and type in "sample size" to find free sample-size calculators.

The following illustrates how to write up and justify sample size, effect, and power.

Example: Sample Size, Effect, and Power

1. Yoga for Chronic Back Pain

   The United Kingdom BEAM (Back Pain Exercise and Manipulation) trial found that a change in the Roland Morris Disability Questionnaire score of 1.57 points was a cost-effective difference. Assuming a standard deviation of 4 points (as in the United Kingdom BEAM sample size), this results in an effect size of 0.39 [1.57 ÷ 4]. To detect this effect size, assuming 80% power and 20% attrition, we required a total of 262 participants (131 per group).

   If you would like to learn more about this study, go to Tilbrook, H. E., Cox, H., Hewitt, C. E., Kang'ombe, A. R., Chuang, L.-H., Jayakody, S., et al. (2011). Yoga for chronic low back pain. *Annals of Internal Medicine*, *155*(9), 569–578.

2. Effectiveness of Telephone Counseling for Smoking Cessation in Parents

   Based on similar studies, we expected a 6% difference in 7-day point prevalence abstinence rates between the telephone counseling condition and the control condition at 12-months assessment (13% versus 7%, respectively). A statistical power of .80 was targeted.

   Hypotheses will be tested at a two-sided significance level of .05. The calculated sample size was corrected for participants who will be lost to attrition. Additionally, the sample size was corrected to allow for supplementary analyses of mediation and moderation.

   If you would like to learn more about this study, go to Schuck, K., Otten, R., Kleinjan, M., Bricker, J., & Engels, R. (2011). Effectiveness of proactive telephone counselling for smoking cessation in parents: Study protocol of a randomized controlled trial. *BMC Public Health*, *11*(1), 732.

Response Rate

The response rate is the number of persons who respond (numerator) divided by the number of eligible respondents (denominator). If 100 people are eligible and 75 completed surveys are available for analysis, the response rate is 75%.

All surveys hope for a high response rate. No single rate is considered the standard, however. In some surveys, between 95% and 100% is expected; in others, 70% is adequate.

Here are some tips to improve the response rate.

Tips for Improving Response Rate

- Know your respondents. Make certain the questions are understandable to them, to the point, and not insensitive to their sociocultural values.
- Use trained personnel to recruit respondents and conduct surveys. Set up a quality assurance system for monitoring quality and retraining.
- Identify a larger number of eligible respondents than you need in case you do not get the sample size desired. Be careful to pay attention to the costs.
- Use surveys only when you are fairly certain that respondents are interested in the topic.
- Keep survey responses confidential or anonymous.
- Send reminders to nonresponders.
- Provide options for completing the survey. For instance, some people prefer online surveys, whereas others prefer paper-and-pencil surveys that they complete and return by mail or fax.
- Provide gift or cash incentives.
- Be realistic about the eligibility criteria.
- Anticipate in advance the proportion of respondents who may not be able to participate because of survey circumstances (such as incorrect addresses) or by chance (they suddenly get ill).
- Tell respondents how you will use the survey's data.
- Be precise in describing how privacy is safeguarded.

Weighting

Surveys rarely obtain information from everyone within the population of interest. To make up for the discrepancy, and to ensure that the survey results are representative of the population, surveyors often assign weights or values to each case. Suppose you survey 100 people who are between 15 to 64 years of age and you obtain responses from 60 females and 40 males. For that age range, you should have responses from 50 males and 50 females. To correct for this difference, you assign a weight to each male's answers so that they contribute more (more weight) than do each female's.

The two most common types of weights are design weights and poststratification or nonresponse weights. Design weights are used to compensate for oversampling or undersampling of specific cases or for disproportionate stratification. Say you want to compare 50 members of Group A and 50 members of Group B. You learn that Group A has many people who often refuse to take surveys. You decide to sample more people in Group A (oversample) than in Group B, hoping that even with a relatively large refusal rate, you will end up with equal numbers of completed surveys in both groups. If you oversampled Group A at a rate 4 times greater than the rate for Group B, than the design weight for Group A is ¼, and for Group B, it is 1.

Poststratification or nonresponse weights are also used to compensate for the fact that persons with certain characteristics are not equally as likely to respond. Most general-population surveys have substantially more female than male respondents (often 60/40), although there are often more males in the population. Because the survey will overrepresent females and underrepresent males, you assign a weight to make up for this bias. There are many respondent characteristics that are likely to be related to the tendency to respond, including age, education, race/ethnicity, sex, and place of residence. It is usually more difficult to come up with poststratification than design weights because you need supplementary information about the population. This means identifying and analyzing very large databases (such as a country's census or a university's enrollment records) to get estimates of the demographic variables that interest you. These databases provide answers to questions like the following: What is the actual proportion of men and women in the population? What is the prevalence of this disease in younger versus older people? What percentage of people live in single-family homes compared to other living arrangements?

■

Here is a hypothetical example of how poststratification weights are calculated:

Gender	*Population Proportion*[x]	*Sample Proportion*[xx]	*Population/ Sample*	*Weight*[xxx]
Female	.5	.6	.5/.6	**.8333**
Male	.5	.4	.5/.4	**1.25**
Total	1	1		

x The census provides information on the proportion of males and females in the population. It is equally distributed: 50% male and 50% female.

xx The survey's respondents are 60% female.

xxx The obtained proportions are different from the population's proportions, with a greater proportion of female respondents. Males are assigned a greater weight to make up for the difference.

SUMMING UP

Three of the most commonly used random sampling methods are the following:

1. Simple random sampling
2. Stratified random sampling
3. Simple random cluster sampling

Random sampling gives everyone who is eligible to participate in the survey a fair chance of selection.

You can sample individuals or larger units such as schools, offices, and hospitals. These larger sampling units contain clusters (of students, employees, nurses, physicians, patients), so the technique is called cluster sampling.

Nonrandom samples are often easier to assemble than random samples. But gains in ease can be met with losses in generalizability.

With systematic samples, every *n*th (e.g., every 5th or 500th) unit (individuals, schools, factories) is selected from a list of units. If *n* is randomly selected, systematic sampling becomes like random sampling.

In convenience sampling, you select everyone who is available when you need them if they meet the criteria for your survey (right age or reading level, voted in the last election, have lived in the community for at least 1 year, etc.). Other nonprobability sampling methods include snowball and quota sampling.

Other nonrandom sampling strategies include the following:

Snowball samples. Previously identified members of a group identify other members. For example, you select CEOs and ask them to nominate others.

Quota samples. The group is divided into subgroups in specific proportions. This is similar to stratified sampling.

Focus groups. Ten to 20 people are brought together to answer specific questions. A trained leader conducts the sessions. A transcriber is usually present to summarize the discussion.

Panels. Respondents agree to complete surveys if they are contacted. The respondents are prescreened for their demographic and other characteristics (e.g., frequent Internet shoppers, people who never use Twitter).

Larger samples tend to reduce sampling errors when the samples are randomly selected. The statistic used to describe sampling error is called the standard error of the mean. It is the standard deviation of the distribution of sample estimates of means that could be formed if an infinite number of samples of a given size were drawn.

Surveys rarely obtain information from everyone within the population of interest. To make up for loss, and to ensure that the survey results are representative of the population, surveyors assign weights or values to each case. There are two types: design and poststratification, or nonresponse, weights.

THINK ABOUT THIS

1. Locate the surveys in the articles below and then do the following:
   A. Discuss the type of sampling that the surveyors use.
   B. Describe the survey's eligibility criteria by listing the characteristics of respondents who are included and excluded.
2. Find any three surveys that are part of larger studies and discuss their sampling plans. Are the sampling plans discussed in adequate detail? That is, do you think you have enough information about the sampling strategy and sampling size to redo the survey if necessary? What additional information would you need?
3. List at least five tips for improving a survey's response rate.
4. Explain the following terms as they are used when talking about surveys: *standard error, confidence interval, power, effect size, level of significance,* and *normal distribution.*

ARTICLES

Holman, E. A., Silver, R. C., Poulin, M., Andersen, J., Gil-Rivas, V., & McIntosh, D. N. (2008). Terrorism, acute stress, and cardiovascular health: A 3-year national study following the September 11th attacks. *Archives of General Psychiatry, 65*(1), 73–80.

O'Connor, T. M., Yang, S.-J., & Nicklas, T. A. (2006). Beverage intake among preschool children and its effect on weight status. *Pediatrics, 118*(4), 2005–2348.

Schuck, K., Otten, R., Kleinjan, M., Bricker, J., & Engels, R. (2011). Effectiveness of proactive telephone counselling for smoking cessation in parents: Study protocol of a randomized controlled trial. *BMC Public Health, 11*(1), 732.

Skoffer, B. (2007). Low back pain in 15- to 16-year-old children in relation to school furniture and carrying of the school bag. *Spine, 32*(24), E713–E717.

Tilbrook, H. E., Cox, H., Hewitt, C. E., Kang'ombe, A. R., Chuang, L.-H., Jayakody, S., et al. (2011). Yoga for chronic low back pain. *Annals of Internal Medicine, 155*(9), 569–578.

Wehr-Flowers, E. (2006). Differences between male and female students' confidence, anxiety, and attitude toward learning jazz improvisation. *Journal of Research in Music Education, 54*(4), 337–349.

5 SURVEY DESIGN

Environmental Control

OVERVIEW

A survey can be given to just one group or to many groups of people once or several times. For example, you can survey preschool children in one school just before and immediately after they participate in a new music program, or you can survey preschool children in 10 schools just before they participate in the program and every 2 years after until they complete high school. A survey's design refers to its frequency (one time or more often), sequence (just before and immediately after a new program), and the number of groups involved (all students in one school district or all students in each of 10 districts).

A cross-sectional design provides a portrait of things as they are at a single point in time. A poll of voters' preferences 1 month before an election and a survey of the income and age of people who vote in a particular election both use cross-sectional designs.

Longitudinal surveys are used to find out about change. If you take a sample of children who were sixth graders in 2012 and follow them with a survey every year for the next 5 years, you are using a longitudinal design called a cohort design.

Other study designs include comparisons. If respondents are assigned to groups that are randomly constituted, you have a true experiment. If you use a table of random numbers or a statistical algorithm to assign 250 workers to an experimental program and 250 to a comparison and survey their satisfaction, you have the makings of a true experiment. True experiments are also referred to as randomized controlled trials. These designs may be able to establish that Program A caused Outcome A, especially if the survey groups are randomly selected and assigned. The ideal is to "blind" all participants so that they are unaware of whether they are in the experimental or control group, which is easier said than done.

Quasi-experimental designs are comparison group designs in which the groups are constituted by volunteers or convenience. Suppose 10 classes agree to participate in a research study investigating an innovative approach to reading. If you compare these 10 against another 10 classes that continue with their regular reading program, you have a quasi-experimental design. With these designs, the surveyor must be on the lookout for groups that differ in their motivation and education to begin with.

A case control design is one in which groups of individuals are chosen because they have (the case) or do not have (the control) the condition being studied, and the groups are compared with respect to existing or past attitudes, habits, beliefs, or demographic factors judged to be of relevance to the causes of the condition. You have a case control design, for example, when you want to compare people who read a book a week or more with those who do not. Do they differ in education? Early childhood exposure to books in the home? Age?