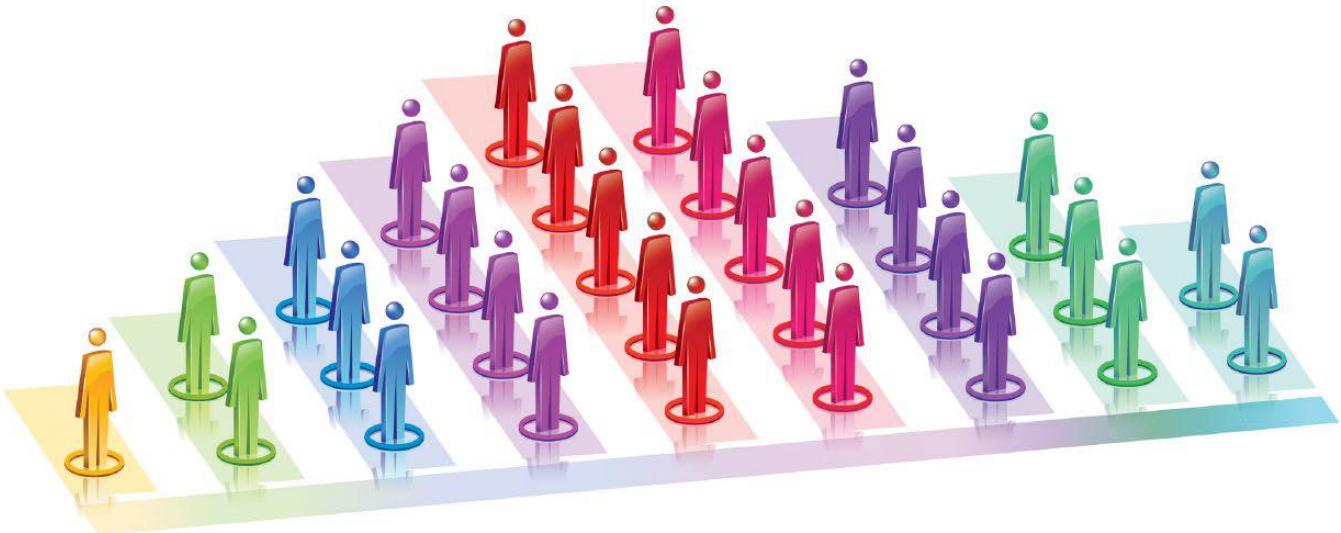
Second Edition

Introductory STATISTICS Using SPSS



HERSCHEL KNAPP

Introductory STATISTICS Using SPSS®

Second Edition

For Mildred & Helen

SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish over 900 journals, including those of more than 400 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, and video. SAGE remains majority-owned by our founder, and after Sara's lifetime will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

Introductory STATISTICS Using SPSS®

Second Edition

HERSCHEL KNAPP

University of Southern California

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

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: Helen Salmon
eLearning Editor: Katie Ancheta
Editorial Assistant: Chelsea Pearson
Production Editor: Libby Larson
Copy Editor: Jim Kelly
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Alison Syring
Indexer: Maria Sosnowski
Marketing Manager: Susannah Goldes

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Names: Knapp, Herschel, author.

Title: Introductory statistics using SPSS / Herschel Knapp.

Description: Second edition. | Thousand Oaks, California : SAGE, [2017] | Includes index.

Identifiers: LCCN 2016022121 | ISBN 978-1-5063-4100-2 (pbk. : alk. paper)

Subjects: LCSH: SPSS for Windows. | Social sciences—Statistical methods—Computer programs.

Classification: LCC HA32 .K59 2016 | DDC 005.5/5—dc23 LC record available at https://lccn.loc.gov/2016022121

Brief Contents

Detailed Contents

PART II: STATISTICAL PROCESSES

Preface

Somewhere, something incredible is waiting to be known.

—Carl Sagan

DOWNLOADABLE DIGITAL LEARNING RESOURCES

Download (and unzip) the digital learning resources for this book from the website **study.sagepub.com/knappstats2e**. This website contains tutorial videos, prepared data sets, and the solutions to all of the odd-numbered exercises. These resources will be discussed in further detail toward the end of the Preface.

OVERVIEW OF THE BOOK

This book covers the statistical functions most frequently used in scientific publications. This should not be considered a complete compendium of useful statistics, however. In other technological fields that you are likely already familiar with (e.g., word processing, spreadsheet calculations, presentation software), you have probably discovered that the "90/10 rule" applies: You can get 90% of your work done using only 10% of the functions available. For example, if you were to thoroughly explore each submenu of your word processor, you would likely discover more than 100 functions and options; however, in terms of actual productivity, 90% of the time, you are probably using only about 10% of them to get all of your work done (e.g., load, save, copy, delete, paste, font, tab, center, print, spell-check). Back to statistics: If you can master the statistical processes contained in this text, it is expected that this will arm you with what you need to effectively analyze the majority of your own data and confidently interpret the statistical publications of others.

This book is not about abstract statistical theory or the derivation or memorization of statistical formulas; rather, it is about *applied* statistics. This book is designed to provide you with practical answers to the following questions: (a) *What statistical test should I use for this kind of data?* (b) *How do I set up the data?* (c) *What parameters should I specify when ordering the test?* and (d) *How do I interpret the results?*

In terms of performing the actual statistical calculations, we will be using IBM® SPSS® *Statistics, an efficient statistical processing software package. This facilitates

*SPSS is a registered trademark of International Business Machines Corporation.

speed and accuracy when it comes to producing quality statistical results in the form of tables and graphs, but SPSS is not an automatic program. In the same way that your word processor does not write your papers for you, SPSS does not know what you want done with your data until you tell it. Fortunately, those instructions are issued through clear menus. Your job will be to learn what statistical procedure suits which circumstance, to configure the data properly, to order the appropriate tests, and to mindfully interpret the output reports.

The 10 chapters are grouped into three parts:

Part I: Statistical Principles

This set of chapters provides the basis for working in statistics.

Chapter 1: Research Principles focuses on foundational statistical concepts, delineating what statistics are, what they do, and what they do not do.

Chapter 2: Sampling identifies the rationale and methods for gathering a relatively small bundle of data to better comprehend a larger population or a specialized subpopulation.

Chapter 3: Working in SPSS orients you to the SPSS (also known as PASW, or Predictive Analytics Software) environment, so that you can competently load existing data sets or configure it to contain a new data set.

Part II: Statistical Processes

These chapters contain the actual statistical procedures used to analyze data.

Chapter 4: Descriptive Statistics provides guidance on comprehending the values contained in continuous and categorical variables.

Chapter 5: *t* Test and Mann-Whitney *U* Test: The ***t* test** is used in two-group designs (e.g., treatment vs. control) to detect if one group significantly outperformed the other. In the event that the data are not fully suitable to run a *t* test, the **Mann-Whitney *U* test** provides an alternative.

Chapter 6: ANOVA and Kruskal-Wallis Test: Analysis of Variance (ANOVA) is similar to the *t* test, but it is capable of processing more than two groups. In the event that the data are not fully suitable to run an ANOVA, the **Kruskal-Wallis test** provides an alternative.

Chapter 7: Paired *t* Test and Wilcoxon Test: The **paired *t* test** is generally used to gather data on a variable before and after an intervention to determine if performance on the posttest is significantly better than that on the pretest. In the event that the data are not fully suitable to run a paired *t* test, the **Wilcoxon test** provides an alternative.

Chapter 8: Correlation and Regression—Pearson and Spearman uses the **Pearson** statistic to assess the relationship between two continuous variables. In the event that the data are not fully suitable to run a Pearson analysis, the **Spearman** test provides an alternative. The Spearman statistic can also be used to assess the relationship between two ordered lists.

Chapter 9: Chi-Square assesses the relationship between categorical variables.

Part III: Data Handling

This chapter demonstrates supplemental techniques in SPSS to enhance your capabilities, versatility, and data processing efficiency.

Chapter 10: Supplemental SPSS Operations explains how to generate random numbers, sort and select cases, recode variables, import non-SPSS data, and practice appropriate data storage protocols.

After you have completed Chapters 4 through 9, the following table will help you navigate this book to efficiently select the statistical test(s) best suited to your (data) situation. For now, it is advised that you skip this table, as it contains statistical terminology that will be covered thoroughly in the chapters that follow.

Overview of Statistical Functions

Chapter	Statistics	When to Use	Results
4	Descriptive statistics	Any continuous or categorical variable	Generates a summary of a variable using figures and graphs
5	*t* test and Mann-Whitney *U* test	Two groups with continuous variables	Indicates if there is a statistically significant difference between the two groups (G_1 : G_2)
6	ANOVA and Kruskal-Wallis test	Similar to the *t* test, except that these tests can process three or more groups	Similar to the *t* test, except all pairs of groups are compared (G_1 : G_2, G_1 : G_3, G_2 : G_3)
7	Paired *t* test and Wilcoxon test	Compares pretest with posttest (continuous variables within one group)	Indicates if there is a statistically significant difference between the pretest and posttest (Pre : Post)
8	Correlation and regression: Pearson and Spearman	Two continuous variables for each subject/record or two ranked lists	Indicates the strength and direction of the relationship between two variables
9	Chi-square	Two categorical variables	Indicates if there is a statistically significant difference between two categories

Parametric Versus Nonparametric (Pronounced *pair-uh-metric*)

In the prior table ("Overview of Statistical Functions"), you may have noticed that Chapters 5 through 8 each contain two statistical tests.

Chapter	Parametric	Nonparametric
5. *t* Test and Mann-Whitney *U* Test	*t* test	Mann-Whitney *U* test
6. ANOVA and Kruskal-Wallis Test	ANOVA	Kruskal-Wallis test
7. Paired *t* Test and Wilcoxon Test	Paired *t* test	Wilcoxon test
8. Regression and Correlation—Pearson and Spearman	Pearson regression	Spearman correlation

The first (*parametric*) statistical test is used when the data are *normally distributed*, meaning that the variable(s) being processed contain *some* very low values and *some* very high values, but *most* of the data land somewhere in the middle—in most instances, data are arranged in this fashion. In cases where one or more of the variables involved are not normally distributed, or other pretest criteria are not met, the second (*nonparametric*) statistic test is the better choice.

The procedure for determining if a variable contains data that are normally distributed is covered thoroughly in Chapter 4 ("Descriptive Statistics").

LAYERED LEARNING

This book is arranged in a progressive fashion, with each concept building on the previous material. As discussed, Chapters 5, 6, 7, and 8 contain two statistics each: The first (parametric) statistic is explained and demonstrated thoroughly, followed by the second (nonparametric) version of the statistic, so that after comprehending the first statistic, the second is only a short step forward; it should not feel like a double workload.

Additionally, Chapter 5 provides the conceptual basis for Chapter 6. Specifically, Chapter 5 ("*t* Test and Mann-Whitney *U* Test") shows how to process a two-group design (e.g., Treatment : Control) to determine if one group outperformed the other. Chapter 6 builds on that concept, but instead of comparing just two groups with each other (e.g., Treatment : Control), the ANOVA and the Kruskal-Wallis tests can compare three or more groups with each other (e.g., Treatment_1 : Treatment_2 : Control) to determine which group(s) outperformed which. Essentially, this is just one step up from what you will already understand from having mastered the *t* test and Mann-Whitney *U* test in Chapter 5, so the learning curve is not as steep.

The point is, you will not be starting from square one as you enter Chapter 6; you will see that you are already more than halfway there to understanding the new

statistics, based on your comprehension of the prior chapter. This form of *layered learning* is akin to simply adding one more layer to an already existing cake, hence the layer cake icon.

DOWNLOADABLE LEARNING RESOURCES

The exercises in Chapter 3 ("Working in SPSS") include the data definitions (codebooks) and corresponding concise data sets printed in the text for manual entry; this will enable you to learn how to set up SPSS from the ground up. This is an essential skill for conducting original research.

Chapters 4 through 10 teach each statistical process using an appropriate example and a corresponding data set. The practice exercises at the end of these chapters provide you with the opportunity to master each statistical process by analyzing actual data sets. For convenience and accuracy, these prepared SPSS data sets are available for download.

The website for this book is **study.sagepub.com/knappstats2e**, which contains the fully developed solutions to all of the odd-numbered exercises so that you can self-check the quality of your learning, along with the following resources.

Videos

The (.mp4) videos provide an overview of each statistical process, along with directions for processing the pretest checklist criteria, ordering the statistical test, and interpreting the results.

Data Set

The downloadable files also contains prepared data sets for each example and exercise to facilitate prompt and accurate processing.

The examples and exercises in this text were processed using Version 18 of the software and should be compatible with most other versions.

RESOURCES FOR INSTRUCTORS

Password-protected instructor resources are available on the website for this book at **study.sagepub.com/knappstats2e** and include the following:

- All student resources (listed above)
- Fully developed solutions to all exercises
- Editable PowerPoint presentations for each chapter

MARGIN ICONS

The following icons provide chapter navigation (in this order) in Chapters 4 to 9:

Video†—Tutorial video demonstrating the **Overview**, **Pretest Checklist**, **Test Run**, and **Results**

Overview—Summary of what a statistical test does and when it should be used

Data Set†—Specifies which prepared data set to load

Pretest Checklist—Instructions to check that the data meet the criteria necessary to run a statistical test

Test Run—Procedures and parameters for running a statistical test

Results—Interpreting the output from the **Test Run**

Hypothesis Resolution—Accepting/rejecting hypotheses based on the **Results**

Documenting Results—Write-up based on the **Hypothesis Resolution**

†Go to **study.sagepub.com/knappstats2e** and download the tutorial videos, prepared data sets, and solutions to all of the odd-numbered exercises.

The following icons are used on an as-needed basis:

Reference Point—This point is referenced elsewhere in the text (think of this as a bookmark)

Key Point—Important fact

Layered Learning—Identifies chapters and statistical tests that are conceptually connected

Technical Tip—Helpful data processing technique

Formula—Useful formula that SPSS does not perform but can be easily processed on any calculator

Acknowledgments

SAGE and the author acknowledge and thank the following reviewers, whose feedback contributed to the development of this text:

Mike Duggan – Emerson College

Tina Freiburger – University of Wisconsin–Milwaukee

Lydia Eckstein Jackson – Allegheny College

Javier Lopez-Zetina – California State University, Long Beach

Lina Racicot, EdD – American International College

Linda M. Ritchie – Centenary College

Christopher Salvatore – Montclair State University

Barbara Teater – College of Staten Island, City University of New York

We extend special thanks to Ann Bagchi for her skillful technical proofreading, to better ensure the precision of this text. We also gratefully acknowledge the contribution of Dean Cameron, whose cartoons enliven this book.

About the Author

Herschel Knapp, PhD, MSSW, has more than 25 years of experience as a health science researcher; he has provided project management for innovative interventions designed to improve the quality of patient care via multisite health science implementations. He teaches master's-level courses at the University of Southern California; he has also taught at the University of California, Los Angeles, and California State University, Los Angeles. Dr. Knapp has served as the lead statistician on a longitudinal cancer research project and managed the program evaluation metrics for a multisite nonprofit children's center. His clinical work includes emergency/trauma psychotherapy in hospital settings. Dr. Knapp has developed and implemented innovative telehealth systems, using videoconferencing technology to facilitate optimal health care service delivery to remote patients and to coordinate specialty consultations among health care providers, including interventions to diagnose and treat people with HIV and hepatitis, with special outreach to the homeless. He is currently leading a nursing research mentorship program and providing research and analytic services to promote excellence within a health care system. The author of numerous articles in peer-reviewed health science journals, he is also the author of *Intermediate Statistics Using SPSS* (2018), *Practical Statistics for Nursing Using SPSS* (2017), *Introductory Statistics Using SPSS* (1st ed., 2013), *Therapeutic Communication: Developing Professional Skills* (2nd ed., 2014), and *Introduction to Social Work Practice: A Practical Workbook* (2010).

PART I

Statistical Principles

CHAPTER 1

Research Principles

- Rationale for Statistics
- Research Questions
- Treatment and Control Groups
- Rationale for Random Assignment
- Hypothesis Formulation
- Reading Statistical Outcomes
- Accept/Reject Hypothesis
- Levels of Measure

The scientific mind does not so much provide the right answers as ask the right questions.

—Claude Lévi-Strauss

LEARNING OBJECTIVES

Upon completing this chapter, you will be able to:

Discuss the rationale for using statistics
Identify various forms of research questions
Differentiate between *treatment* and *control* groups
Comprehend the rationale for random assignment
Understand the basis for hypothesis formulation
Understand the fundamentals of reading statistical outcomes
Appropriately accept or reject hypotheses based on statistical outcomes
Understand the four levels of measure
Determine the variable type: *categorical* or *continuous*

OVERVIEW—RESEARCH PRINCIPLES

This chapter introduces statistical concepts that will be used throughout this book. Applying statistics involves more than just processing tables of numbers; it involves being curious and assembling mindful questions in an attempt to better understand what is going on in a setting. As you will see, statistics extends far beyond simple averages and head counts. Just as a toolbox contains a variety of tools to accomplish a variety of diverse tasks (e.g., a screwdriver to place or remove screws, a saw to cut materials), there are a variety of statistical tests, each suited to address a different type of research question.

RATIONALE FOR STATISTICS

While statistics can be used to track the status of an *individual*, answering questions such as *What is my academic score over the course of the term?* or *What is my weight from week to week?*, this book focuses on using statistics to understand the characteristics of *groups* of people.

Descriptive statistics, described in Chapter 4, are used to comprehend one variable at a time, answering questions such as *What is the average age of people in this group?* or *How many females and males are there in this group?* Chapters 5 to 9 cover *inferential* statistics, which enable us to make determinations such as *Which patrolling method is best for reducing crime in this neighborhood? Which teaching method produces the highest test scores? Is Treatment A better than Treatment B for a particular disorder? Is there a relationship between salary and happiness?* and *Are female or male students more likely to graduate?*

Statistics enables professionals to implement evidence-based practice (EBP), meaning that instead of simply taking one's best guess at the optimal choice, one can use statistical results to help inform such decisions. Statistical analyses can aid in (more) objectively determining the most effective patrolling method, the best available teaching method, or the optimal treatment for a specific disease or disorder.

EBP involves researching the (published) statistical findings of others who have explored a field you are interested in pursuing; the statistical results in such reports provide evidence as to the effectiveness of such implementations. For example, suppose a researcher has studied 100 people in a sleep lab and now has statistical evidence showing that people who listened to soothing music at bedtime fell asleep faster than those who took a sleeping pill. Such evidence-based findings have the potential to inform professionals regarding best practices—in this case, how to best advise someone who is having problems falling asleep.

EBP, which is supported by statistical findings, helps reduce the guesswork and paves the way to more successful outcomes with respect to assembling more plausible requests for proposals (RFPs), independent proposals for new implementations, and plans for quality improvement, which could involve modifying or enhancing existing implementations (quality improvement), creating or amending policies, or assembling best-practices guidelines for a variety of professional domains.

Even with good intentions, without EBP, we risk adopting implementations that may have a neutral, suboptimal, or even negative impact, hence failing to serve or possibly harming the targeted population.

Additionally, statistics can be used to evaluate the performance of an existing program, which people may be simply assuming is effective, or statistics can be built in to a proposal for a new implementation as a way of monitoring the performance of the program on a progressive basis. For example, instead of simply launching a new program designed to provide academic assistance to students with learning disabilities, one could use EBP methods to design the program, such that the program that would be launched would be composed of elements that have demonstrated efficacy. Furthermore, instead of just launching the program and hoping for the best, the design could include periodic grade audits, wherein one would gather and statistically review the grades of the participants at the conclusion of each term to determine if the learning assistance program is having the intended impact. Such findings could suggest which part(s) of the program are working as expected, and which require further development.

Consider another concise example, wherein a school has implemented an evidence-based strategy aimed at reducing absenteeism. Without a statistical evaluation, administrators would have no way of knowing if the approach worked or not. Alternatively, statistical analysis might reveal that the intervention has reduced absences except on Fridays—in which case, a supplemental attendance strategy could be considered, overall, or the strategy could be adapted to include some special Friday incentives.

RESEARCH QUESTIONS

A statistician colleague of mine once said, "I want the numbers to tell me a story." Those nine words elegantly describe the mission of statistics. Naturally, the story depends on the nature of the statistical question. Some statistical questions render descriptive (summary) statistics, such as: *How many people visit a public park on weekends? How many cars cross this bridge per day? What is the average age of students at a school? How many accidents have occurred at this intersection? What percentage of people in a geographical region have a particular disease? What is the average income per household in a community? What percentage of students graduate from high school?* Attempting to comprehend such figures simply by inspecting them visually may work for a few dozen numbers, but visual inspection of these figures would not be feasible if there were hundreds or even thousands of numbers to consider. To get a reasonable idea of the nature of these numbers, we can mathematically and graphically summarize them and thereby better understand any amount of figures using a concise set of **descriptive statistics**, as detailed in Chapter 4.

Another form of research question involves comparisons; often this takes the form of an experimental outcome. Some questions may involve comparisons of scores between two groups, such as: *In a fourth grade class, do girls or boys do better on math tests? Do smokers sleep more than nonsmokers? Do students whose parents are teachers have better test scores than students whose parents are not teachers? In a two-group clinical trial, one*

group was given a new drug to lower blood pressure, and the other group was given an existing drug; does the new drug outperform the old drug in lowering blood pressure? These sorts of questions, involving the scores from two groups, are answered using the ***t* test** or the **Mann-Whitney *U* test**, which are covered in Chapter 5.

Research questions and their corresponding designs may involve several groups. For example, in a district with four elementary schools, each uses a different method for teaching spelling; is there a statistically significant difference in spelling scores from one school to another? Another example would be a clinical trial aimed at discovering the optimal dosage of a new sleeping pill. Group 1 gets a placebo, Group 2 gets the drug at a 10-mg dose, and Group 3 gets the drug at a 15-mg dose; is there a statistically significant difference among the groups in terms of number of hours of sleep per night? Questions involving analyzing the scores from more than two groups are processed using **ANOVA (analysis of variance)** or the **Kruskal-Wallis test**, which are covered in Chapter 6.

Some research questions involve assessing the effectiveness of a treatment by administering a pretest, then the treatment, then a posttest to determine if the group's scores improved after the treatment. For example, suppose it is expected that brighter lighting will enhance mood. To test for this, the researcher administers a mood survey under normal lighting to a group, which renders a score (e.g., 0 = *very depressed*, 10 = *very happy*). Next, the lighting is brightened, after which that group is asked to retake the mood test. The question is: *According to the pretest and posttest scores, did the group's mood (score) increase significantly after the lighting was changed?* Consider another example: Suppose it is expected that physical exercise enhances math scores. To test this, a fourth grade teacher administers a multiplication test to each student. Next, the students are taken out to the playground to run to the far fence and back three times, after which the students immediately return to the classroom to take another multiplication test. The question is: *Is there a statistically significant difference between the test scores before and after the physical activity?* Questions involving before-and-after scores within a group are processed with the **paired *t* test** or the **Wilcoxon test**, which are covered in Chapter 7.

Another kind of research question may seek to understand the (co)relation between two variables. For example: *What is the relationship between the number of homework hours per week and grade?* We might expect that as homework hours go up, grades would go up as well. Similarly, we might ask: *What is the relationship between exercise and weight (if exercise goes up, does weight go down)? What is the relationship between mood and hours of sleep per night (when mood is low, do people sleep less)?* Alternatively, we may want to assess how similarly (or dissimilarly) two lists are ordered. Questions involving the correlation between two scores are processed with **correlation** and **regression** using the **Spearman** or **Pearson** test, which are covered in Chapter 8.

Research questions may also involve comparisons between categories. For example: *Is there a difference in ice cream preference (chocolate, strawberry, vanilla) based on gender (male, female)—in other words, does gender have any bearing on ice cream flavor selection?* We could also investigate questions such as: *Does the marital status of parents (divorced, not divorced) have any bearing on their children's graduation from high school*

(graduated, not graduated)? Questions involving comparisons among categories are processed using **chi-square** (*chi* is pronounced *k-eye*), which is covered in Chapter 9.

As you can see, even at this introductory level, a variety of statistical questions can be asked and answered. An important part of knowing which statistical test to reach for involves understanding the nature of the question and the type of data at hand.

TREATMENT AND CONTROL GROUPS

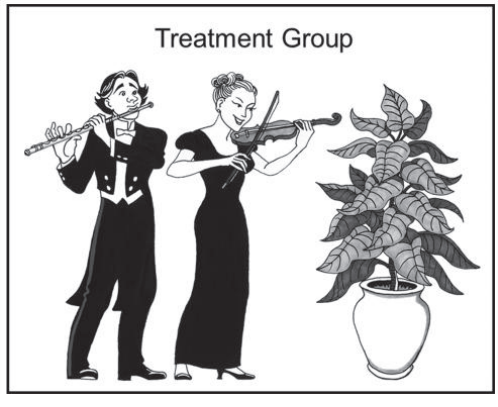
Even if you are new to statistics, you have probably heard of *treatment* and *control* groups. To understand the rationale for using this two-group design, we will explore the results of four examples aimed at answering the research question "Does classical music enhance plant growth?"

Example 1 (Figure 1.1) is a one-group design consisting of a treatment group only (no control group), wherein a good seed is planted in quality soil in an appropriate planter. The plant is given proper watering, sunlight, and 8 hours of classical music per day for 6 months.

At 6 months, the researcher will measure the plant's growth by counting the number of leaves. In this case, the plant produced 20 full-sized healthy leaves, leading the researchers to reason that *classical music facilitates quality plant growth*.

Anyone who is reasonably skeptical might ponder, "The plant had a lot of things going for it—a quality seed, rich soil, the right planter, regular watering and sunlight, and classical music. So, how do we really know that it was the *classical music* that made the plant grow successfully? Maybe it would have done fine without it." Example 2 (Figure 1.2) uses

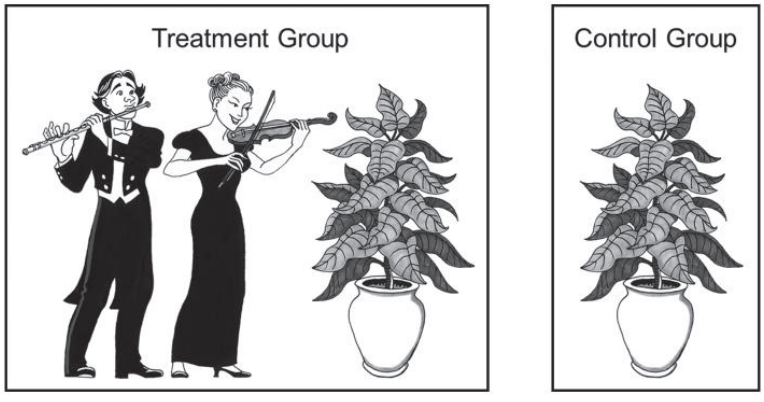
Figure 1.1 One group: treatment group only (positive treatment effect presumed).



a two-group design, consisting of a treatment group and a control group to address that reasonable question.

Notice that in Example 2, the treatment group is precisely the same as in Example 1, which involves a plant grown with a quality seed, rich soil, the right planter, regular watering and sunlight, and classical music. The exact same protocol is given to the other plant, which is placed in the control group, except for one thing: The control plant will receive *no music*. In other words, everything is the same in these two groups except that one plant gets the music and the other does not—this will help us isolate the effect of the music.

Figure 1.2 Two groups: treatment group performs the same as control group (neutral treatment effect).

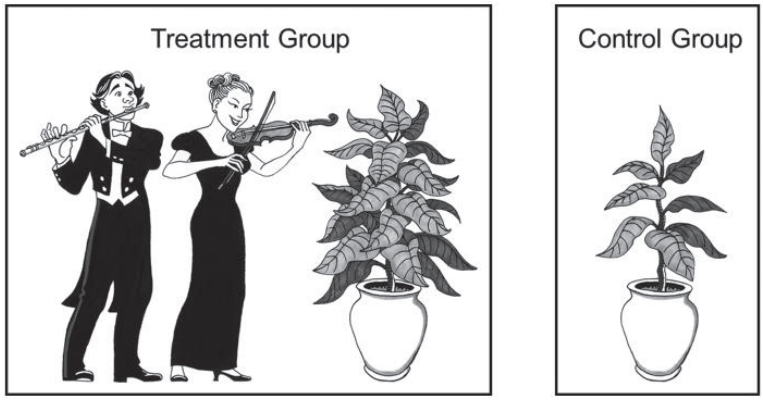


At 6 months, the researcher would then assess the plant growth for each group: In this case, the treatment plant produced 20 leaves, and the control plant also produced 20 leaves. Now we are better positioned to answer the question "How do we really know that it was the *classical music* that made the plant grow successfully?" The control group is the key to answering that question. Both groups were handled identically except for one thing: The treatment group got classical music and the control group did not. Since the control plant received no music but did just as well as the plant that did get the music, we can reasonably conclude that the classical music had a *neutral* effect on the plant growth. Without the control group, we may have mistakenly concluded that classical music had a *positive* effect on plant growth, since the (single) plant did so well in producing 20 leaves.

Next, consider Example 3, which is set up the same as Example 2: a treatment group, in which the plant gets music, and a control group, in which the plant gets no music (Figure 1.3).

In Example 3, we see that the plant in the treatment group produced 20 leaves, whereas the plant in the control group produced only 8 leaves. Since the only difference

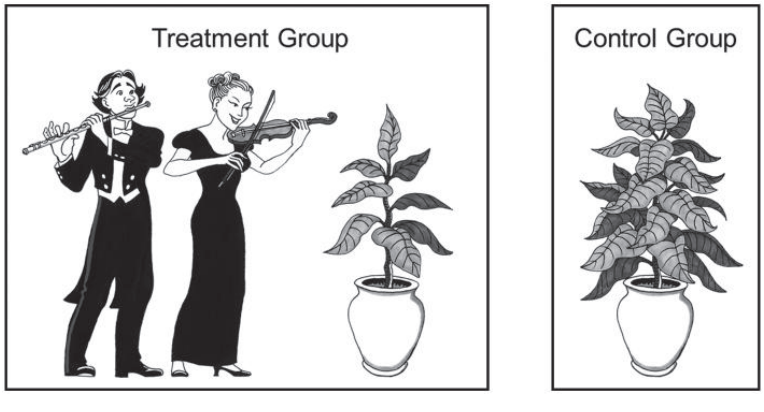
Figure 1.3 Two groups: treatment group outperforms control group (positive treatment effect).



between these two groups is that the treatment group got the music and the control group did not, the results of this experiment suggest that the music had a *positive* effect on plant growth.

Finally, Example 4 (Figure 1.4) shows that the plant in the treatment group produced only 8 leaves, whereas the control plant produced 20 healthy leaves; these results suggest that the classical music had a *negative* effect on plant growth.

Figure 1.4 Two groups: control group outperforms treatment group (negative treatment effect).



Clearly, having the control group provides a comparative basis for more realistically evaluating the outcome of the treatment group. As in this set of examples, in the best circumstances, the treatment group and the control group should begin as identically as possible in every respect, except that the treatment group will get the specified treatment, and the control group proceeds without the treatment. Intuitively, to determine the effectiveness of an intervention, we are looking for substantial differences in the performance between the two groups—is there a significant difference between the results of those in the treatment group compared with the control group?

The statistical tests covered in this text focus on different types of procedures for evaluating the difference(s) between groups (treatment : control) to help determine the effectiveness of the intervention—whether the treatment group significantly outperformed the control group.

To simplify the foregoing examples, the illustrations were drawn with a single plant in each group. If this had been an actual experiment, the design would have been more robust if each group contained multiple plants (e.g., about 30 plants per group); instead of counting the leaves on a single plant, we would compute an average (mean) number of leaves in each group. This would help protect against possible anomalies; for example, the results of a design involving only one plant per group could be compromised if, unknowingly, a good seed were used in one group and a bad seed were used in the other group. Such adverse effects such as this would be minimized if more plants were involved in each group. The rationale and methods for having larger sample sizes (greater than one member per group) are covered in Chapter 2 (“Sampling”).

RATIONALE FOR RANDOM ASSIGNMENT

Understanding the utility of randomly assigning participants to treatment or control groups is best explained by example: Dr. Zinn and Dr. Zorders have come up with Q-Math, a revolutionary system for teaching multiplication. The Q-Math package is shipped out to schools in a local district to determine if it is more effective than the current teaching method. The instructions specify that each fourth grade class should be divided in half and routed to separate rooms, with students in one room receiving the Q-Math teaching and students in the other room getting their regular math lesson. At the end, both groups are administered a multiplication test, and the results of both groups are compared. The question is: *How should the class be divided into two groups?* This is not such a simple question. If the classroom is divided into boys and girls, this may influence the outcome, because gender may be a relevant factor in math skills—if by chance we send the gender with stronger math skills to receive the Q-Math intervention, this may serve to inflate those scores. Alternatively, suppose we decided to slice the class in half by seating. This introduces a different potential confound—what if the half who sit near the front of the classroom are naturally more attentive than those who sit in the back half of the classroom? Again, this grouping method may confound the findings of the study. Finally, suppose the teacher splits the class by age. This presents yet another potential confound—maybe older students are able to perform math better than younger students. In addition, it is unwise to allow participants

to self-select which group they want to be in; it may be that more proficient math students, or students who take their studies more seriously, may systemically opt for the Q-Math group, thereby potentially influencing the outcome.

Through this simple example, it should be clear that the act of selectively assigning individuals to (treatment or control) groups can unintentionally affect the outcome of a study; it is for this reason that we often opt for random assignment to assemble more balanced groups. In this example, the Q-Math instructions may specify that a coin flip be used to assign students to each of the two groups: Heads assigns a student to Q-Math, and tails assigns a student to the usual math teaching method. This random assignment method ultimately means that regardless of factors such as gender, seating position, age, math proficiency, and academic motivation, each student will have an equal (50/50) chance of being assigned to either group. The process of random assignment will generally result in roughly the same proportion of girls and boys, the same propotion of math-smart students, the same proportion of front- and back-of-the-room students, and the same proportion of older and younger students being assigned to each group. If done properly, random assignment helps cancel out factors inherent to participants that may have otherwise biased the findings one way or another.

HYPOTHESIS FORMULATION

Everyone has heard of the word *hypothesis*; hypotheses simply spell out each of the anticipated possible outcomes of an experiment. In simplest terms, before we embark on the experiment, we need one hypothesis that states that nothing notable happened, because sometimes experiments fail. This would be the *null hypothesis* (H_0), basically meaning that the treatment had a null effect—nothing notable happened.

Another possibility is that something notable did happen (the experiment worked), so we would need an *alternative hypothesis* (H_1) that accounts for this.

Continuing with the above example involving Q-Math, we first construct the null hypothesis (H_0); as expected, the null hypothesis states that the experiment produced null results—basically, the experimental group (the group that got Q-Math) and the control group (the group that got regular math) performed about the same; essentially, that would mean that Q-Math was no more effective than the traditional math lesson. The alternative hypothesis (H_1) is phrased indicating that the treatment (Q-Math) group outperformed the control (regular math lesson) group. Hypotheses are typically written in this fashion:

H_0: Q-Math and regular math teaching methods produce equivalent test results.

H_1: Q-Math produces higher test results compared with regular teaching methods.

When the results are in, we would then know which hypothesis to reject and which to accept; from there, we can document and discuss our findings.

Remember: In simplest terms, the statistics we will be processing are designed to answer the question: *Do the members of the treatment group (who get the innovative*

treatment) significantly outperform the members of the control group (who get no treatment, a placebo, or treatment as usual)? As such, the hypotheses need to reflect each possible outcome. In this simple example, we can anticipate two possible outcomes: H_0 states that there is *no* significant difference between the treatment group and the control group, suggesting that *the treatment was ineffective*. On the other hand, we need another hypothesis that anticipates that the treatment will significantly outperform the control condition; as such, H_1 states that there *is* a significant difference in the outcomes between the treatment and control conditions, suggesting that *the treatment was effective*. The outcome of the statistical test will point us to which hypothesis to accept and which to reject.

READING STATISTICAL OUTCOMES

Statistical tests vary substantially in terms of the types of research questions each is designed to address, the format of the source data, their respective equations, and the content of their results, which can include figures, tables, and graphs. Although there are some similarities in reading statistical outcomes (e.g., means, alpha [α] levels, *p* values), these concepts are best explained in the context of working examples; as such, how to read statistical outcomes will be thoroughly explained as each emerges in Chapters 4 through 9.

ACCEPT OR REJECT HYPOTHESES

As is the case with reading statistical outcomes, the decision to accept or reject a hypothesis depends on the nature of the test and, of course, the results: the alpha (α) level, *p* value, and, in some cases, the means. Just as with reading statistical outcomes, instructions for accepting or rejecting hypotheses for each test are best discussed in the context of actual working examples; these concepts will be covered in Chapters 5 through 9.

VARIABLE TYPES AND LEVELS OF MEASURE

Comprehending the types of variables involved in a data set or research design is essential when it comes to properly selecting, running, and documenting the results of statistical tests. There are two types of variables: *continuous* and *categorical*. Each has two levels of measure; continuous variables may be either *interval* or *ratio*, and categorical variables may be either *nominal* or *ordinal*.

Basically, you will need to be able to identify the types of variables you will be processing (*continuous* or *categorical*), which will help guide you in selecting and running the proper statistical analyses.

Continuous

Continuous variables contain the kinds of numbers we are accustomed to dealing with in counting and mathematics. A continuous variable may be either *interval* or *ratio*.

Interval

Interval variables range from $-\infty$ to $+\infty$, like numbers on a number line. These numbers have equal spacing between them; the distance between 1 and 2 is the same as the distance between 2 and 3, which is the same as the distance between 3 and 4, and so on. Such variables include bank account balance (which could be negative) and temperature (e.g., –40° to 85°), as measured on either the Fahrenheit or Celsius scale. Interval variables are considered continuous variables.

Ratio

Ratio variables are similar to interval variables, except that interval variables can have negative values, whereas ratio variables cannot be less than zero. The zero value in a ratio variable indicates that there is none of that variable; temperature measured in degrees Celsius or Fahrenheit is not a ratio variable, because zero degrees Celsius does not mean there is no temperature. Finally, by definition, when comparing two ratio variables, you can look at the *ratio* of two measurements; an adult who weighs 160 pounds weighs twice as much as a child who weighs 80 pounds. Examples of ratio variables include weight, distance, income, calories, academic grade (0% to 100%), number of pets, number of pencils in a pencil cup, number of siblings, or number of members in a group. Ratio variables are considered continuous variables.

Learning tip: Notice that the word *ratio* ends in *o*, which looks like a *zero*.

Categorical

Categorical variables (also known as discrete variables) involve assigning a number to an item in a category. A categorical variable may be either *nominal* or *ordinal*.

Nominal

Nominal variables are used to represent categories that defy ordering. For example, suppose you wish to code eye color, and there are six choices: amber, blue, brown, gray, green, and hazel. There is really no way to put these in any order; for coding and computing purposes, we could assign 1 = amber, 2 = blue, 3 = brown, 4 = gray, 5 = green, and 6 = hazel. Since order does not matter among nominal variables, these eye colors could have just as well been numbered differently: 1 = blue, 2 = green, 3 = hazel, 4 = gray, 5 = amber, and 6 = brown. Nominal variables may be used to represent

categorical variables such as gender (1 = female, 2 = male), agreement (1 = yes, 2 = no), religion (1 = atheist, 2 = Buddhist, 3 = Catholic, 4 = Hindu, 5 = Jewish, 6 = Taoist, etc.), or marital status (1 = single, 2 = married, 3 = separated, 4 = divorced, 5 = widow or widower).

Since the numbers are arbitrarily assigned to labels within a category, it would be inappropriate to perform traditional arithmetic calculations on such numbers. For example, it would be foolish to compute the average marital status (e.g., would 1.5 indicate a *single married* person?). The same principle applies to other nominal variables, such as gender or religion. There are, however, appropriate statistical operations for processing nominal variables that will be discussed in Chapter 4 ("Descriptive Statistics"). In terms of statistical tests, nominal variables are considered categorical variables.

Learning tip: There is no order among the categories in a nominal variable; notice that the word *nominal* starts with *no*, as in *no order*.

Ordinal

Ordinal variables are similar to nominal variables in that numbers are assigned to represent items within a category. Whereas nominal variables have no real rank order to them (e.g., amber, blue, brown, gray, green, hazel), the values in an ordinal variable can be placed in a ranked order. For example, there is an order to educational degrees (1 = high school diploma, 2 = associate's degree, 3 = bachelor's degree, 4 = master's degree, 5 = doctoral degree). Other examples of ordinal variables include military rank (1 = private, 2 = corporal, 3 = sergeant, etc.) and meals (1 = breakfast, 2 = brunch, 3 = lunch, 4 = dinner, 5 = late-night snack). In terms of statistical tests, ordinal variables are considered categorical variables.

Learning tip: Notice that the root of the word *ordinal* is *order*, suggesting that the categories have a meaningful *order* to them.

Summary of Variable Types

Continuous	Interval	(…–3, –2, –1, 0, 1, 2, 3…)
	Ratio	(0, 1, 2, 3…)
Categorical	Nominal	(Red, Blue, Green)
	Ordinal	(Breakfast, Lunch, Dinner)

GOOD COMMON SENSE

As we explore the results of multiple statistics throughout this text, keep in mind that no matter how precisely we proceed, the process of statistics is not perfect. Our findings do not *prove* or *disprove* anything; rather, statistics helps us reduce uncertainty—to help us better comprehend the nature of those we study.

Additionally, what we learn from statistical findings speaks to the *group* we studied on an *overall basis*, not any one *individual*. For instance, suppose we find that the average age within a group is 25; this does not mean that we can just point to any one person in that group and confidently proclaim "You are 25 years old."

Key Concepts

- Rationale for statistics
- Research question
- Treatment group
- Control group
- Random assignment
- Hypotheses (null, alternative)
- Statistical outcomes
- Accepting or rejecting hypotheses
- Types of data (continuous, categorical)
- Level of data (continuous: interval, ratio; categorical: nominal, ordinal)

Practice Exercises

Each of the following exercises describes the basis for an experiment that would render data that could be processed statistically.

Exercise 1.1

It is expected that aerobic square dancing during the 30-minute recess at an elementary school will help fight childhood obesity.

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses.

Exercise 1.2

Recent findings suggest that nursing home residents may experience fewer depressive symptoms when they participate in pet therapy with certified dogs for 30 minutes per day.

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses.

Exercise 1.3

A chain of retail stores has been experiencing substantial cash shortages in cashier balances across 10 of its stores. The company is considering installing cashier security cameras.

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses.

Exercise 1.4

Anytown Community wants to determine if implementing a neighborhood watch program will reduce vandalism incidents.

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses.

Exercise 1.5

Employees at Acme Industries, consisting of four separate buildings, are chronically late. An executive is considering implementing a "get out of Friday free" lottery; each day an employee is on time, he or she gets one token entered into the weekly lottery.

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses.

Exercise 1.6

The Acme Herbal Tea Company advertises that its product is "the tea that relaxes."

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses.

Exercise 1.7

Professor Madrigal has a theory that singing improves memory.

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses.

Exercise 1.8

Mr. Reed believes that providing assorted colored pens will prompt his students to write longer essays.

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses.

Exercise 1.9

Ms. Fractal wants to determine if working with flash cards helps students learn the multiplication table.

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses

Exercise 1.10

A manager at the Acme Company Call Center wants to see if running a classic movie on a big screen (with the sound off) will increase the number of calls processed per hour.

a. State the research question.

b. Identify the control and experimental group(s).

c. Explain how you would randomly assign participants to groups.

d. State the hypotheses (H_0 and H_1).

e. Discuss the criteria for accepting or rejecting the hypotheses.

CHAPTER 2

Sampling

Who needs the whole population when a ***sample*** *will do nicely?*

- Rationale for Sampling
- Sampling Terminology
- Representative Sample
- Probability Sampling
- Nonprobability Sampling
- Sampling Bias

Ya gots to work with what you gots to work with.

—Stevie Wonder

LEARNING OBJECTIVES

Upon completing this chapter, you will be able to:

Comprehend the rationale for sampling: time, cost, feasibility, extrapolation
Understand essential sampling terminology: population, sample frame, sample
Derive a representative sample to facilitate external validity
Select an appropriate method to conduct probability sampling: simple random sampling, stratified sampling, proportionate sampling, disproportionate sampling, systematic sampling, area sampling
Select an appropriate method to conduct nonprobability sampling: convenience sampling, purposive sampling, quota sampling, snowball sampling
Understand techniques for detecting and reducing sample bias
Optimize sample size

OVERVIEW—SAMPLING

Statistics is about processing numbers in a way to produce concise, readily consumable information. One statistic you are probably already familiar with is the average. Suppose you wanted to know the average age of students in a classroom. The task would be fairly simple—you could ask each person to write down his or her age on a slip of paper and then proceed with the calculations. In the relatively small setting of a classroom, it is possible to promptly gather the data on everyone, but what if you wanted to know the age of all enrolled students or all students in a community? Now the mission becomes more time-consuming, complex, and probably expensive. Instead of trying to gather data on everyone, as in the census survey, another option is to gather a sample. Gathering a sample of a population is quicker, easier, and more cost-effective than gathering data on everyone, and if done properly, the findings from your sample can provide you with quality information about the overall population.

You may not realize it, but critical decisions are made based on samples all the time. Laboratories process thousands of blood samples every day. On the basis of the small amount of blood contained in the test tube, a qualified health care professional can make determinations about the overall health status of the patient from whom the blood was drawn. Think about that for a moment: A few cubic centimeters of blood is sufficient to carry out the tests to make quality determinations. The laboratory did not need to drain the entire blood supply from the patient, which would be time-consuming, complicated, expensive, and totally impractical—it would kill the patient. Just as a small sample of blood is sufficient to represent the status of the entire blood supply, proper sampling enables us to gather a small and manageable bundle of data from a population of interest, statistically process those data, and reasonably comprehend the larger population from which it was drawn.

RATIONALE FOR SAMPLING

Moving beyond the realm of a statistics course, statistics takes place in the real world to answer real-world questions. As with most things in the real world, gathering data involves the use of scarce resources; key concerns involve the time, cost, and feasibility associated with gathering quality data. With these very real constraints in mind, it is a relief to know that it is not necessary to gather *all* of the data available; in fact, it is rare that a statistical data set consists of figures from the entire population (such as the census). Typically, we proceed with a viable sample and extrapolate what we need to know to better comprehend the larger population from which that sample was drawn. Let us take a closer look at each of these factors.

Time

Some consider time to be the most valuable asset. Time cannot be manufactured or stored—it can only be used. Time spent doing one thing means that other things must

wait. Spending an exorbitant amount of time gathering data from an entire population precludes the accomplishment of other vital activities. For example, suppose you are interested in people's opinions (yes or no) regarding the death penalty for a paper you are drafting for a course. Every minute you spend gathering data postpones your ability to proceed with the completion of the paper, and that paper has a firm due date. Additionally, there are other demands competing for your time (e.g., other courses, work, family, friends, rest, recreation). Sampling reduces the amount of time involved in gathering data, enabling you to statistically process the data more promptly and proceed with the completion of the project within the allotted time.

Another aspect of time is that some (statistical) answers are time sensitive. Political pollsters must use sampling to gather information in a prompt fashion, leaving sufficient time to interpret the findings and adjust campaign strategies prior to the election. They simply do not have time to poll all registered voters—a well-drawn sample is sufficient.

Cost

Not all data are readily available for free. Some statistical data may be derived from experiments or interviews, which involves multiple costs, including a recruitment advertising budget, paying staff members to screen and process participants, providing reasonable financial compensation to study participants, facility expenses, and so on. Surveys are not free either; expenses may include photocopying, postage, website implementation charges, telephone equipment, and financial compensation to study participants and staff members. Considering the costs associated with data collection, one can see the rationale for resorting to sampling as opposed to attempting to gather data from an entire population.

Feasibility

Data gathering takes place in the real world; hence, real-world constraints must be reckoned with when embarking on such research. Because of time and budgetary constraints, it is seldom feasible or necessary to gather data on a population-wide basis; sampling is a viable option. In the case involving the blood sample, clearly it is neither necessary nor feasible to submit the patient's entire blood supply to the lab for testing—quality determinations can be made based on well-drawn samples. In addition, if a research project focuses on a large population (e.g., all students in a school district) or a population spanning a large geographical region, it may not be feasible to gather data on that many people; hence, sampling makes sense.

Extrapolation

It turns out that by sampling properly, it is unnecessary to gather data on an entire population to achieve a reasonable comprehension of it. Extrapolation involves using sampling methods and statistics to analyze the sample of data that was drawn from the population. If done properly, such findings help us (better) understand not only the smaller (sample) group but also the larger group from which it was drawn.

SAMPLING TERMINOLOGY

As in any scientific endeavor, the realm of sampling has its own language and methods. The following terms and types of sampling methods will help you comprehend the kinds of sampling you may encounter in scientific literature and provide you with viable options for carrying out your own studies. We will begin with the largest realm (the *population*) and work our way down to the smallest (the *sample*).

Population

The *population* is the entire realm of people (or items) that could be measured or counted. A population is not simply all people on the planet; the researcher specifies the population, which consists of the entire domain of interest. For example, the population may be defined as all students who are currently enrolled at a specified campus. Additional examples of populations could be all people who reside in a city, all people who belong to a club, all people who are registered voters in an election district, or all people who work for a company. As you might have surmised, the key word here is *all*.

Sample Frame

If the population you are interested in is relatively small (e.g., the 5 people visiting the public park, the 16 people who are members of a club), then gathering data from the entire population is potentially doable. More often, the population is larger than you can reasonably accommodate, or you may be unable to attain a complete list of the entire population you are interested in (e.g., all students enrolled in a school, every registered voter in an election district). The *sample frame*, sometimes referred to as the *sampling frame*, is the part of a population you could potentially access. For example, Acme University publishes a list of student names and e-mail addresses in the form of a downloadable report on the school's public website. If this list included every single student enrolled, it would represent the entire population of the school; however, students have the privilege to opt out of this list, meaning that each student can go to his or her online profile and check a box to include or exclude his or her name and e-mail address from this public roster. Suppose the total population of the university consists of 30,000 enrolled students, and 70% chose to have their names appear on this list; this would mean that the sample frame, the list from which you could potentially select subjects, consists of 21,000 students (30,000 × .70).

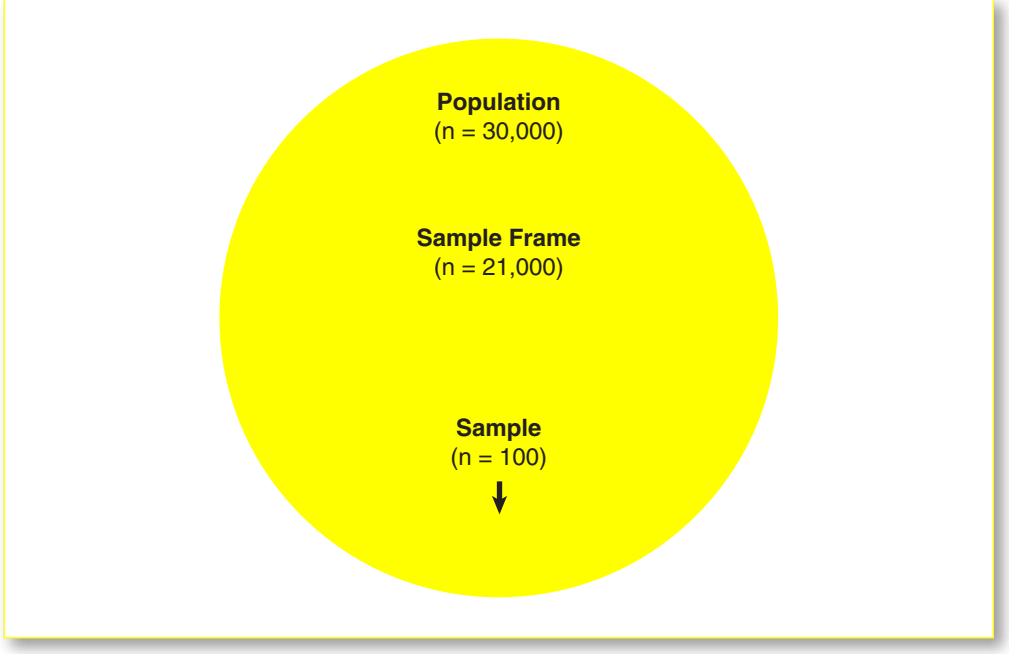
Sample

A *sample* is a portion of individuals selected from the sample frame. Certainly 21,000 is considerably less than 30,000, but that may still be an unwieldy amount for your purposes. Consider that your investigation involves conducting a 1-hour interview with participants and that each participant will be compensated $10 for his or her time; the

subject fee budget for this study would be $210,000, and assuming you conducted back-to-back interviews for 8 hours a day, 7 days a week, you would have your data set in a little over 7 years. Considering the constraints mentioned earlier (time, cost, and feasibility), you can probably already see where this is going: (a) Is a $210,000 budget for subjects really feasible? (b) Do you really have 7 years to gather your findings? (c) Most of the students on this list will not be students 7 years from now. (d) After students graduate, their e-mail addresses may change. In this case, accessing the entire sample frame is untenable, but the sample frame is still useful; instead of attempting to recruit the 21,000 students, you may choose to gather information from a subset of 100 students from this sample frame. These 100 students will constitute the sample. Selecting a sample of 100 students from the sample frame of 21,000 means that your subject fee budget is reduced from $210,000 to $1,000, and instead of taking more than 7 years to gather the data, using the same interviewing schedule, you would have your complete data set in less than 2 weeks. In terms of time, cost, and feasibility, sampling is clearly the way to go. As for *how* to select that sample of 100 students from among the sample frame of 21,000, there are a variety of techniques covered in the sections on probability sampling and nonprobability sampling.

Just to recap, you can think of the population, sample frame, and sample as a hierarchy (Figure 2.1):

Figure 2.1 Three tiers of sampling: population, sample frame, and sample.



- The *population* is the entire realm of those in a specified set (e.g., every person who lives in a city, all members of an organization or club, all students enrolled on a campus).
- The *sample frame* is the list of those who could be potentially accessed from a population.
- The *sample* is the sublist of those selected from the sample frame whom you will (attempt to) gather data from.

REPRESENTATIVE SAMPLE

You may not realize it, but you already understand the notion of a *representative sample*. Suppose you are at a cheese-tasting party, and the host brings out a large wheel of cheese from Acme Dairy. You are served a small morsel of the cheese that is less than 1% of the whole cheese and, based on that, you decide if you like it enough to buy a hunk of it or not. The assumption that you are perhaps unknowingly making is that the rest of that big cheese will be exactly like the tiny sample you tasted. You are presuming that the bottom part of the cheese is not harder, that the other side of the cheese is not sharper, that the middle part of the cheese is not runnier, and so on. Essentially, you are assuming that the sample of cheese you tasted is representative of the flavor, color, and consistency of the whole wheel of cheese. This is what a representative sample is all about: The small sample you drew is proportionally representative of the overall population (or big cheese) from which it was taken. Often, it is the goal of researchers to select a representative sample, thereby facilitating *external validity*—meaning that what you discover about the sample can be viably generalized to the overall population from which the sample was drawn.

Sampling is about gathering a manageable set of data so that you can learn something about the larger population through statistical analysis. The question remains: *How do you get from the population, to the sample frame, to the actual representative sample?* Depending on the nature of the information you are seeking and the availability of viable participants and data, you may opt to use *probability sampling* or *nonprobability sampling* methods.

PROBABILITY SAMPLING

You can think of probability sampling as equal-opportunity sampling, meaning that each potential element (person or data record) has the same chance of being selected for your sample. There are several ways of conducting probability sampling, which include simple random sampling, systematic sampling, stratified sampling, proportionate or disproportionate sampling, and area sampling.

Simple Random Sampling

Simple random sampling begins with gathering the largest sample frame possible and then numbering each item or person (1, 2, 3, . . . 60). For this example, let us assume that there are 60 people (30 women and 30 men) on this list, and you want to recruit

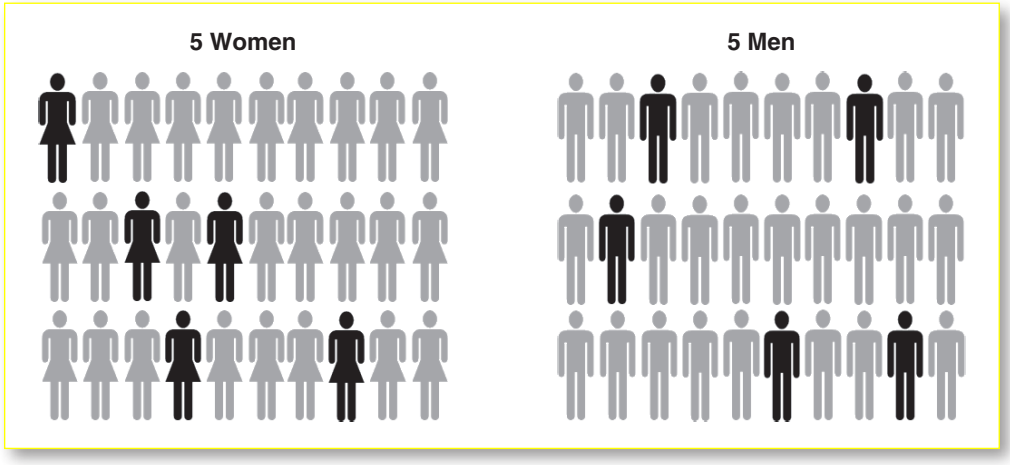
Figure 2.2 Simple random sampling. The researcher randomly selects a sample of 10 from a sample frame of 60.

10 participants (Figure 2.2); you could use SPSS to generate 10 random numbers ranging from 1 to 60. It is not essential that you perform this procedure at this time; Chapter 10 ("Supplemental SPSS Operations") has a section that provides step-by-step instructions for generating random numbers to your specifications.

Stratified Sampling

In the above example, simple random sampling rendered a sample consisting of 80% women and 20% men, which may not suit your needs. Suppose you still want a sample of 10 from the sample frame of 60, but instead of leaving the gender counts to random chance, you want to specifically control for the numbers of women and men in your sample; stratified sampling would ensure that your sample is balanced by gender. To draw a stratified sample based on gender, divide your sample frame into two lists (strata): women and men. In this case, the initial sample frame of 60 is divided into two separate strata: 30 women and 30 men. Suppose you still want to draw a (total) sample of 10; you would use simple random sampling to randomly select 5 participants from the female stratum and another 5 participants from the male stratum (Figure 2.3).

Figure 2.3 Stratified sampling. The researcher splits the sample frame into two strata, women and men, and randomly selects a sample of five from each stratum.



NOTE: Systematic sampling (which will be discussed on page 28) could also be used to make selections within each strata.

Proportionate and Disproportionate Sampling

Within the realm of stratified sampling, you can further specify if you want to gather a *proportionate* or *disproportionate* sample. Continuing with the gender stratification example, suppose you were conducting a survey of people that consists of 30 women and 10 men, and for the purposes of this survey, gender is relevant.

The first step is to split the sample frame into two strata (lists) based on gender: women and men. You then have the option to draw a proportionate sample or a disproportionate sample. To draw a proportionate sample, you would draw the same percentage (in this case, 10%) from each stratum: 3 from the 30 in the female stratum and 1 from the 10 in the male stratum (Figure 2.4).

When the count within one or more strata is relatively low, proportionate sampling will expectedly produce a sample of the stratum that may be too small to be viable; in the above case, sampling 10% from the 10 in the male strata means selecting only 1 male participant. In such instances, disproportionate sampling may be a better choice.

To gather a disproportionate sample, randomly select the same number of participants from each stratum, regardless of the size of the stratum. In this case (Figure 2.5), three are being drawn from each stratum: three women and three men. Although the sample sizes from the two strata are now equal (three from each stratum), the proportions are now different; 10% of the women have been selected, whereas 30% of the men have been selected.

Figure 2.4 Proportionate stratified sampling. The researcher randomly selects the same percentage from each strata.

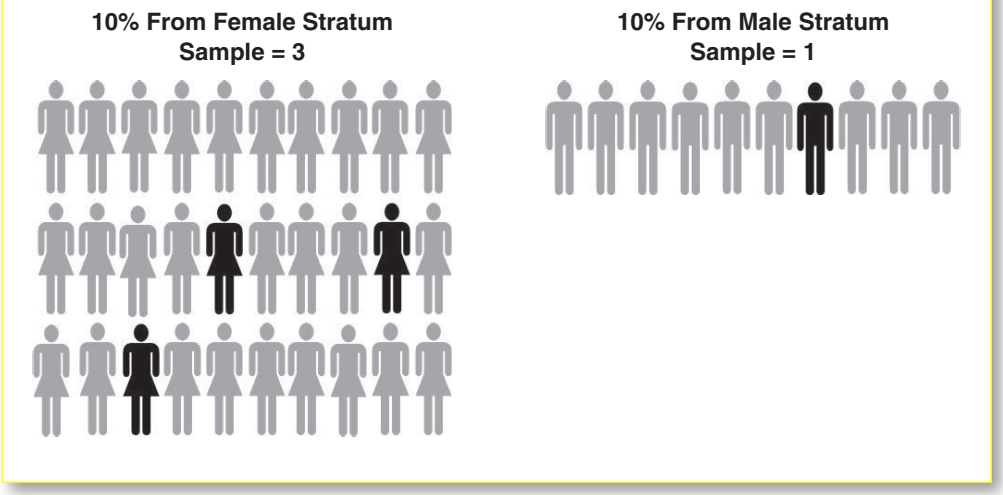
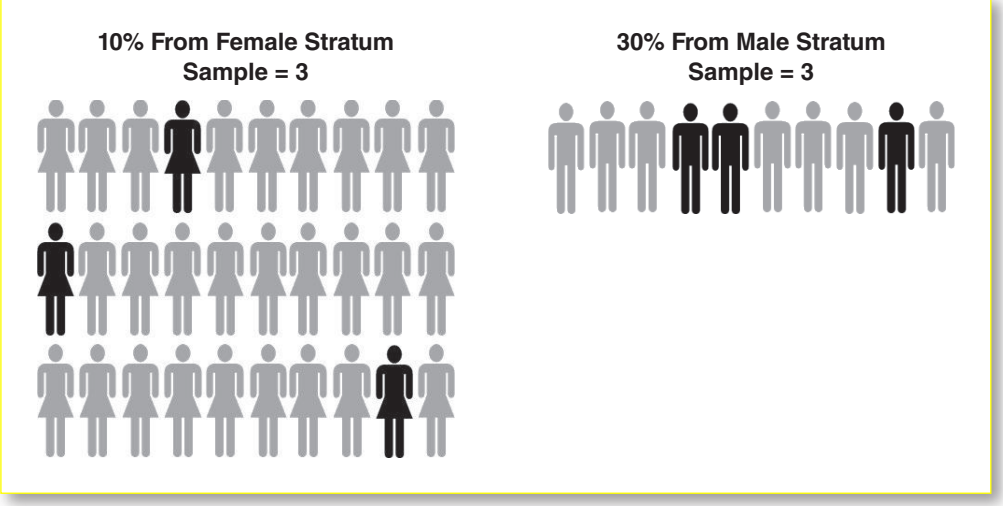


Figure 2.5 Disproportionate stratified sampling. The researcher randomly selects the same (total) number from each stratum.



Systematic Sampling

Whereas simple random sampling may produce a sample wherein the items may be drawn from similar proximity (e.g., several participants who are next to one another may all be selected), systematic sampling uses a periodic selection process that draws the sample more evenly throughout the sample frame.

Suppose you have 60 people, and you decide that the target sample size will be 15 participants. Begin by dividing the sample frame by the target sample size (60 ÷ 15 = 4); the solution (4) is the "*k*" or skip term. Next, you need to identify the start point; this will be a random number between 1 and *k*. For this example, suppose the randomly derived start point number is 3. The process begins with selecting the third person and then skips ahead *k* (4) people at a time to select each additional participant who will be included in the sample. In this case, the following 15 participants would be selected: 3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, and 59 (Figure 2.6).

Figure 2.6 Systematic sampling. The researcher selects a periodic sample from the sample frame: *k* (skip term) = 4 (*k* = *sample frame* ÷ target *sample* size: 60 ÷ 15 = 4). Start = 3 (start = a random number between 1 and *k*).