Statway™

A statistics pathway for college students

Module 1: Statistical Studies and Overview of the Data Analysis Process
Module 2: Summarizing Data Graphically and Numerically
Module 3: Reasoning About Bivariate Numerical Data—Linear Relationships
Module 4: Modeling Nonlinear Relationships
Module 5: Reasoning About Bivariate Categorical Data and Introduction to Probability
Module 6: Formalizing Probability and Probability Distributions
Module 7: Linking Probability to Statistical Inference
Module 8: Inference for One Proportion
Module 9: Inference for Two Proportions
Module 10: Inference for Means
Module 11: Chi-Squared Tests
Module 12: Other Mathematical Content

Version 1.0

A resource from
The Charles A. Dana Center at
The University of Texas at Austin

July 2011
Copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin

Outside the license described below, no part of this resource shall be reproduced, stored in a retrieval system, or transmitted by any means—electronically, mechanically, or via photocopying, recording, or otherwise, including via methods yet to be invented—without express written permission from the Foundation and the University.

The original version of this work was created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching.

STATWAY™ / Statway™ is a trademark of the Carnegie Foundation for the Advancement of Teaching.

***

This copyright notice is intended to prohibit unlicensed commercial use of the Statway materials.

License for use

Statway Version 1.0, developed by the Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported license.

To view the details of this license, see creativecommons.org/licenses/by-nc-sa/3.0. In general, under this license

You are free:

to Share—to copy, distribute, and transmit the work

to Remix—to adapt the work

Under the following conditions:

Attribution—You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). We request you attribute the work thus:

The original version of this work was developed by the Charles A. Dana Center at the University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. This work is used (or adapted) under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0) license: creativecommons.org/licenses/by-nc-sa/3.0. For more information about Carnegie’s work on Statway, see www.carnegiefoundation.org/statway; for information on the Dana Center’s work on The New Mathways Project, see www.utdanacenter.org/mathways.

Noncommercial—You may not use this work for commercial purposes.

Share Alike—If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

The Charles A. Dana Center at the University of Texas at Austin, as well as the authors and editors, assume no liability for any loss or damage resulting from the use of this resource. We have made extensive efforts to ensure the accuracy of the information in this resource, to provide proper acknowledgement of original sources, and to otherwise comply with copyright law. If you find an error or you believe we have failed to provide proper acknowledgment, please contact us at dana-txshop@utlists.utexas.edu.
The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.

The Charles A. Dana Center
The University of Texas at Austin
1616 Guadalupe Street, Suite 3.206
Austin, TX 78701-1222
Fax: 512-232-1855
dana-txshop@utlists.utexas.edu
www.utdanacenter.org

The Carnegie Foundation for the Advancement of Teaching
51 Vista Lane
Stanford, California, 94305
Phone: 650-566-5110
pathways@carnegiefoundation.org
www.carnegiefoundation.org

About the development of this resource
The content for this full version of Statway was developed under a November 30, 2010, agreement by a team of faculty authors and reviewers contracted and managed by the Charles A. Dana Center at the University of Texas at Austin with funding from the Carnegie Foundation for the Advancement of Teaching.

This resource was produced in Microsoft Word 2008 and 2011 for the Mac. The content of these 12 modules was developed and produced (that is, written, reviewed, edited, and laid out) by the Charles A. Dana Center at The University of Texas at Austin and delivered by the Dana Center to the Carnegie Foundation for the Advancement of Teaching on June 30, 2011.

Some issues to be aware of:
- PDF files need to be viewed with Adobe Acrobat for full functionality. If viewed through Preview, which is the default on some computers, the URLs may not be correct.
- The file names indicate the lesson number and whether the document is the instructor or student version or the out-of-class experience.

The Dana Center is engaged in a process of revising and improving these materials to create the Dana Center Statistics Pathway. We welcome feedback from the community as part of our course revision process. If you would like to discuss these materials or learn more about the Dana Center’s plans for this course, contact us at mathways@austin.utexas.edu.

About the Charles A. Dana Center at The University of Texas at Austin
The Dana Center collaborates with local and national entities to improve education systems so that they foster opportunity for all students, particularly in mathematics and science. We are dedicated to nurturing students’ intellectual passions and ensuring that every student leaves school prepared for success in postsecondary education and the contemporary workplace—and for active participation in our modern democracy.

The Center was founded in 1991 in the College of Natural Sciences at The University of Texas at Austin. Our original purpose—which continues in our work today—was to raise student achievement in K–16 mathematics and science, especially for historically underserved populations. We carry out our work by supporting high standards and building system capacity; collaborating with key state and national organizations to address emerging issues; creating and delivering professional supports for educators and education leaders; and writing and publishing education resources, including student supports.

Our staff of more than 80 researchers and education professionals has worked intensively with dozens of school systems in nearly 20 states and with 90 percent of Texas’s more than 1,000 school districts. As one of the College’s largest research units, the Dana Center works to further the university’s mission of achieving excellence in education, research, and public service. We are committed to ensuring that the accident of where a student attends school does not limit the academic opportunities he or she can pursue.

For more information about the Dana Center and our programs and resources, see our homepage at www.utdanacenter.org. To access our resources (many of them free) please see our products index at www.utdanacenter.org/products. To learn about Dana Center professional development sessions, see our professional development site at www.utdanacenter.org/pd.
Acknowledgments

The original version of this work was created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching.


Leadership—Charles A. Dana Center at the University of Texas at Austin

Uri Treisman, director
Susan Hudson Hull, program director of mathematics national initiatives

Leadership—Carnegie Foundation for the Advancement of Teaching

Anthony S. Bryk, president
Bernadine Chuck Fong, senior managing partner
Louis Gomez, senior fellow
Paul LeMahieu, senior fellow
James Stigler, senior fellow
Uri Treisman, senior fellow
Guadalupe Valdés, senior fellow

Statway Project Leads

Kristen Bishop, former team lead for the New Mathways Project, the Charles A. Dana Center at the University of Texas at Austin
Thomas J. Connolly, project lead, Statway, the Charles A. Dana Center at the University of Texas at Austin
Karon Klipple, director of Statway, the Carnegie Foundation for the Advancement of Teaching
Jane Muhich, director of Quantway, the Carnegie Foundation for the Advancement of Teaching

Project Staff—Charles A. Dana Center at the University of Texas at Austin

Richard Blount, advisor
Kathi Cook, project director, online services team
Jenna Cullinane, research associate
Steve Engler, lead editor and production editor
Amy Getz, team lead for the New Mathways Project
Susan Hudson Hull, program director of mathematics national initiatives
Joseph Hunt, graduate research assistant
Rachel Jenkins, consulting editor
Erica Moreno, program coordinator
Carol Robinson, administrative associate
Cathy Seeley, senior fellow
Rachele Seifert, administrative associate
Lilly Soto, senior administrative associate
Phil Swann, senior designer
Laura Torres, graduate research assistant
Thomas Wiegel, freelance formatter and proofreader
Authors Contracted by the Dana Center
Roxy Peck, professor emerita of statistics, California Polytechnic State University, San Luis Obispo, California
Beth Chance, professor of statistics, California Polytechnic State University, San Luis Obispo, California
Robert C. delMas, associate professor of educational psychology, University of Minnesota, Minneapolis, Minnesota
Scott Guth, professor of mathematics, Mt. San Antonio College, Walnut, California
Rebekah Isaak, graduate research student, University of Minnesota, Minneapolis, Minnesota
Leah McGuire, assistant professor, University of Minnesota, Minneapolis, Minnesota
Jiyoon Park, graduate research student, University of Minnesota, Minneapolis, Minnesota
Brian Kotz, associate professor of mathematics, Montgomery College, Germantown, Maryland
Chris Olsen, assistant professor of mathematics and statistics, Grinnell College, Grinnell, Iowa
Mary Parker, professor of mathematics, Austin Community College, Austin, Texas
Michael A. Posner, associate professor of statistics, Villanova University, Villanova, Pennsylvania
Thomas H. Short, professor, John Carroll University, University Heights, Ohio
Penny Smeltzer, teacher of statistics, Westwood High School, Austin, Texas
Myra Snell, professor of mathematics, Los Medanos College, Pittsburg, California
Laura Ziegler, graduate research student, University of Minnesota, Minneapolis, Minnesota

Reviewers Contracted by the Dana Center
Michelle Brock, American River College, Sacramento, California
Thomas J. Connolly, the Charles A. Dana Center at the University of Texas at Austin
Andre Freeman, Capital Community College, Hartford, Connecticut
Karon Klipple, the Carnegie Foundation for the Advancement of Teaching
Roxy Peck, professor emerita of statistics, California Polytechnic State University, San Luis Obispo, California
Jim Smart, Tallahassee Community College, Tallahassee, Florida
Myra Snell, Los Medanos College, Pittsburg, California

Committee for Statistics Learning Outcomes
Rose Asera, formerly of the Carnegie Foundation for the Advancement of Teaching
Kristen Bishop, formerly of the Charles A. Dana Center at the University of Texas at Austin
Richelle (Rikki) Blair, American Mathematical Association of Two-Year Colleges (AMATYC); Lakeland Community College, Ohio
David Bressoud, Mathematical Association of America (MAA); Macalester College, Minnesota
John Climent, American Mathematical Association of Two-Year Colleges (AMATYC); Cecil College, Maryland
Peg Crider, Lone Star College, Tomball, Texas
Jenna Cullinane, the Charles A. Dana Center at the University of Texas at Austin
Robert C. delMas, Consortium for the Advancement of Undergraduate Statistics Education (CAUSE); University of Minnesota, Minneapolis, Minnesota
Bernadine Chuck Fong, the Carnegie Foundation for the Advancement of Teaching
Karen Givvin, the University of California, Los Angeles
Larry Gray, American Mathematical Society (AMS); University of Minnesota
Susan Hudson Hull, the Charles A. Dana Center at the University of Texas at Austin
Rob Kimball, American Mathematical Association of Two-Year Colleges (AMATYC); Wake Technical Community College, North Carolina
Dennis Pearl, Consortium for the Advancement of Undergraduate Statistics Education (CAUSE); The Ohio State University
Roxy Peck, American Statistical Association (ASA); Consortium for the Advancement of Undergraduate Statistics Education (CAUSE); California Polytechnic State University, San Luis Obispo, California

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Myra Snell, American Mathematical Association of Two-Year Colleges (AMATYC); Los Medanos College, Pittsburg, California
Jim Stigler, the Carnegie Foundation for the Advancement of Teaching; the University of California, Los Angeles
Daniel Teague, Mathematical Association of America (MAA); North Carolina School of Science and Mathematics, Durham
Uri Treisman, the Carnegie Foundation for the Advancement of Teaching; the Charles A. Dana Center at the University of Texas at Austin

Version 1.0 of Statway was developed in collaboration with faculty from the following colleges, the “Collaboratory,” who advised on the development of the course. These Collaboratory colleges are:

**California**
American River College, Sacramento, California
Foothill College, Los Altos Hills, California
Mt. San Antonio College, Walnut, California
Pierce College, Woodland Hills, California
San Diego City College, San Diego, California

**California State University System**
CSU Northridge
Sacramento State University
San Jose State University

**Connecticut**
Capital Community College, Hartford, Connecticut
Gateway Community College, New Haven, Connecticut
Housatonic Community College, Bridgeport, Connecticut
Naugatuck Valley Community College, Waterbury, Connecticut

**Florida**
Miami Dade College, Miami, Florida
Tallahassee Community College, Tallahassee, Florida
Valencia Community College, Orlando, Florida

**Texas**
Austin Community College, Austin, Texas
El Paso Community College, El Paso, Texas
Houston Community College, Houston, Texas
Northwest Vista College, San Antonio, Texas
Richland College, Dallas, Texas

**Washington**
Seattle Central Community College, Seattle, Washington
Tacoma Community College, Tacoma, Washington

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Table of Contents

Module 1: Statistical Studies and Overview of the Data Analysis Process

Lesson 1.1.1: The Statistical Analysis Process
Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions
Lesson 1.2.1: Collecting Data by Sampling
Lesson 1.2.2: Random Sampling
Lesson 1.2.3: Other Sampling Strategies
Lesson 1.2.4: Sources of Bias in Sampling
Lesson 1.3.1: Collecting Data by Conducting an Experiment
Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos
Lesson 1.4.1: Drawing Conclusions from Statistical Studies

Module 2: Summarizing Data Graphically and Numerically

Lesson 2.1.1: Dotplots, Histograms, and Distributions for Quantitative Data
Lesson 2.1.2: Constructing Histograms for Quantitative Data
Lesson 2.1.3: Comparing Distributions of Quantitative Data in Two Independent Samples
Lesson 2.2.1: Quantifying the Center of a Distribution—Sample Mean and Sample Median
Lesson 2.2.2: Constructing Histograms for Quantitative Data
Lesson 2.3.1: Quantifying Variability Relative to the Median
Lesson 2.4.1: Quantifying Variability Relative to the Mean
Lesson 2.4.2: The Sample Variance

Module 3: Reasoning About Bivariate Numerical Data—Linear Relationships

Lesson 3.1.1: Introduction to Scatterplots and Bivariate Relationships
Lesson 3.1.2: Developing an Intuitive Sense of Form, Direction, and Strength of the Relationship Between Two Measurements
Lesson 3.1.3: Introduction to the Correlation Coefficient and Its Properties
Lesson 3.1.4: Correlation Formula
Lesson 3.1.5: Correlation Is Not Causation
Lesson 3.2.1: Using Lines to Make Predictions
Lesson 3.2.2: Least Squares Regression Line as Line of Best Fit
Lesson 3.2.3: Investigating the Meaning of Numbers in the Equation of a Line
Lesson 3.2.4: Special Properties of the Least Squares Regression Line
Lesson 3.3.1: Using Residuals to Determine If a Line Is a Good Fit
Lesson 3.3.2: Using Residuals to Determine If a Line Is an Appropriate Model

Module 4: Modeling Nonlinear Relationships
Lesson 4.1.1: Investigating Patterns in Data
Lesson 4.1.2: Exponential Models
Lesson 4.1.3: Assessing How Well a Model Fits the Data

Module 5: Reasoning About Bivariate Categorical Data and Introduction to Probability
Lesson 5.1.1: Reasoning About Risk and Chance
Lesson 5.1.2: Defining Risk
Lesson 5.1.3: Interpreting Risk
Lesson 5.1.4: Comparing Risks
Lesson 5.1.5: More on Conditional Risks

Module 6: Formalizing Probability and Probability Distributions
Lesson 6.1.1: Probability
Lesson 6.1.2: Probability Rules
Lesson 6.1.3: Simulation, Discrete Random Variables, and Probability Distributions
Lesson 6.2.1: Probability Distributions of Continuous Random Variables
Lesson 6.2.2: Z-Scores and Normal Distributions
Lesson 6.2.3: Using Normal Distributions to Find Probabilities and Critical Values

Module 7: Linking Probability to Statistical Inference
Lesson 7.1.1: Predicting an Election—Statistics and Sampling Variability
Lesson 7.1.2: Sampling from a Population
Lesson 7.1.3: Testing Statistical Hypotheses
Lesson 7.2.1: Two Types of Inferential Procedures—Estimation and Hypothesis Testing
Lesson 7.2.2: Connecting Sampling Distributions and Confidence Intervals
Lesson 7.2.3: Connecting Sampling Distributions and Hypothesis Testing
Module 8: Inference for One Proportion
   Lesson 8.1.1: Sampling Distribution of One Proportion
   Lesson 8.1.2: Sampling Distribution of One Proportion
   Lesson 8.2.1: Estimation of One Proportion
   Lesson 8.2.2: Estimation of One Proportion
   Lesson 8.3.1: Estimation of One Proportion
   Lesson 8.3.2: Hypothesis Testing for One Proportion

Module 9: Inference for Two Proportions
   Lesson 9.1.1: Sampling Distribution of Differences of Two Proportions
   Lesson 9.1.2: Using Technology to Explore the Sampling Distribution of the Differences in Two Proportions
   Lesson 9.2.1: Confidence Intervals for the Difference in Two Population Proportions
   Lesson 9.2.2: Computing and Interpreting Confidence Intervals for the Difference in Two Population Proportions
   Lesson 9.3.1: A Statistical Test for the Difference in Two Population Proportions
   Lesson 9.3.2: A Statistical Test for the Difference in Two Population Proportions
   Lesson 9.3.3: Conducting a Statistical Test for the Difference in Two Population Proportions

Module 10: Inference for Means
   Lesson 10.1.1: The Sampling Distribution of the Sample Mean
   Lesson 10.1.2: Using an Applet to Explore the Sampling Distribution of the Mean with Focus on Shape
   Lesson 10.2.1: Estimating a Population Mean
   Lesson 10.2.2: $T$-Statistics and $T$-Distributions
   Lesson 10.2.3: The Confidence Interval for a Population Mean
   Lesson 10.3.1: Testing Hypotheses About a Population Mean
   Lesson 10.3.2: Test Statistic and $P$-Values, One-Sample $T$-Test
   Lesson 10.4.1: Inferences About the Difference Between Two Population Means
   Lesson 10.4.2: Inference for Paired Data
   Lesson 10.4.3: Two-Sample $T$-Test
Module 11: Chi-Squared Tests
Lesson 11.1.1: Introduction to Chi-Square Tests for One-Way Tables
Lesson 11.1.2: Executing the Chi-Square Test for One-Way Tables (Goodness-of-Fit)
Lesson 11.1.3: The Chi-Square Distribution and Degrees of Freedom
Lesson 11.2.1: Introduction to Chi-Square Tests for Two-Way Tables
Lesson 11.2.2: Executing the Chi-Square Test for Independence in Two-Way Tables
Lesson 11.2.3: Executing the Chi-Square Test for Homogeneity in Two-Way Tables

Module 12: Other Mathematical Content
Lesson 12.1.1: Statistical Linear Relationships and Mathematical Models of Linear Relationships
Lesson 12.1.2: Mathematical Linear Models
Lesson 12.1.3: Contrasting Mathematical and Statistical Linear Relationships
Lesson 12.1.4: Proportional Models
Lesson 12.2.1: Multiple Representations of Exponential Models
Lesson 12.2.2: Linear Models—Answering Various Types of Questions Algebraically
Lesson 12.2.3: Power Models
Lesson 12.2.4: Solving Inequalities
Initiating Lesson 1.1.1: The Statistical Analysis Process

Estimated number of 50-minute class sessions: 2
Tasks 1 and 2 together equal 1 to 1.5 class sessions. Task 3 fits the remaining time.

Learning Goals
Students will begin to understand
- that statistical investigations are an iterative cycle of forming a research question, collecting relevant data, data analysis, statistical inference, and drawing appropriate conclusions.
- the role that chance variability plays in the statistical decision-making process.

Students will begin to be able to
- describe the steps in the statistical analysis process.
- describe a categorical data distribution in terms of category frequencies and relative frequencies.
- construct a dotplot.
- evaluate the strength of evidence against a claim about a population proportion.

Materials Required
- Handout for the rich task (Student Handout 2)
- Three index cards per student for Task 2

Rich Task 1: Signs of the Zodiac and Personality Traits [Student Handout, estimated time: 25 minutes]

In astrology, your sign is determined by your birthday. Many people believe that the 12 signs of the zodiac correspond to 12 personality types and that characteristics of your personality are related to which sign your birthday falls under. What do you think? Are characteristics of your personality really predictable from your birthday?

To investigate this, you will look at descriptions written by astrologists of the personality traits that are supposed to be associated with each sign. The handout provided by your instructor gives three lists of personality traits for the 12 signs. For each sign, one list (Choice 1, 2, or 3) is the one that is supposed to describe people whose birthday falls in that sign; the other two lists were selected at random from the descriptions of the other 11 signs.

(Note: Students work through Questions 1–5 in small groups for about 10 minutes. This type of reasoning may be new to them, so encourage their struggle and do not jump in too quickly with the right answer. Students may have different ways to approach this, and you want to foster that and then discuss the different approaches later.)

(1) Locate the sign that corresponds to your birthday. Look at the three lists of personality traits and pick the one (Choice 1, 2, or 3) you think comes closest to describing you.

(Note: You will reveal the right answers to students much later. Right now, you want them to be comfortable with this matching process.)
(2) If the theory that personality characteristics are related to sign is not correct (meaning birth date has nothing to do with personality traits), will some classmates still pick the matching sign? About what fraction of the students in the class do you expect to pick the description that astrologists say matched their sign? Why do you think this?

(Note: You want students to realize that if there really is nothing to this theory and that all students are simply guessing at random, then in the long run about 1/3 of students will pick the correct one and 2/3 of students the wrong one. Make sure they do not focus on 1/2 because they either match or they do not. In addition, make sure students realize that if everyone is just guessing, then this fraction is pretty predictable. Students may realize that you have to make this assumption across the board, not that some signs have a better match than others.)

(3) Later in class you will see how many picked the correct sign according to the astrologists’ theory. What values for the count of classmates that picked the correct sign would you not consider convincing support for the theory that personality characteristics are related to sign?

(Note: If the number is about 1/3 of the class size, these types of outcomes are consistent with the “just-guessing” outcomes. If the number is much less than 1/3, this also indicates that people actually do worse than guessing at the matching sign. Some students may begin to realize that a number just above 1/3 of the class size is in the right direction but may not yet be large enough to be convincing. At this stage, students may just say things like 0 and 1, but encourage them to consider more than one outcome.)

(4) How many students (or what fraction) of your class need to pick the correct description for you to be convinced that there is something to this theory?

(Note: This may be a bit more difficult for students, and you may see a wide range of responses, which is fine. Some students will consider anything more than 1/3 of the class size, others may want the entire class or just about the entire class. Let them know that this is a rather subjective question for now and people may have different answers. Get students to consider whether there are outcomes above 1/3 but less than 1 that they would be surprised by. If an individual group is really struggling, ask it to consider 30 or 40 coin tosses and whether there is some number of heads in a row it needs to get before it begins to be suspicious that there is something funny going on. The point is that even in the face of randomness and chance, students can still consider some outcomes unusual and use that information to make tentative conclusions.)

(5) Suppose half the students in your class select the correct description. Does this guarantee that there is something to this theory? If not, suggest another explanation for why so many picked the correct sign.

(Note: This is a critical question for students to consider. If there are more than 1/3 matches, there are two possible reasons:

- There is something to the theory, and the data reflect this.
- Everyone was just guessing, and the class just happened by chance to have more than 1/3 of students pick the correct sign.

The key here is for students to consider the “it was just luck” explanation. If an individual group is struggling, try the legal trial analogy with it: If someone has evidence against him or her, does that...
Initiating Lesson 1.1.1: The Statistical Analysis Process

guarantee he or she is guilty? In this case, what alternative explanation could the “defense attorney” give for the higher number of matches than expected by chance?

This is the last question of the group discussion. Make sure students have thought hard about these questions before you bring them back together for group discussion. Before moving to Question 6, review and discuss the groups’ answers, not insisting on specific answers but encouraging them to consider each other’s arguments.)

(6) There are two possible explanations for why more than 1/3 of the class might choose the matching description:

- The theory is correct, and the class data reflect that the theory is correct.
- The theory is incorrect, and the class just got lucky, resulting in more than 1/3 picking the matching description.

Describe a strategy that might allow you to rule out the second explanation. ([Hint: Can you think of a way to collect new data that could help you decide whether your class data [once you see it] are consistent with the “got lucky” explanation? How can you learn about what outcomes you expect to see when there is nothing to theory?])

(Note: Let students discuss this in groups and potentially develop very different strategies that can be evaluated as a class. Make sure they are focusing on the second explanation [What will the prosecutor do in response to the argument that it was just “luck of the draw”?]. The following are some possible prompts:

How do you rule out chance? To do that, you need to know what chance looks like. So how do you get an idea for what chance looks like here? How do you create a situation where you know that all that is going on is chance? [Like assuming someone is innocent, you assume the outcome is chance until convinced otherwise.]

Bring up the coin-flipping analogy again:

How did you settle on that cutoff for the number of heads in a row that you could get by chance? How could you check that intuition? How can you determine what outcomes are unlikely to occur just by chance?

Have different groups share and critique their strategies. Keep asking what they will learn from their strategy.)

Wrap-Up for Rich Task [Instructor-led discussion]

The overall goal is to give students a sense of what a statistical analysis is all about. You can talk about the data collection plan and whether it appears sound (e.g., Are the responses by this class representative of a larger population? Is asking people to pick one of the three descriptions a good way to answer this research question? Did everyone follow the instructions?). However, the larger focus here is actually on the inferential process. Will you be able to say something about the sign theory in general based on the class results, beyond simply the class results? You know a proportion larger than 1/3 is evidence, but is it convincing enough evidence that students did not just happen to guess correctly more often than expected. You can collect pretty easily new information where you know the outcomes you are getting are purely by chance. This allows you
to decide whether the data are consistent with chance. If you can rule out chance as an
explanation for a class result higher than expected, you have stronger evidence to support the
theory. In the next tasks, you get more specific about how you can decide whether the class data
support this claim. The goal here is to expose students to the reasoning process and ideally help
them generate the reasoning process themselves. If students do not see the big picture, you can
return to the coin example; ask students whether/why they would be surprised for you to toss 10
heads in a row. They should be able to respond “because that would not happen by chance.” You
can add “assuming I was not doing anything funny when tossing the coin,” but it is the same
issue—could this result have happened by chance? If not (or highly unlikely), you consider this
evidence that something other than chance is at play.

Supporting Task 2: Ruling Out Chance [Estimated time: 35–40 minutes]
(Note: This task should be done as a whole-group activity/discussion, although it could also be done in
small groups. Use the material in the Introduction to begin this task.)

Introduction
In Task 1, you considered that when sign has nothing to do with personality traits, you expect the
fraction of students in the class who read the three personality trait lists and then pick the one
that happens to be associated with his or her sign to be around 1/3.

Discussion Question A: Why do you expect the fraction to be around 1/3? That is, why is the
fraction 1/3 used?
(Note: Students were given three descriptions to choose from. If they are choosing at random, you
expect them to choose the correct description about one out of three times.)

Discussion Question B: Why “around 1/3” and not “exactly 1/3”? 
(Note: You do not expect a class of students to hit the 1/3 mark exactly. You may get a little more
or a little less than 1/3 just by chance. The 1/3 applies over thousands and thousands of
attempts.)

To be convinced that there is a relationship between personality characteristics and sign, you
certainly need to see more than 1/3 choosing the matching description. In fact, even if students
in the class picked one of the three descriptions at random, it would not be surprising to see
somewhat more than 1/3 picking the matching description just by chance. So, for convincing
evidence that personality characteristics are related to sign, you need a fraction that is not just
greater than 1/3 but sufficiently greater than 1/3 that you do not think it could be explained just
by chance. To rule out chance as a possible explanation, you need to have some idea of what
values can be expected for the fraction of the class who pick the matching description just by
chance.

Student Handout
Next, you will investigate what kinds of outcomes are consistent with the assumption that sign has
nothing to do with personality traits. To do this, you will generate data in a way that you know for
sure that the choice of the personality trait description has nothing to do with sign.

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under
sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of
Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie
Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Initiating Lesson 1.1.1: The Statistical Analysis Process

Your instructor has provided you with three index cards. On one side of one card, write the word *match*. On one side of the other two cards, write *no match*. Turn the cards over so that your classmates cannot see what is written, mix the cards well, and then select one card at random.

Record the following information:

Number of students in the class:

Number in the class who selected the card that said *match*:

Fraction of students who chose the match card:

Proportion of students who chose the match card:

**Note:** The proportion is a decimal number between 0 and 1 (including 0 and 1) that is computed by dividing the number who chose the match card by the total number of students in the class.

(Answer: Results vary from class to class, but make sure students understand the distinctions between *count*, *fraction*, and *proportion*. Be very consistent in your language when you refer to proportions, which students will eventually mostly deal with.)

(7) Is the proportion of students in the class who picked the match card equal to 0.333? If not, is it greater than or less than 0.333?

(Answer: Results vary.)

(8) If your class repeated this process a second time, would you get exactly the same proportion of matches for the class? Why or why not?

(Answer: Students should see that the random selection of cards gives potentially different results from trial to trial.)

(9) Each classmate mixed the cards and picked one, and then you computed the proportion of students in the class that picked the match card. Explain why knowing this one proportion is not enough to tell you the values of the proportion of matches you would expect to observe just by chance when people are picking at random.

(Answer: This just tells you one value, but what you really need to see is a pattern of values. This could have been an unusual outcome. You want to look at many outcomes to be able to quantify how common/unusual different values are.)

(10) What might you do next to get a better understanding of what values of the proportion of matches would not be surprising to observe when people are picking at random? What values would be surprising in that they rarely occur if people are picking at random?

(Answer: Make sure that the discussion ends up with “we need to try this more times.”)

To understand what kinds of class match proportions are consistent with picking one of the three lists at random, you can repeat the process of generating purely “by chance” outcomes a large number of times, each time recording the class match proportion. Begin by writing the class proportion from Task 2 in the following table for Trial 1. Then work with your classmates to repeat this process until you have proportions for 10 trials.
Initiating Lesson 1.1.1: The Statistical Analysis Process

<table>
<thead>
<tr>
<th>Trial</th>
<th>Observed Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

(Note: Have students shuffle the cards, pick one card, compute the class match proportion, and enter value into table. Repeat eight times.

Introduce a way of graphically displaying the data [the dotplot] using the following material:

A convenient way to summarize all the observed proportions in the table is to construct a graphical display. You will learn about graphical displays in Module 2, but for now we will construct a simple graphical display called a *dotplot*. In a dotplot, each numerical value is plotted as a point along a number line.

Plot the 10 observed class proportions on the board for students to see.)

(11) The dotplot constructed by the class shows observed match proportions for 10 trials where students are picking one of three choices at random.

(a) What was the smallest match proportion observed?

(b) What was the largest match proportion observed?

(c) Did the match proportion differ much from trial to trial?

To really understand how the observed proportion varies from one trial to another, you need to look at many more trials. This can be a tedious process, so you will turn to technology to perform more trials.

(Note: *This applet is under development.* You will use an applet to carry out many more trials and build up a dotplot that describes chance variability in the match proportion. To use the applet, enter the class size and then click the First Trial button. This shows the cards shuffling and the total number of matches and the proportion of matches; it adds a dot to the dotplot of the proportions. Then click the Repeat 1 Time button a couple of times to add dots to the dotplot. Talk about what is happening after each trial. Do this a few times until students understand what is happening. Then move to the Repeat 10 Times button and click this button until you have at least
Initiating Lesson 1.1.1: The Statistical Analysis Process

100 observed values in the dotplot. If students have access to a computer in class, they can do this part themselves.)

(Answer: Results vary by class size but may look something like this, centered around 1/3.

Confirm with students that they understand that each dot represents the proportion of matches in one set of \( n \) picks \([n = \text{class size}]\) where each pick was made completely at random.)

(Note: Once the applet has been used to create the dotplot, this next part can be done as a whole-class discussion, as a minilecture, or in small groups. Be sure that the following questions are addressed.)

(12) Summarize what your dotplot tells you. Why did you create this graph?

(Answer: Remind students that they wanted to know about the chance variation in the proportion of matches under the condition that everyone was picking at random.)

(13) How does this graph tell you whether the actual class data (where students picked the description that they thought was most accurate to test the astrology theory) support the sign theory?

(Answer: This graph gives you something to compare the class’s number of matches to. If the class proportion is not consistent with the outcomes that occur by chance alone, you have evidence that something else was behind the data. If the class proportion looks like the chance proportions, you can conclude that there is not anything to the sign theory—the result was also just due to chance.)

(14) How would you complete the following sentence?

If students in the class were picking at random, it would be unusual to see a match proportion as large as ______________.

How does the number you used to complete the sentence relate to what you see in the dotplot?

(Answer: Class results vary, but they should be out in the “tail” of the distribution generated in the applet, like above 0.45 or 0.5.)

(15) Suppose the match proportion was 0.40 when your classmates looked at the personality characteristic lists and selected the one that they thought best described themselves. Is this consistent with thinking that personality characteristics are unrelated to sign? How can you tell? 

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s front matter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
(Answer: Although 0.4 is above 0.333, the dotplot from the applet shows that it is not extremely far above. It does not appear to be unusual to get 40% correct matches even when you know all the selections are made completely at random, with no correspondence between the sign and personality characteristics. From the dotplot, students should get a sense that 0.4 is closer to the middle of the distribution rather than way out in the tail.

Some students may go on to count how often that happens. For example, in the previous example results, you get 40% or more about 16% of the time. You may not want to take students too far down this path right now as they will want to know why you look at or more and where the cutoff is. Why is 15% not unusual? What would be unusual?)

(16) How large does the match proportion have to be to convince you that personality characteristics might be related to sign? How is this related to your answer to Question 13?

(Answer: If you are around 0.45 or 0.5 or higher, this is a very unusual outcome when the personality characteristics are not related to sign [according to the previous graph that showed the outcomes under random chance alone]. So, if the class results are that large, this gives evidence that students did better than expected just by chance and there could be something to the sign theory.)

Making a Decision! [Student Handout]

Now it is time to take a look at the actual class data and see whether you think that it provides evidence for a relationship between personality traits and sign. Which of the three choices for your sign did you pick?

Choice that you picked as best description:

Choice that is a match for your sun sign:
(Your instructor will provide this information.)

Number of students in the class:

Number of matches in the class:

Match proportion for the class:

(Answer: Correct matches are

<table>
<thead>
<tr>
<th>Sign</th>
<th>Choice 1</th>
<th>Choice 2</th>
<th>Choice 2</th>
<th>Choice 3</th>
<th>Choice 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taurus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gemini</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virgo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(17) Now it all comes down to this! Does the class match proportion provide convincing evidence that personality characteristics could be related to sign? Why or why not? Use the simulation results in your reasoning!

(Answer: Students need to compare the class results to the dotplot generated by the applet and see whether the class results are in the tail of the distribution. If the class proportion is in the tail, this provides evidence that personality characteristics might be related to sign or at least convincing evidence that random chance alone is not a sufficient counter argument.)

Wrap-Up [Instructor-led discussion]
Discuss the overall process students have experienced here—state a theory, collect data to test the theory, and consider what kind of evidence provides support for this theory. Students then evaluated two possible explanations—the theory is correct or the theory is not correct and the observed result is due to chance variability. To conclude that there is evidence in support of the theory, they need to rule out chance as a plausible explanation for what was observed. Students do this by generating results assuming the theory is false and what they are seeing is only due to chance variability; they then compare the observed results to see whether they can be considered consistent with chance.

Task 3: Statistical Investigations [Student Handout, estimated time: 30–40 minutes]
The tasks you have just completed with the astrology example illustrate a general process that is common to many statistical investigations. One way to describe this process is as a sequence of four steps, as shown in the table below.

Steps in a Statistical Investigation

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ask a question that can be answered by collecting data.</td>
</tr>
<tr>
<td>2.</td>
<td>Decide what to measure and then collect data.</td>
</tr>
<tr>
<td>3.</td>
<td>Summarize and analyze the data.</td>
</tr>
<tr>
<td>4.</td>
<td>Draw a conclusion and communicate the results.</td>
</tr>
</tbody>
</table>

Keep in mind that this is an iterative process. Analyzing the results of one study often leads researchers to consider other research questions and then to conduct additional research. Thinking more carefully about how the data were collected may also lead them to consider ways to improve the data collection process.

(Note: You may use whole-group discussion to fill in the following table for the astrology example. Answers are italicized and do not appear in the Student Handout.)
Identify each step for the astrology investigation to complete the following table.

<table>
<thead>
<tr>
<th>Steps in a Statistical Investigation</th>
<th>For the Astrology Investigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ask a question that can be answered by collecting data.</td>
<td>Are personality characteristics related to sign?</td>
</tr>
<tr>
<td>2. Decide what to measure and then collect data.</td>
<td>Measure whether students select the description that is supposed to match their sign when presented with three different lists of personality characteristics. For each student, whether the chosen description was a match was recorded.</td>
</tr>
<tr>
<td>3. Summarize and analyze the data.</td>
<td>The total number of matches for students in the class was determined, and then the proportion of matches was computed. We then decided whether it was likely that the observed proportion was by chance when there was not a relationship between personality characteristics and sign.</td>
</tr>
<tr>
<td>4. Draw a conclusion and communicate the results.</td>
<td>There is/is not evidence that personality characteristics are related to sign among students.</td>
</tr>
</tbody>
</table>
Initiating Lesson 1.1.1: The Statistical Analysis Process

(19) Look at the following two study descriptions. For each study, identify the four steps of the statistical investigation process to complete the table on the following page.

**Study 1**
Researchers at the Center for Reproductive Medicine at Brigham and Women’s Hospital wondered what proportion of women who visit a fertility clinic would want the opportunity to choose the sex of a future child. They also wondered if those that would like to choose the sex were more likely to want a boy or girl. The researchers mailed a survey containing 19 questions to women who had visited the Center. One question on the survey asked women whether they would like the option of choosing the sex of a future child. If the response to that question was yes, a follow-up question asked whether they would choose a boy or girl. Of the 229 women who wanted to choose, 89 (38.9%) said they would choose a boy and 140 (61.1%) said they would choose a girl. Based on their statistical analysis of these data, the researchers concluded that there is convincing evidence of an overall preference for girls among women wanting to choose the sex of a future child. This conclusion is based on the fact that it is very unusual to observe a percentage as high as 61.1% in a sample of size 229 women just by chance if there really is no preference for girls in the population of women who would like to select the sex of a future child.

**Study 2**
Psychologists believe that people are less likely to do something if they think it will require a lot of effort. But how do people decide what things they think will be hard and what things they think will be easy? Researchers at the University of Michigan wondered whether how difficult it was to read the instructions for how to perform a task would affect how hard people thought the task would be. To investigate this, they performed an experiment. Twenty students were divided at random into two groups of 10 students each. One group received instructions for an exercise routine printed in a font that was easy to read, and the other group received the same set of instructions printed in a font that was difficult to read. A sample of each font appears below. Each student read the instructions, and then they were asked how many minutes they thought the exercise routine would take. For the group that read the instructions printed in an easy-to-read font, the average for the number of minutes they thought the routine would take was 8.23. For the group that read the same instructions printed in the font that was hard to read, the average was 15.1 minutes. Based on the study data, the researchers concluded that the difference between these two averages was not likely to be due to chance and that there was evidence that people think a task will be harder when the instructions are hard to read.

This is the easy-to-read font that was used in the study.
This is the hard-to-read font that was used in the study.
(Note: You may want to have students read the description of Study 1 and then talk about it as a class to make sure that they understand the scenario described. Students could then fill in the Study 1 column. The process can then be repeated for Study 2. This is probably easier for students than asking them to read both studies at once. Answers are italicized and do not appear in the Student Handout.)

<table>
<thead>
<tr>
<th>Steps in a Statistical Investigation</th>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ask a question that can be answered by collecting data.</td>
<td>Do women visiting fertility clinics who want to choose the sex of the child tend to choose one gender more than the other?</td>
<td>Do people think a task will be harder if the instructions for the task are harder to read.</td>
</tr>
<tr>
<td>2. Decide what to measure and then collect data.</td>
<td>Surveys were sent to women to see which sex option they would pick.</td>
<td>University of Michigan students were asked to estimate how long a task would take using different fonts, where one was harder to read.</td>
</tr>
<tr>
<td>3. Summarize and analyze the data.</td>
<td>61.1% of 229 respondents said they would choose a girl.</td>
<td>The average time estimate for the task with the hard-to-read font was 15.1 minutes compared to 8.23 minutes with the easier-to-read font.</td>
</tr>
<tr>
<td>4. Draw a conclusion and communicate the results.</td>
<td>There is evidence of an overall preference for girls.</td>
<td>There is evidence that people think a task will be harder when the instructions are harder to read.</td>
</tr>
</tbody>
</table>

(20) Even though these two studies both follow the same general process, they are different in some ways. What are two ways that these studies are different?

(Note: Focus on the type of question that is asked [one asks about characteristics of a population, and one asks about the effect of some treatment] and the manner the data were collected [sampling versus experimentation]. This leads in to the next lessons.)

Lesson Wrap-Up [Instructor-led discussion]

The main point here is for students to get an overview of the types of studies they will see and will be able to analyze. Emphasize that the reasoning process to judge whether a result is significant follows very similar lines to the astrology study. Also emphasize that the other steps (deciding what to measure, obtaining valid measurements, reporting conclusions) are also critical to this process and should always be given careful attention as well.
Initiating Lesson 1.1.1: The Statistical Analysis Process

Homework

Introduction
The following is an introduction to the Homework section that you can choose to deliver to the class:

For the astrology investigation, you looked at how you could decide whether the data provided evidence in support of the claim that personality characteristics are related to sign. To do this, you looked at whether students in your class were more likely to match the claimed description that they would expect by chance alone when there was no relationship between personality characteristics and sign. In the lesson, hypothetical data were generated in a way that you could be sure that there was no relationship between personality characteristics and sign by having students pick one of the three descriptions at random and then noting the proportion that happened to pick the matching description just by chance. You compared these hypothetical results to the actual result for the class to see whether the higher proportion in the class was consistent with that chance variation. If the class result was unlikely to occur by chance alone, this gave some evidence to the sign theory.

Student Handout

(1) Suppose the investigation had given each student a choice of four personality descriptions for each sign. If there is no relationship between personality characteristics and sign, about what fraction of the students in the class do you expect to pick the description that astrologists say matched their sign? Why do you think this?

(Answer: Now there is one correct choice of four matches, so a correct match occurs by chance about one time in four, for a match proportion [in the long-run] of about 0.25.)

(2) One dotplot shown below was constructed by asking 40 students to pick one of four descriptions at random (using index cards) and then computing the proportion that chose the description matching their sign just by chance. This process was repeated a large number of times to generate the data used to construct the dotplot. Which dotplot do you think is the one that was constructed in this way? Why did you pick this dotplot?
Initiating Lesson 1.1.1: The Statistical Analysis Process

(Answer: These data represent what happens when people are just guessing, so the proportions should center around 0.25. The correct choice is Dotplot 1.)

(3) Suppose each of these 40 students then made their choices from the list of four personality types. What proportion of the 40 students need to match correctly to provide convincing evidence that there is a connection between sign and personality type? (Hint: Use your answer from Question 2.)

(Answer: Results vary a bit, but a proportion of 0.35 or larger seems reasonable as this puts the observed proportion in the tail of the distribution [Dotplot 1]. With Dotplot 2, they might say anything above 0.6.)

(4) Explain your reasoning for your choice in Question 3.

(Answer: The reasoning should look at comparing the observed proportion to those values generated by chance alone. This helps students decide whether an observed proportion is larger than what they expect to happen by chance. They should focus on a value that appears inconsistent with the “by chance alone” results.)

(5) Consider the following study:

Researchers at Minnesota State University wanted to learn whether middle-aged adults who used the Wii Fit video games actually exercised with enough intensity to meet fitness recommendations found in the U.S. Surgeon General report on physical activity. The 20 middle-aged adults participating in the study were given training on the Wii Fit video games. On the day after the training session, participants completed a 20-minute Wii Fit session. Total energy expenditure (in calories) was measured. The average energy expenditure was computed to be 116 calories for the 20-minute session. The energy expenditure recommended to stay fit is 150 to 400 calories per day. Based on the results of the study, the researchers concluded that the Wii Fit video games could be an effective form of exercise for middle-aged adults, but that to meet this requirement, the length of the exercise session should be increased from 20 minutes to 30 minutes.

(a) Identify the four steps of the statistical investigation process for the study in the table.

(Note: Answers are italicized and do not appear in the Student Handout.)

<table>
<thead>
<tr>
<th>Steps in a Statistical Investigation</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ask a question that can be answered by collecting data.</td>
<td>Does the Wii Fit video game burn enough calories to be considered suitable exercise?</td>
</tr>
<tr>
<td>2. Decide what to measure and then collect data.</td>
<td>The total energy expenditure was gathered for a 20-minute session.</td>
</tr>
<tr>
<td>3. Summarize and analyze the data.</td>
<td>The average energy expenditure was 116 calories. This is below the recommended 150 to 400 calories.</td>
</tr>
<tr>
<td>4. Draw a conclusion and communicate the results.</td>
<td>The Wii Fit video game does not appear to burn enough calories in a 20-minute session, but a 30-minute session would possibly be enough.</td>
</tr>
</tbody>
</table>
(b) Based on the results of this study, what next steps do you recommend to the researchers investigating this issue.

(Answer: Researchers could next measure the 30-minute sessions to see whether the calories burned are now in line with the recommendations. They could also try this with a larger and/or more diverse group of subjects [advantages to this are discussed in the next lesson]. Students may also suggest expanding beyond middle-age adults.)
Task 1: Signs of the Zodiac and Personality Traits

In astrology, your sign is determined by your birthday. Many people believe that the 12 signs of the zodiac correspond to 12 personality types and that characteristics of your personality are related to which sign your birthday falls under. What do you think? Are characteristics of your personality really predictable from your birthday?

To investigate this, you will look at descriptions written by astrologists of the personality traits that are supposed to be associated with each sign. The handout provided by your instructor gives three lists of personality traits for the 12 signs. For each sign, one list (Choice 1, 2, or 3) is the one that is supposed to describe people whose birthday falls in that sign; the other two lists were selected at random from the descriptions of the other 11 signs.

(1) Locate the sign that corresponds to your birthday. Look at the three lists of personality traits and pick the one (Choice 1, 2, or 3) you think comes closest to describing you.

(2) If the theory that personality characteristics are related to sign is not correct (meaning birth date has nothing to do with personality traits), will some classmates still pick the matching sign? About what fraction of the students in the class do you expect to pick the description that astrologists say matched their sign? Why do you think this?

(3) Later in class you will see how many picked the correct sign according to the astrologists’ theory. What values for the count of classmates that picked the correct sign would you not consider convincing support for the theory that personality characteristics are related to sign?

(4) How many students (or what fraction) of your class need to pick the correct description for you to be convinced that there is something to this theory?
(5) Suppose half the students in your class select the correct description. Does this guarantee that there is something to this theory? If not, suggest another explanation for why so many picked the correct sign.

(6) There are two possible explanations for why more than 1/3 of the class might choose the matching description:

- The theory is correct, and the class data reflect that the theory is correct.
- The theory is incorrect, and the class just got lucky, resulting in more than 1/3 picking the matching description.

Describe a strategy that might allow you to rule out the second explanation. (Hint: Can you think of a way to collect new data that could help you decide whether your class data [once you see it] are consistent with the “got lucky” explanation? How can you learn about what outcomes you expect to see when there is nothing to theory?)
Initiating Lesson 1.1.1: The Statistical Analysis Process

Task 2: Ruling Out Chance

Next, you will investigate what kinds of outcomes are consistent with the assumption that sign has nothing to do with personality traits. To do this, you will generate data in a way that you know for sure that the choice of the personality trait description has nothing to do with sign.

Your instructor has provided you with three index cards. On one side of one card, write the word *match*. On one side of the other two cards, write *no match*. Turn the cards over so that your classmates cannot see what is written, mix the cards well, and then select one card at random.

Record the following information:

- Number of students in the class:
- Number in the class who selected the card that said *match*:
- Fraction of students who chose the match card:
- Proportion of students who chose the match card:

**Note:** The proportion is a decimal number between 0 and 1 (including 0 and 1) that is computed by dividing the number who chose the match card by the total number of students in the class.

(7) Is the proportion of students in the class who picked the match card equal to 0.333? If not, is it greater than or less than 0.333?

(8) If your class repeated this process a second time, would you get exactly the same proportion of matches for the class? Why or why not?

(9) Each classmate mixed the cards and picked one, and then you computed the proportion of students in the class that picked the match card. Explain why knowing this one proportion is not enough to tell you the values of the proportion of matches you would expect to observe just by chance when people are picking at random.
(10) What might you do next to get a better understanding of what values of the proportion of matches would not be surprising to observe when people are picking at random? What values would be surprising in that they rarely occur if people are picking at random?

To understand what kinds of class match proportions are consistent with picking one of the three lists at random, you can repeat the process of generating purely “by chance” outcomes a large number of times, each time recording the class match proportion. Begin by writing the class proportion from Task 2 in the following table for Trial 1. Then work with your classmates to repeat this process until you have proportions for 10 trials.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Observed Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

(11) The dotplot constructed by the class shows observed match proportions for 10 trials where students are picking one of three choices at random.

(a) What was the smallest match proportion observed?

(b) What was the largest match proportion observed?

(c) Did the match proportion differ much from trial to trial?
Initiating Lesson 1.1.1: The Statistical Analysis Process

To really understand how the observed proportion varies from one trial to another, you need to look at many more trials. This can be a tedious process, so you will turn to technology to perform more trials.

(12) Summarize what your dotplot tells you. Why did you create this graph?

(13) How does this graph tell you whether the actual class data (where students picked the description that they thought was most accurate to test the astrology theory) support the sign theory?

(14) How would you complete the following sentence?

If students in the class were picking at random, it would be unusual to see a match proportion as large as _____________.

How does the number you used to complete the sentence relate to what you see in the dotplot?

(15) Suppose the match proportion was 0.40 when your classmates looked at the personality characteristic lists and selected the one that they thought best described themselves. Is this consistent with thinking that personality characteristics are unrelated to sign? How can you tell?
Initiating Lesson 1.1.1: The Statistical Analysis Process

(16) How large does the match proportion have to be to convince you that personality characteristics might be related to sign? How is this related to your answer to Question 13?

Making a Decision!

Now it is time to take a look at the actual class data and see whether you think that it provides evidence for a relationship between personality traits and sign. Which of the three choices for your sign did you pick?

Choice that you picked as best description:

Choice that is a match for your sun sign:
(Your instructor will provide this information.)

Number of students in the class:

Number of matches in the class:

Match proportion for the class:

(17) Now it all comes down to this! Does the class match proportion provide convincing evidence that personality characteristics could be related to sign? Why or why not? Use the simulation results in your reasoning!
Task 3: Statistical Investigations

The tasks you have just completed with the astrology example illustrate a general process that is common to many statistical investigations. One way to describe this process is as a sequence of four steps, as shown in the table below.

<table>
<thead>
<tr>
<th>Steps in a Statistical Investigation</th>
<th>For the Astrology Investigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ask a question that can be answered by collecting data.</td>
<td></td>
</tr>
<tr>
<td>2. Decide what to measure and then collect data.</td>
<td></td>
</tr>
<tr>
<td>3. Summarize and analyze the data.</td>
<td></td>
</tr>
<tr>
<td>4. Draw a conclusion and communicate the results.</td>
<td></td>
</tr>
</tbody>
</table>

Keep in mind that this is an iterative process. Analyzing the results of one study often leads researchers to consider other research questions and then to conduct additional research. Thinking more carefully about how the data were collected may also lead them to consider ways to improve the data collection process.

(18) Identify each step for the astrology investigation to complete the following table.
(19) Look at the following two study descriptions. For each study, identify the four steps of the statistical investigation process to complete the table on the following page.

### Study 1
Researchers at the Center for Reproductive Medicine at Brigham and Women’s Hospital wondered what proportion of women who visit a fertility clinic would want the opportunity to choose the sex of a future child. They also wondered if those that would like to choose the sex were more likely to want a boy or girl. The researchers mailed a survey containing 19 questions to women who had visited the Center. One question on the survey asked women whether they would like the option of choosing the sex of a future child. If the response to that question was yes, a follow-up question asked whether they would choose a boy or girl. Of the 229 women who wanted to choose, 89 (38.9%) said they would choose a boy and 140 (61.1%) said they would choose a girl. Based on their statistical analysis of these data, the researchers concluded that there is convincing evidence of an overall preference for girls among women wanting to choose the sex of a future child. This conclusion is based on the fact that it is very unusual to observe a percentage as high as 61.1% in a sample of size 229 women just by chance if there really is no preference for girls in the population of women who would like to select the sex of a future child.

### Study 2
Psychologists believe that people are less likely to do something if they think it will require a lot of effort. But how do people decide what things they think will be hard and what things they think will be easy? Researchers at the University of Michigan wondered whether how difficult it was to read the instructions for how to perform a task would affect how hard people thought the task would be. To investigate this, they performed an experiment. Twenty students were divided at random into two groups of 10 students each. One group received instructions for an exercise routine printed in a font that was easy to read, and the other group received the same set of instructions printed in a font that was difficult to read. A sample of each font appears below. Each student read the instructions, and then they were asked how many minutes they thought the exercise routine would take. For the group that read the instructions printed in an easy-to-read font, the average for the number of minutes they thought the routine would take was 8.23. For the group that read the same instructions printed in the font that was hard to read, the average was 15.1 minutes. Based on the study data, the researchers concluded that the difference between these two averages was not likely to be due to chance and that there was evidence that people think a task will be harder when the instructions are hard to read.

This is the easy-to-read font that was used in the study.

This is the hard-to-read font that was used in the study.
### Initiating Lesson 1.1.1: The Statistical Analysis Process

<table>
<thead>
<tr>
<th>Steps in a Statistical Investigation</th>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ask a question that can be answered by collecting data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Decide what to measure and then collect data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Summarize and analyze the data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Draw a conclusion and communicate the results.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(20) Even though these two studies both follow the same general process, they are different in some ways. What are two ways that these studies are different?
Initiating Lesson 1.1.1: The Statistical Analysis Process

Homework

(1) Suppose the investigation had given each student a choice of four personality descriptions for each sign. If there is no relationship between personality characteristics and sign, about what fraction of the students in the class do you expect to pick the description that astrologists say matched their sign? Why do you think this?

(2) One dotplot shown below was constructed by asking 40 students to pick one of four descriptions at random (using index cards) and then computing the proportion that chose the description matching their sign just by chance. This process was repeated a large number of times to generate the data used to construct the dotplot. Which dotplot do you think is the one that was constructed in this way? Why did you pick this dotplot?
(3) Suppose each of these 40 students then made their choices from the list of four personality types. What proportion of the 40 students need to match correctly to provide convincing evidence that there is a connection between sign and personality type? (Hint: Use your answer from Question 2.)

(4) Explain your reasoning for your choice in Question 3.
(5) Consider the following study:

Researchers at Minnesota State University wanted to learn whether middle-aged adults who used the Wii Fit video games actually exercised with enough intensity to meet fitness recommendations found in the U.S. Surgeon General report on physical activity. The 20 middle-aged adults participating in the study were given training on the Wii Fit video games. On the day after the training session, participants completed a 20-minute Wii Fit session. Total energy expenditure (in calories) was measured. The average energy expenditure was computed to be 116 calories for the 20-minute session. The energy expenditure recommended to stay fit is 150 to 400 calories per day. Based on the results of the study, the researchers concluded that the Wii Fit video games could be an effective form of exercise for middle-aged adults, but that to meet this requirement, the length of the exercise session should be increased from 20 minutes to 30 minutes.

(a) Identify the four steps of the statistical investigation process for the study in the table.

<table>
<thead>
<tr>
<th>Steps in a Statistical Investigation</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ask a question that can be answered by collecting data.</td>
<td></td>
</tr>
<tr>
<td>2. Decide what to measure and then collect data.</td>
<td></td>
</tr>
<tr>
<td>3. Summarize and analyze the data.</td>
<td></td>
</tr>
<tr>
<td>4. Draw a conclusion and communicate the results.</td>
<td></td>
</tr>
</tbody>
</table>

(b) Based on the results of this study, what next steps do you recommend to the researchers investigating this issue.
## Initiating Lesson 1.1.1: The Statistical Analysis Process

### Astrology Investigation

<table>
<thead>
<tr>
<th>Zodiac Sign</th>
<th>Choice 1</th>
<th>Choice 2</th>
<th>Choice 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strengths</td>
<td>Weaknesses</td>
<td>Strengths</td>
</tr>
</tbody>
</table>
## Initiating Lesson 1.1.1: The Statistical Analysis Process

<table>
<thead>
<tr>
<th>Zodiac Sign</th>
<th>Choice 1</th>
<th>Choice 2</th>
<th>Choice 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strengths</td>
<td>Weaknesses</td>
<td>Strengths</td>
</tr>
<tr>
<td>Cancer (6/22 to 7/22)</td>
<td>Emotional</td>
<td>Changeable</td>
<td>Patient</td>
</tr>
<tr>
<td></td>
<td>Loving</td>
<td>Moody</td>
<td>Reliable</td>
</tr>
<tr>
<td></td>
<td>Intuitive</td>
<td>Overemotional</td>
<td>Warmhearted</td>
</tr>
<tr>
<td></td>
<td>Imaginative</td>
<td>Touchy</td>
<td>Loving</td>
</tr>
<tr>
<td></td>
<td>Shrewd</td>
<td>Clinging</td>
<td>Persistent</td>
</tr>
<tr>
<td></td>
<td>Cautious</td>
<td>Unable to let go</td>
<td>Determined</td>
</tr>
<tr>
<td></td>
<td>Protective</td>
<td></td>
<td>Placid</td>
</tr>
<tr>
<td></td>
<td>Sympathetic</td>
<td></td>
<td>Placid</td>
</tr>
<tr>
<td>Leo (7/23 to 8/21)</td>
<td>Emotional</td>
<td>Changeable</td>
<td>Generous</td>
</tr>
<tr>
<td></td>
<td>Loving</td>
<td>Moody</td>
<td>Warmhearted</td>
</tr>
<tr>
<td></td>
<td>Intuitive</td>
<td>Overemotional</td>
<td>Creative</td>
</tr>
<tr>
<td></td>
<td>Imaginative</td>
<td>Touchy</td>
<td>Enthusiastic</td>
</tr>
<tr>
<td></td>
<td>Shrewd</td>
<td>Clinging</td>
<td>Broad-minded</td>
</tr>
<tr>
<td></td>
<td>Cautious</td>
<td>Unable to let go</td>
<td>Expansive</td>
</tr>
<tr>
<td></td>
<td>Protective</td>
<td></td>
<td>Faithful</td>
</tr>
<tr>
<td></td>
<td>Sympathetic</td>
<td></td>
<td>Loving</td>
</tr>
<tr>
<td>Virgo (8/22 to 9/23)</td>
<td>Practical</td>
<td>Pessimistic</td>
<td>Adaptable</td>
</tr>
<tr>
<td></td>
<td>Prudent</td>
<td>Fatalistic</td>
<td>Versatile</td>
</tr>
<tr>
<td></td>
<td>Ambitious</td>
<td>Miserable</td>
<td>Communicative</td>
</tr>
<tr>
<td></td>
<td>Disciplined</td>
<td>Grudging</td>
<td>Witty</td>
</tr>
<tr>
<td></td>
<td>Patient</td>
<td></td>
<td>Intellectual</td>
</tr>
<tr>
<td></td>
<td>Careful</td>
<td></td>
<td>Eloquent</td>
</tr>
<tr>
<td></td>
<td>Humorous</td>
<td></td>
<td>Youthful</td>
</tr>
<tr>
<td></td>
<td>Reserved</td>
<td></td>
<td>Lively</td>
</tr>
</tbody>
</table>
### Initiating Lesson 1.1.1: The Statistical Analysis Process

<table>
<thead>
<tr>
<th>Zodiac Sign</th>
<th>Choice 1: Strengths</th>
<th>Weaknesses</th>
<th>Choice 2: Strengths</th>
<th>Weaknesses</th>
<th>Choice 3: Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
</table>

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
## Initiating Lesson 1.1.1: The Statistical Analysis Process

<table>
<thead>
<tr>
<th>Zodiac Sign</th>
<th>Choice 1</th>
<th>Choice 2</th>
<th>Choice 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strengths</td>
<td>Weaknesses</td>
<td>Strengths</td>
</tr>
</tbody>
</table>

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Estimated number of 50-minute class sessions: 2

Learning Goals
Students will begin to understand

- the role that appropriate data collection methods play in the types of conclusions they can draw from data.
- the importance of clearly stating the question to be answered by a statistical study and deciding what data will be collected to answer the question.
- the differences between an observational study and an experiment.
- the applicability of statistical methods to a variety of disciplines.

Students will begin to be able to

- given a description of a statistical study, identify the steps in the statistical analysis process.
- determine when it is reasonable to generalize conclusions from a statistical study.
- determine whether a statistical study is an observational study or an experiment.
- after reviewing the study design, critique the conclusions drawn in a study.

Task 1: Research Questions that Can Be Answered by Collecting Data [Student Handout, estimated time: 15 minutes]

(Note: Task 1 can be carried out with students working individually or with a partner, or it can be done as a whole-group discussion. Give students time to think of their own research questions. Then ask them to share their responses with the class. Model good answers and help students see some of the subtle distinctions in correct and incorrect responses. The first three tasks in this lesson can probably be completed in a 50-minute class session. The main goal is to expose students to the types of research questions they will be able to analyze; another goal is encouraging them to think about the need to carefully collect data so that they can draw legitimate conclusions from it.)

In Lesson 1.1.1, a four-step process that is common to many statistical studies was introduced. Step 1 is, “Ask a question that can be answered by collecting data.” The questions that are the focus of statistical studies are usually one of the following types: research questions about populations or research questions about the effect of some variable on a response.

Research Questions About Populations

A population is the entire group of individuals or objects that you are interested in learning about. Usually, it is not possible to study the entire population, so you collect data on just a part of the population. For example, if you were interested in learning about the proportion of households in Boston that routinely recycle plastic containers, it would be both time consuming and costly to collect data from every household. You might only be able to collect data for 1,000 households in Boston, but if the 1,000 households were selected carefully, you can still learn a lot about plastic recycling in Boston. The part of the population for which data are collected is called a sample.
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

(Note: If you think that students do not see the distinction between population and sample, show them the following figures:

Population

Sample Shown in Black

Sample"
Research questions about populations might involve

- estimating a numerical population characteristic or testing a claim about a population (e.g., “Do more than half of all households in Boston recycle plastic containers?”).
- comparing two or more populations (e.g., “Is plastic recycling more common in Los Angeles than in Boston?”).
- asking whether two variables are related in the population of interest (e.g., “Is there a relationship between household income and which households in Boston recycle plastics?”)

**Research Questions About the Effect of Some Variable on a Response**

This type of question can take the form, “What is the effect of ...?” or “What happens when...?” For example, “What is the effect of room temperature on exam performance?” or “What happens to the distance a car can travel on 1 gallon of gasoline if 100 pounds of additional weight are put in the trunk of the car?”

**Additional Examples of Research Questions About Populations**

**Instructor-Led Introduction**

Another example of a research question about a population is, “What is the average amount of money spent of textbooks per semester for full-time students at your college?” Here the population of interest is all full-time students at your college, and you could collect data on a representative sample of students at your school that allows you to estimate this characteristic (the average amount of money spent on textbooks) for the entire population.

**Student Handout**

(1) Give an example of a research question that involves estimating a population characteristic where the population of interest is all full-time students at your college.

**Answer:** Many answers are possible. Most students will find it straightforward to talk about estimating a mean or a proportion. Make sure their responses are clearly about all full-time students at the school.

- What is the average commute time for full-time students at the school?
- What proportion of full-time students at the school are employed at least 20 hours a week off-campus?

Make sure students present a research question about a numerical population characteristic [e.g., a population average or a population proportion] and not only a variable [e.g., “How long do students commute to school?” or “What is your commute time?” or “Do students work off-campus?”]. Encourage them to insert words such as *on average or typical*.

Students may also be interested in extremes in the data set [e.g., “What is the longest commute time by a student at this school?”]. Such values can also be estimated based on a sample of the population.

(2) Give an example of a research question that involves estimating a population characteristic for a different population. What is the population of interest for your question?
**Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions**

(Answer: Many students will take the research question from Question 1 and change the population of interest, such as “all college students in the state” or “all full- and part-time students” or “all students and faculty at the school.”)

**Instructor-Led Introduction**

Another example of a research question about a population is, “Is there evidence that more than half of the students at your college registered to vote?” Here the population of interest is all students at your college (not just full-time students, like in Question 1). You could collect data that allows you to decide whether there is convincing evidence to support the claim that more than half of the students at the college are registered to vote. Here you are not focused on estimating the actual value of the population proportion, but rather on how it compares to some hypothesized value. Or you may compare the same numerical characteristic but for two different populations.

**Student Handout**

(3) Give an example of a research question that involves testing a claim about a population characteristic where the population of interest is all students at your college.

(Answer: If students are struggling, point them toward thinking about a tendency or majority in their comparison [e.g., “Do most students at this school commute less than 30 minutes?” or “Do students at this school tend to work more than 20 hours per week off-campus?”]. Help students see the distinction between estimating how many hours students work on average and comparing this average to a particular value like 20, though this is not a major consideration. Mostly make sure students are making statements about a larger population [versus what was the average commute time for students that were surveyed] and are going beyond describing the variable measured of each individual [e.g., “How long is a student’s commute time?” versus “How long is a typical commute time for students at this school?”].)

(4) Give an example of a research question that involves testing a claim comparing two populations. What are the populations of interest for your question?

(Answer: Encourage students to compare two populations: “Do males tend to work more hours off-campus than females?” or “Are upperclassmen more likely to work off-campus than lowerclassmen?” Make sure the populations of interest are clearly defined.)

**Instructor-Led Introduction**

Another example of a research question about a population is, “Is there a relationship between grade point average and the number of hours spent studying in a typical week for students at your college?” By collecting data on both grade point average and the number of hours studied for one representative sample of students at your college, you could investigate whether there is convincing evidence of a relationship between these two variables in the population as a whole.
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Student Handout

(5) Give an example of a research question that involves investigating the relationship between two variables where the population of interest is all students at your college.

(Answer: If students are struggling, remind them that they should think about asking each person in the sample two questions. The research question is then simply whether there is a relationship between those two variables [i.e., Does knowing the outcome of one variable help you predict the other?]. For example, “Is there a relationship between high school GPA and first-year college GPA for students at this school?” or “Is there a relationship between playing collegiate sports and number of units taken this quarter?”

Discourage students from making causal-sounding statements such as, “Belonging to a fraternity lowers GPA.” This type of research question is discussed next. It may be worth keeping a list on the board of action words like causes, leads to, and enhances that many interpret as sounding casual. Words like relationship, association, connection, and even more likely are less likely to be interpreted this way.)

(6) Give an example of a research question that involves investigating the relationship between two variables for a different population. What is the population of interest for your question?

Additional Examples of Research Questions About the Effect of Some Variable on a Response

Instructor-Led Introduction

An example of a research question about the effect of some variable on a response is, “Does the font size of text affect how quickly a person can read a paragraph?” This could be rephrased as, “What is the effect of font size on the time required to read a paragraph?” or “Does the font size make a difference in reading times?” or “Does a paragraph tend to take longer to read because it is in a smaller font?” Here the response you are interested in is the time to read a paragraph and the variable that you think might affect the response is font size. You could gather data that allows you to answer this question by asking some people to read a paragraph that is printed in a small font size. You then ask other people to read the same paragraph of text, but where the paragraph is printed in a large font size. If you measure how long it takes people to read the paragraph, you then compare the times for those that read the small font and the times for those that read the large font to draw a conclusion about whether the font size had an effect on reading time.

Note: If you are collecting students’ written responses, watch for the proper use of effect (the noun) and affect (the verb).

Student Handout

(7) Give an example of a research question about the effect of some variable on a response that you think would be interesting to investigate. What is the response? What is the variable that you think might affect the response?
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

(Answer: Encourage students to focus on the cause-and-effect nature of the research question, such as “Does drinking coffee when studying for an exam tend to improve exam scores?” or “Does regular exercise reduce the chance of catching a cold during the semester?” or “Are students more likely to pass an exam written on blue paper than on yellow paper?” You may remind students to focus on words like average and tendency so that it is clear that the research question is not just about how one person might respond.)

Task 2: Types of Statistical Studies [Student Handout, estimated time: 20–25 minutes]

(Note: After going over the initial definitions, the next few questions allow students to practice making the distinction between observational studies and experiments and to think about potential issues in conducting such studies. Encourage students to work in pairs to critique the study designs before sharing answers as a class. You will probably want to bring the class together after Question 12 as well as at the end of this task.)

There are two common types of statistical studies—observational studies and experiments. Observational studies are typically used to learn about populations, and experiments are typically used to learn about the effect of some variable on a response.

A study is an observational study if it involves observing characteristics of a group (such as the amount of money spent on textbooks), usually for a sample selected from one or more existing populations (such as the population of all full-time students at your college). Because the goal of an observational study is typically to learn about the population, it is important that the sample be representative of the population.

A study is an experiment if it involves observing how a response (such as reading time) behaves under different experimental conditions (such as font sizes) that are actively manipulated by the researcher (e.g., giving some students one font size and other students a different font size). Because the goal of an experiment is to learn about the effect of the different experimental conditions, it is important to have comparable groups for each of the different experimental conditions.

Suppose you are interested in learning about the proportion of students at your college who favor a $20-per-semester fee increase to keep the college library open in the evenings and on weekends. You plan to select a sample of 100 students and ask each student whether he or she favors the fee increase.

(8) Does this involve learning about a population or learning about the effect of some variable on a response?

(Answer: This involves learning about a population—100 students are interviewed and from their responses a conclusion is drawn about the preference for the population of all students. You also have only one variable here—the opinion on the fee increase. You are not investigating whether another variable [like how the question is worded] might affect their responses.)

(9) Is this an observational study or an experiment? If it is an observational study, what is the population of interest? If it is an experiment, what is the response and what is the variable that you think might affect the response?

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

(Answer: This is an observational study because no variables are being manipulated by the researcher. Only existing opinions are measured.)

(10) Suppose you collect data by asking 100 students entering the library whether they favor the fee. Is this a good way to collect data? Why or why not?

(Answer: Students entering the library are library users and probably more interested in keeping the library open, in general, than students not entering the library.)

(11) Suppose you collect data by asking 100 students enrolled in freshman English whether they favor the fee. Is this a good way to collect data? Why or why not?

(Answer: Freshman students, especially those in need of numerous resources and/or long periods of quiet time, may not be representative of all students when it comes to willingness to pay an additional fee to keep the library open longer hours.)

(12) If the goal is to obtain a sample of 100 students that is representative of students at the college, suggest a way to select the 100 students that you think is better than the two methods described in Questions 10 and 11. Why is your proposed way better?

(Answer: The main goal is to not have the selection mechanism in any way related to library use. Instead of trying different locations on campus, it would be better to have a list of all enrolled students at the university and then select a subset from that list. Their own behavior does not influence whether they are selected.)

Instructor-Led Wrap-Up

Students may not suggest the list of all students, but encourage them to eliminate any self-selection in their sampling or anything related to library use that could increase or decrease the chance a student ends up in the sample. Remind students that good ways to select samples come in Topic 1.2, including how to carry out random selection.

Suppose you are interested in learning whether jogging for 30 minutes three times a week for six weeks decreases resting heart rate more than jogging for only 15 minutes three times a week for six weeks. You plan to use 100 college students who do not currently jog and have volunteered to participate as subjects in this study. The resting heart rate of each subject is measured at the start of the study. Fifty students participate in a jogging program where they get together three times a week and jog for 30 minutes. The other 50 students get together three times a week, but only jog for 15 minutes. At the end of six weeks, resting heart rate is measured again.

(13) Does this involve learning about a population or learning about the effect of some variable on a response?

(Answer: This involves learning about the effect of some variable [amount of time jogging] on a response [resting heart rate].)

(14) Is this an observational study or an experiment? If it is an observational study, what is the population of interest? If it is an experiment, what is the response and what is the variable that you think might affect the response?
**Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions**

(Answer: This is an experiment because the experimental group [30 minutes or 15 minutes of jogging] for each participant was assigned by the research [rather than letting the students decide themselves how long they jog].)

(15) Suppose you create the two groups for this study (the 30- and 15-minute jogging groups) by putting the 50 youngest volunteers into the 30-minute group and the oldest 50 into the 15-minute group. Is this a good idea? Why or why not?

(Answer: This is not recommended because if the 30-minute group has a larger change in resting heart rate, you are not able to decide whether it was due to jogging time or age. Therefore, although you can see whether there was a difference between the two groups, you cannot conclude whether jogging time was affecting the heart rate. The following diagram might help students see that the groups compared differ on both age and jogging time:

![Diagram](image)

(16) Suppose you create the two groups for this study (the 30- and 15-minute jogging groups) by putting the 50 volunteers weighing the most into the 30-minute group and the 50 weighing the least into the 15-minute group. Is this a good idea? Why or why not?

(Answer: Again, this creates two things [weight and jogging time] that are different between the jogging groups, and you cannot separate the effects of these two things. If the 30-minute group has a higher or lower change in resting heart rate, it could be due to the heavier weight rather than the jogging. You are not able to conclude that jogging time was affecting resting heart rate.)

(17) If the goal is to divide the 100 volunteers into two groups that result in a fair comparison of the two different jogging times, suggest a way to divide the 100 volunteers into two groups that you think is better than the two methods described in Questions 15 and 16. Why is your proposed way better?

(Answer: Flip a coin to decide which group someone is in [this is reasonable but actually not entirely preferred as the group sizes will probably be unequal or there will be some individuals at the end of the assignment that may be grouped together] or put the volunteers’ names on slips of paper, mix them up in a hat, and draw 50 to be in the 30-minute group. As students will see in Topic 1.3, this random assignment allows you to attribute any difference in heart rate change between the two groups to the jogging time and not some other factor.)
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Instructor-Led Wrap-Up

Emphasize that you want an impartial (random) mechanism for dividing into groups (like a coin toss), rather than self-selection. The goal is for there to be no preexisting differences between the experimental groups. Remind them that good ways to set up experiments come in Topic 1.3, including how to carry out random assignment.

Task 3: Drawing Conclusions from Statistical Studies [Student Handout, estimated time: 10 minutes]

(Note: This task involves a class discussion to wrap up this 50-minute session. Students are then given an out-of-class assignment that asks them to bring in an example for the next class. These student-provided examples will be used in Task 5.)

There are two types of conclusions that might be made from a statistical study that involve saying something beyond the data that are observed. One type of conclusion involves generalizing from what you see in a sample to some larger population. The other type of conclusion involves reaching cause-and-effect conclusions about the effect of some variable on a response. When is it reasonable to draw such conclusions? The answer depends on the type of study (observational study or experiment) and also on the design of the study.

(Note: Conduct a discussion about the table below, which summarizes when each of these types of conclusions is reasonable.

<table>
<thead>
<tr>
<th>Type of Conclusion</th>
<th>Reasonable When</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalize from sample to population</td>
<td>Random selection is used to obtain the sample</td>
</tr>
<tr>
<td>Change in response is caused by experimental conditions (cause-and-effect conclusion)</td>
<td>Random assignment to experimental conditions</td>
</tr>
</tbody>
</table>

Tie this into the discussion from Task 2. The study asking students’ opinion about library hours belongs in the first row and the study looking at jogging effects on resting heart rate belongs in the second row. Remind students that they will learn more about random selection and random assignment in later lessons, but that there is an important distinction between them. In particular, emphasize the distinction in their goals—obtaining a representative sample of a larger population versus creating comparable experimental groups—and the consequences in terms of the type of conclusions you can draw.

This implies the following:

- For observational studies, you should avoid making cause-and-effect conclusions, but it is possible to generalize from the sample to the population of interest if the study design incorporated random selection from the population.
- For experiments, it is possible to reach cause-and-effect conclusions if the study design uses random assignment to create the experimental groups.
- If an experiment uses both random assignment to create experimental groups and random selection from some population, it is possible to reach cause-and-effect conclusions and generalize these conclusions to the population.
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Also notice that a study that does not involve either random selection or random assignment is of limited value. Keep this and the three observations above in mind when you are evaluating statistical studies and deciding whether the conclusions drawn from them are reasonable. This is explored further in Tasks 4 and 5 later in this lesson.

(18) Out-of-class assignment in preparation for Tasks 4 and 5: Find a description of a statistical study that you think illustrates the steps in a statistical investigation (from Lesson 1.1.1) and bring a copy to the next class meeting. Possible sources of examples include your local newspaper, popular magazines, or online sources such as the following:

- LiveScience (www.livescience.com)
- USA Today (www.usatoday.com)
- the research section of Nature magazine (www.nature.com/nature/current_issue.html)
- ScienceDaily (www.sciencedaily.com)

Recall the steps in a statistical investigation:

**Steps in a Statistical Investigation**

1. Ask a question that can be answered by collecting data.
2. Decide what to measure and then collect data.
3. Summarize and analyze the data.
4. Draw a conclusion and communicate the results.

*(Note: Encourage students to look for an example that is of interest to them. As an alternative, find a small number of examples that you can provide students with and ask them to choose one to consider further.)*

**Task 4: Considering Study Design** [Student Handout, estimated time: 20 minutes]

*(Note: The next two tasks will probably make a 50-minute period. For Task 4, students can be given time to work on these questions individually or in pairs before they are discussed as a class.)*

The SAT exam (administered by an organization called The College Board) is used in admissions decisions by many four-year colleges and universities. Prior to 2005, the exam consisted of only multiple-choice questions, but in 2005 a writing section was added to the exam. Students taking the exam are given 25 minutes to write an essay on a particular topic. In 2006, The College Board conducted a study of 6,498 essays selected at random from the more than 1.4 million SAT exams taken in the 2005–2006 academic year. For this sample of essays, 15% were written in cursive and 85% were printed in block letters. It was found that the average score for essays written in cursive was higher than the average score for essays that were printed.

(19) Is this study an observational study or an experiment?

*(Answer: This is an observational study because variables were recorded on individuals but not manipulated by the researchers. In particular, the researchers did not control which type of writing was used by students; the students decided that themselves.)*
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

(20) Is it reasonable to conclude that writing the essay in cursive was the cause of the higher scores? Why or why not?

(Answer: Not really. There could be other differences between students who use cursive writing and students who use block letters. For example, those who use cursive writing may be more likely to attend private schools and test preparation courses than those who use print writing. You do not know for sure this is the case, but the potential for something else to be different between the two groups exists. Therefore, you should not be comfortable drawing any cause-and-effect conclusions from this observational study.)

An article published in the Archives of Pediatrics and Adolescent Medicine (“Display of Health Risk Behaviors on MySpace by Adolescents,” 2009) described a study in which researchers looked at a random sample of 500 publicly accessible MySpace web profiles posted by 18-year-olds. The content of each profile was analyzed. One conclusion reported was that displaying sport or hobby involvement was associated with decreased references to risky behavior (sexual references or references to substance abuse or violence).

(21) Is it reasonable to generalize the stated conclusion to all 18-year-olds with a publicly accessible MySpace web profile? What aspect of the study supports your answer?

(Answer: Yes, because those in the study were randomly selected from the population of interest [18-year-olds with a publicly accessible MySpace web profile], you can generalize information from this sample to that larger population.)

(22) Not all MySpace users have a publicly accessible profile. Is it reasonable to generalize the stated conclusion to all 18-year-old MySpace users? Why or why not?

(Answer: No, there could be something distinct about those with public profiles, especially with respect to information about high-risk behavior.)

The April 20, 2009, issue of the magazine Sports Illustrated reported that the Oklahoma City Thunder, a professional basketball team, had a win-loss record for the 2008–2009 season that was actually worse for home games that were sold out (3 wins and 15 losses) than for home games that were not sold out (12 wins and 11 losses).

(23) Based on these data, is it reasonable to conclude that a sellout crowd is the cause of the team’s poor performance at home games? Can you think of another explanation for why the win-loss record might be worse for sold-out games than for games that are not sold out?

(Answer: No, this was an observational study, not an experiment—the researchers did not impose which games would be sellouts, but observed them as they occurred. Reinforce to students that citing the observational nature of the study is sufficient, or encourage them to suggest an alternative explanation for the observed relationship. For example, sold-out games may have been the home games that were against really good visiting teams.)

(24) Did random selection or random assignment play any role in the collection of these data?

(Answer: No, this was all games in one season.)
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Task 5: Students’ Study Examples [Student Handout, estimated time: 30 minutes]
(Note: Form groups of three to five students to complete the steps in this task. Allow about 15 minutes for the group work and about 15 minutes for the groups to present their work.

Have each student in the group take about one minute to briefly describe the statistical study that he or she brought to class. The group then selects one of the studies to focus on for the rest of this task.

Alternatively, if you provide students with a small number of studies from which to choose, form the groups based on the study students were interested in considering.)

For your group’s study, work together to answer the following questions. Choose one group member to present the study and answers to the questions to the class.

**Question A:** What is one question the study was trying to answer?

**Question B:** Is the study an observational study or an experiment? If it is an observational study, what is the population of interest? If it is an experiment, what is the response and what is the variable that you think might affect the response?

**Question C:** Did the study involve random selection of a sample or random assignment to experimental groups?

**Question D:** Is it reasonable to generalize the results of this study to some larger population?

**Question E:** Is it reasonable to draw a cause-and-effect conclusion based on this study?

**Question F:** If the news article attempted to summarize the results of a study published elsewhere, did the publication include all the relevant information to help you answer these questions? Is the headline appropriate?

Lesson Wrap-Up

This instructor-led discussion emphasizes the following points:

- **The distinction between observational studies and experiments:** The main focuses are that the latter involves actively imposing the condition you want to see the effect of and that there is a difference in the types of conclusion that can be drawn.

- **The distinction between random selection and random assignment:** They have different purposes (representative sample, comparable treatment groups) and consequences (generalizability to larger population, potential to draw cause-and-effect conclusions). The scope of conclusions you can legitimately draw from a study (e.g., generalize to larger population, cause and effect) are dependent on how the study is conducted. Poorly conducted studies (e.g., those based on self-selected samples) severely limit what you can conclude beyond the individuals in the study.
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Homework

“Sweet Potatoes Brighten Your Skin” is the headline of an article that appeared in the magazine Woman’s World (November 1, 2010). The article concludes that eating sweet potatoes causes skin to be healthier because it reverses age spots, blocks harmful ultraviolet rays in sunlight, and protects against skin dryness. Consider five hypothetical study designs. For each study, answer the following five questions:

Question A: Is the study described an observational study or an experiment?
Question B: Did the study use random selection from some population?
Question C: Did the study use random assignment to experimental groups?
Question D: Is the conclusion “eating sweet potatoes leads to healthier skin” appropriate given the study description? Explain your reasoning.
Question E: Is it reasonable to generalize conclusions from this study to some larger population? If so, what population?

Study Design 1

Two hundred students were selected at random from those enrolled at a large college in California. Each student in the sample was asked whether he or she ate sweet potatoes more than once in a typical week. A skin specialist rated skin health for each student on a scale of 1 to 10. It was concluded that skin health was significantly better on average for the group that reported eating sweet potatoes more than once a week.

(Answers: A. Observational study. B. Random selection from the population of all students at the large college. C. No, students were asked whether they ate sweet potatoes. D. No, this was an observational study. E. Yes, because the study involved random selection, you can generalize this conclusion to students enrolled at this larger college in California.)

Study Design 2

One hundred people living in Miami volunteered to participate in a statistical study. The volunteers were divided into two experimental groups based on gender, with women in Group 1 and men in Group 2. Those in Group 1 ate 6 ounces of sweet potatoes daily for three months. Those in Group 2 did not eat sweet potatoes for one month. At the end of the three months, a skin specialist rated skin health on a scale of 1 to 10 for each volunteer. It was concluded that skin health was significantly better on average for Group 1.

(Answers: A. Experiment because the researchers told participants what to eat. B. No, volunteers. C. No, participants were split into groups by gender instead of randomly assigned. D. No, even though this was an experiment, the participants were not randomly assigned to the two groups. You cannot separate the effects of gender and diet. E. Not really because the study used volunteers. You might be willing to generalize to people who live in Miami, but even this is risky.)
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Study Design 3
Two hundred people volunteered to participate in a statistical study. Each person was asked how often he or she ate sweet potatoes, and a skin specialist rated skin health on a scale of 1 to 10 for each volunteer. It was concluded that skin health was significantly better for those who ate sweet potatoes more than once a week.

(Answers: A. Observational study. B. No, volunteers. C. No, participants were asked how often they ate sweet potatoes. D. No, this was an observational study. E. This study involved volunteers, and you have no additional information about the participants. Therefore, you cannot comfortably generalize these results to any larger population.)

Study Design 4
One hundred people volunteered to participate in a statistical study. For each volunteer, a coin was tossed and if the coin landed heads up, the volunteer was assigned to Group 1. If the coin landed tails up, the volunteer was assigned to Group 2. Those in Group 1 ate 6 ounces of sweet potatoes daily for three months. Those in Group 2 did not eat sweet potatoes for one month. At the end of the three months, a skin specialist rated skin health on a scale of 1 to 10 for each volunteer. It was concluded that skin health was significantly better on average for those in Group 1.

(Answers: A. Experiment because the researchers assigned the participants to groups and determined how much sweet potatoes would be consumed in each group. B. No, volunteers. C. Yes, the coin-tossing procedure randomly assigned the participants to experimental groups. D. Yes, you can draw a cause-and-effect conclusion because this was a randomized experiment [and the results were statistically significant]. E. This study involved volunteers, and you have no additional information about the participants. Therefore, you cannot comfortably generalize these results to any larger population.)

Study Design 5
One hundred students were selected at random from those enrolled at a large college. Each selected student was asked to participate in a study, and all agreed to participate. For each student, a coin was tossed and if the coin landed heads up, the student was assigned to Group 1. If the coin landed tails up, the student was assigned to Group 2. Those in Group 1 ate 6 ounces of sweet potatoes daily for three months. Those in Group 2 did not eat sweet potatoes for one month. At the end of the three months, a skin specialist rated skin health on a scale of 1 to 10 for each volunteer. It was concluded that skin health was significantly better for those in Group 1.

(Answers: A. Experiment. B. Yes, they were randomly selected from students at the college. C. Yes, the coin-tossing procedure randomly assigned the participants to experimental groups. D. Yes, you can draw a cause-and-effect conclusion because this was a randomized experiment [and the results were statistically significant]. E. Yes, because of the random selection, you can generalize these results to all students at this university.)
Task 1: Research Questions that Can Be Answered by Collecting Data

In Lesson 1.1.1, a four-step process that is common to many statistical studies was introduced. Step 1 is, “Ask a question that can be answered by collecting data.” The questions that are the focus of statistical studies are usually one of the following types: research questions about populations or research questions about the effect of some variable on a response.

Research Questions About Populations

A population is the entire group of individuals or objects that you are interested in learning about. Usually, it is not possible to study the entire population, so you collect data on just a part of the population. For example, if you were interested in learning about the proportion of households in Boston that routinely recycle plastic containers, it would be both time consuming and costly to collect data from every household. You might only be able to collect data for 1,000 households in Boston, but if the 1,000 households were selected carefully, you can still learn a lot about plastic recycling in Boston. The part of the population for which data are collected is called a sample.

Research questions about populations might involve

• estimating a numerical population characteristic or testing a claim about a population (e.g., “Do more than half of all households in Boston recycle plastic containers?”).
• comparing two or more populations (e.g., “Is plastic recycling more common in Los Angeles than in Boston?”).
• asking whether two variables are related in the population of interest (e.g., “Is there a relationship between household income and which households in Boston recycle plastics?”)

Research Questions About the Effect of Some Variable on a Response

This type of question can take the form, “What is the effect of ...?” or “What happens when...?” For example, “What is the effect of room temperature on exam performance?” or “What happens to the distance a car can travel on 1 gallon of gasoline if 100 pounds of additional weight are put in the trunk of the car?”

Additional Examples of Research Questions About Populations

(1) Give an example of a research question that involves estimating a population characteristic where the population of interest is all full-time students at your college.

(2) Give an example of a research question that involves estimating a population characteristic for a different population. What is the population of interest for your question?
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

(3) Give an example of a research question that involves testing a claim about a population characteristic where the population of interest is all students at your college.

(4) Give an example of a research question that involves testing a claim comparing two populations. What are the populations of interest for your question?

(5) Give an example of a research question that involves investigating the relationship between two variables where the population of interest is all students at your college.

(6) Give an example of a research question that involves investigating the relationship between two variables for a different population. What is the population of interest for your question?

Additional Examples of Research Questions About the Effect of Some Variable on a Response

(7) Give an example of a research question about the effect of some variable on a response that you think would be interesting to investigate. What is the response? What is the variable that you think might affect the response?
Task 2: Types of Statistical Studies

There are two common types of statistical studies—observational studies and experiments.

Observational studies are typically used to learn about populations, and experiments are typically used to learn about the effect of some variable on a response.

A study is an observational study if it involves observing characteristics of a group (such as the amount of money spent on textbooks), usually for a sample selected from one or more existing populations (such as the population of all full-time students at your college). Because the goal of an observational study is typically to learn about the population, it is important that the sample be representative of the population.

A study is an experiment if it involves observing how a response (such as reading time) behaves under different experimental conditions (such as font sizes) that are actively manipulated by the researcher (e.g., giving some students one font size and other students a different font size). Because the goal of an experiment is to learn about the effect of the different experimental conditions, it is important to have comparable groups for each of the different experimental conditions.

Suppose you are interested in learning about the proportion of students at your college who favor a $20-per-semester fee increase to keep the college library open in the evenings and on weekends. You plan to select a sample of 100 students and ask each student whether he or she favors the fee increase.

(8) Does this involve learning about a population or learning about the effect of some variable on a response?

(9) Is this an observational study or an experiment? If it is an observational study, what is the population of interest? If it is an experiment, what is the response and what is the variable that you think might affect the response?

(10) Suppose you collect data by asking 100 students entering the library whether they favor the fee. Is this a good way to collect data? Why or why not?
Suppose you collect data by asking 100 students enrolled in freshman English whether they favor the fee. Is this a good way to collect data? Why or why not?

If the goal is to obtain a sample of 100 students that is representative of students at the college, suggest a way to select the 100 students that you think is better than the two methods described in Questions 10 and 11. Why is your proposed way better?

Suppose you are interested in learning whether jogging for 30 minutes three times a week for six weeks decreases resting heart rate more than jogging for only 15 minutes three times a week for six weeks. You plan to use 100 college students who do not currently jog and have volunteered to participate as subjects in this study. The resting heart rate of each subject is measured at the start of the study. Fifty students participate in a jogging program where they get together three times a week and jog for 30 minutes. The other 50 students get together three times a week, but only jog for 15 minutes. At the end of six weeks, resting heart rate is measured again.

Does this involve learning about a population or learning about the effect of some variable on a response?

Is this an observational study or an experiment? If it is an observational study, what is the population of interest? If it is an experiment, what is the response and what is the variable that you think might affect the response?
Suppose you create the two groups for this study (the 30- and 15-minute jogging groups) by putting the 50 youngest volunteers into the 30-minute group and the oldest 50 into the 15-minute group. Is this a good idea? Why or why not?

Suppose you create the two groups for this study (the 30- and 15-minute jogging groups) by putting the 50 volunteers weighing the most into the 30-minute group and the 50 weighing the least into the 15-minute group. Is this a good idea? Why or why not?

If the goal is to divide the 100 volunteers into two groups that result in a fair comparison of the two different jogging times, suggest a way to divide the 100 volunteers into two groups that you think is better than the two methods described in Questions 15 and 16. Why is your proposed way better?
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Task 3: Drawing Conclusions from Statistical Studies

There are two types of conclusions that might be made from a statistical study that involve saying something beyond the data that are observed. One type of conclusion involves generalizing from what you see in a sample to some larger population. The other type of conclusion involves reaching cause-and-effect conclusions about the effect of some variable on a response. When is it reasonable to draw such conclusions? The answer depends on the type of study (observational study or experiment) and also on the design of the study.

This implies the following:

- For observational studies, you should avoid making cause-and-effect conclusions, but it is possible to generalize from the sample to the population of interest if the study design incorporated random selection from the population.
- For experiments, it is possible to reach cause-and-effect conclusions if the study design uses random assignment to create the experimental groups.
- If an experiment uses both random assignment to create experimental groups and random selection from some population, it is possible to reach cause-and-effect conclusions and generalize these conclusions to the population.

Also notice that a study that does not involve either random selection or random assignment is of limited value. Keep this and the three observations above in mind when you are evaluating statistical studies and deciding whether the conclusions drawn from them are reasonable. This is explored further in Tasks 4 and 5 later in this lesson.

(18) Out-of-class assignment in preparation for Tasks 4 and 5: Find a description of a statistical study that you think illustrates the steps in a statistical investigation (from Lesson 1.1.1) and bring a copy to the next class meeting. Possible sources of examples include your local newspaper, popular magazines, or online sources such as the following:

- LiveScience (www.livescience.com)
- USA Today (www.usatoday.com)
- the research section of Nature magazine (www.nature.com/nature/current_issue.html)
- ScienceDaily (www.sciencedaily.com)

Recall the steps in a statistical investigation:

<table>
<thead>
<tr>
<th>Steps in a Statistical Investigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ask a question that can be answered by collecting data.</td>
</tr>
<tr>
<td>2. Decide what to measure and then collect data.</td>
</tr>
<tr>
<td>3. Summarize and analyze the data.</td>
</tr>
<tr>
<td>4. Draw a conclusion and communicate the results.</td>
</tr>
</tbody>
</table>

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Task 4: Considering Study Design

The SAT exam (administered by an organization called The College Board) is used in admissions decisions by many four-year colleges and universities. Prior to 2005, the exam consisted of only multiple-choice questions, but in 2005 a writing section was added to the exam. Students taking the exam are given 25 minutes to write an essay on a particular topic. In 2006, The College Board conducted a study of 6,498 essays selected at random from the more than 1.4 million SAT exams taken in the 2005–2006 academic year. For this sample of essays, 15% were written in cursive and 85% were printed in block letters. It was found that the average score for essays written in cursive was higher than the average score for essays that were printed.

(19) Is this study an observational study or an experiment?

(20) Is it reasonable to conclude that writing the essay in cursive was the cause of the higher scores? Why or why not?

An article published in the Archives of Pediatrics and Adolescent Medicine (“Display of Health Risk Behaviors on MySpace by Adolescents,” 2009) described a study in which researchers looked at a random sample of 500 publicly accessible MySpace web profiles posted by 18-year-olds. The content of each profile was analyzed. One conclusion reported was that displaying sport or hobby involvement was associated with decreased references to risky behavior (sexual references or references to substance abuse or violence).

(21) Is it reasonable to generalize the stated conclusion to all 18-year-olds with a publicly accessible MySpace web profile? What aspect of the study supports your answer?

(22) Not all MySpace users have a publicly accessible profile. Is it reasonable to generalize the stated conclusion to all 18-year-old MySpace users? Why or why not?
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

The April 20, 2009, issue of the magazine *Sports Illustrated* reported that the Oklahoma City Thunder, a professional basketball team, had a win-loss record for the 2008–2009 season that was actually worse for home games that were sold out (3 wins and 15 losses) than for home games that were not sold out (12 wins and 11 losses).

(23) Based on these data, is it reasonable to conclude that a sellout crowd is the cause of the team’s poor performance at home games? Can you think of another explanation for why the win-loss record might be worse for sold-out games than for games that are not sold out?

(24) Did random selection or random assignment play any role in the collection of these data?
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Task 5: Study Examples

For your group’s study, work together to answer the following questions. Choose one group member to present the study and answers to the questions to the class.

**Question A:** What is one question the study was trying to answer?

**Question B:** Is the study an observational study or an experiment? If it is an observational study, what is the population of interest? If it is an experiment, what is the response and what is the variable that you think might affect the response?

**Question C:** Did the study involve random selection of a sample or random assignment to experimental groups?

**Question D:** Is it reasonable to generalize the results of this study to some larger population?

**Question E:** Is it reasonable to draw a cause-and-effect conclusion based on this study?

**Question F:** If the news article attempted to summarize the results of a study published elsewhere, did the publication include all the relevant information to help you answer these questions? Is the headline appropriate?
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Homework

“Sweet Potatoes Brighten Your Skin” is the headline of an article that appeared in the magazine *Woman’s World* (November 1, 2010). The article concludes that eating sweet potatoes causes skin to be healthier because it reverses age spots, blocks harmful ultraviolet rays in sunlight, and protects against skin dryness. Consider five hypothetical study designs. For each study, answer the following five questions:

**Question A:** Is the study described an observational study or an experiment?

**Question B:** Did the study use random selection from some population?

**Question C:** Did the study use random assignment to experimental groups?

**Question D:** Is the conclusion “eating sweet potatoes leads to healthier skin” appropriate given the study description? Explain your reasoning.

**Question E:** Is it reasonable to generalize conclusions from this study to some larger population? If so, what population?

**Study Design 1**

Two hundred students were selected at random from those enrolled at a large college in California. Each student in the sample was asked whether he or she ate sweet potatoes more than once in a typical week. A skin specialist rated skin health for each student on a scale of 1 to 10. It was concluded that skin health was significantly better on average for the group that reported eating sweet potatoes more than once a week.

**Study Design 2**

One hundred people living in Miami volunteered to participate in a statistical study. The volunteers were divided into two experimental groups based on gender, with women in Group 1 and men in Group 2. Those in Group 1 ate 6 ounces of sweet potatoes daily for three months. Those in Group 2 did not eat sweet potatoes for one month. At the end of the three months, a skin specialist rated skin health on a scale of 1 to 10 for each volunteer. It was concluded that skin health was significantly better on average for Group 1.
Supporting Lesson 1.1.2: Types of Statistical Studies and Scope of Conclusions

Study Design 3
Two hundred people volunteered to participate in a statistical study. Each person was asked how often he or she ate sweet potatoes, and a skin specialist rated skin health on a scale of 1 to 10 for each volunteer. It was concluded that skin health was significantly better for those who ate sweet potatoes more than once a week.

Study Design 4
One hundred people volunteered to participate in a statistical study. For each volunteer, a coin was tossed and if the coin landed heads up, the volunteer was assigned to Group 1. If the coin landed tails up, the volunteer was assigned to Group 2. Those in Group 1 ate 6 ounces of sweet potatoes daily for three months. Those in Group 2 did not eat sweet potatoes for one month. At the end of the three months, a skin specialist rated skin health on a scale of 1 to 10 for each volunteer. It was concluded that skin health was significantly better on average for those in Group 1.

Study Design 5
One hundred students were selected at random from those enrolled at a large college. Each selected student was asked to participate in a study, and all agreed to participate. For each student, a coin was tossed and if the coin landed heads up, the student was assigned to Group 1. If the coin landed tails up, the student was assigned to Group 2. Those in Group 1 ate 6 ounces of sweet potatoes daily for three months. Those in Group 2 did not eat sweet potatoes for one month. At the end of the three months, a skin specialist rated skin health on a scale of 1 to 10 for each volunteer. It was concluded that skin health was significantly better for those in Group 1.
Initiating Lesson 1.2.1: Collecting Data by Sampling

Estimated 50-minute class sessions: 1

Materials Required
Each student needs a copy of the following:
- the unnumbered fish picture,
- the paper ruler, and
- either the fish picture with fish lengths added or the fish length key.

These images are provided in a separate file.
You also need flip chart paper or overhead transparencies for the dotplots constructed in the rich task. (These are revisited in Lesson 1.2.2, so you need to keep them.)

Learning Goals
Students will begin to understand
- that the absence of bias is one characteristic of a good sampling plan.
- possible sources of bias that can occur when sampling.

Students will begin to be able to
- distinguish between a sample and a population.
- explain the difference between a census (a study of an entire population) and a sample (a study of a subset of a population).
- summarize numerical data graphically using a dotplot.
- explain why a poor sampling plan can lead to misleading conclusions.

Introduction [Student Handout]
Often organizations are interested in learning about some population, but it is not feasible to collect data on every individual in the population. For example, suppose you want to learn about the average number of units that students at your college are enrolled in this semester. The ideal way to proceed is to contact every student at your college and ask how many units he or she is enrolled in. Then you could compute the average and know the actual value.

- Is this practical for you to do? Why or why not?
  (Answer: This would be a lot of work, and it would take a long time to complete this task.)
- Suggest a different strategy that allows you to learn about the average number of units for students in this population.
  (Answer: If it is impractical to study the entire population, examining a sample selected from the population might be possible. If that sample has similar characteristics, you can use the information from the sample to make inferences about the population as a whole.)

You might be willing to settle for an estimate of the average, rather than needing to know the exact value. If you will accept a bit of inaccuracy, a reasonable approach is collecting data for only a sample of students and then using the data from the sample to calculate an estimate of the average.
Part I: Something Fishy Going On [Student Handout, estimated time: 50 minutes]

Suppose you have set up a large fish tank for 91 fish. With this many fish, you are worried that an even larger tank might be needed. You consult the First Tank Guide (www.firsttankguide.net/capacity.php) and see that it recommends about 1 gallon of water for each 2.5 centimeters of fish length. If you have an estimate of the average length of the fish in the tank and you know that there are a total of 91 fish, you could decide how many gallons of water is the least amount the tank should hold. For example, if your estimate of the average length of the fish is 5 centimeters, the total length of fish in the tank is approximately $91 \times 5 = 455$ centimeters.

(Note: If students have trouble seeing why having an estimate of the average length is helpful, show them that if 1 gallon of water is needed for each 2.5 centimeters of fish length, the tank should hold at least $455/2.5 = 182$ gallons of water. So, to help decide whether a larger tank is needed, you want to learn about the average length of the fish in the tank.)

Let’s start by thinking about the question you want to answer and decide what should be measured to answer this question. The question is, “What is the average length of the fish in the tank?”

To answer this question, you need to know how to measure the length of a fish. For this task, the length of a fish is defined as the length from the tip of the nose to the midpoint of the tail. For example, the length of the two fish below is shown by the line drawn from nose to tail.

Your instructor will now provide you with a picture of the fish in the tank, but keep this picture face down and do not look at it until your instructor tells you to do so.

(Note: Distribute the unnumbered fish picture face down. After all students have the picture, give them 5 seconds to look at it and then ask them to turn the paper over again so that the fish are out of sight.)
Initiating Lesson 1.2.1: Collecting Data by Sampling

Giving students about 5 seconds to look at the fish typically leads to them overestimating the average fish length as they tend to overlook the many small fish.)

(1) Based on your look at the fish population, draw a line in the space below that represents your best guess of the average length of the fish in the tank.

(2) Use the ruler provided by your instructor to measure the length of the line that represents your best guess of the average fish length. Record your value to the nearest tenth of a centimeter.

(Note: Use the paper rulers provided or rulers that allow students to measure the length in centimeters so that the measurement is comparable to those provided later in this lesson and in the next lesson. You may need to project the ruler and show that each small hash mark between the numbered hash marks represents one-tenth of a centimeter. Students should express the measured fish length as a decimal, such as 4.3 centimeters.

While students are doing this, talk briefly about the need for accurate measurements and consistent measurement methods. If the nose-to-tail method had not been outlined, students could use different methods, leading to inconsistency in those measurements.)

(3) Look at the picture again and pick five fish that you think are representative of the fish in the tank. Circle these five fish (your sample).

(4) Using the key provided by your instructor, record the lengths of the five fish you selected. Then compute the average of these five fish lengths.

(Note: You can provide the second picture of the fish where the lengths are indicated or the key that gives the length of each type of fish. Another option is having students use their ruler to measure the five fish they selected. If a student selects a fish that is partially hidden [making it impossible to measure length], he or she should measure another fish of the same type.)

(5) Starting with your best guess of the average fish length from Question 2, add your guess to the dotplot that displays the guesses for all students in the class.

(Note: Construct the class dotplot on a piece of flip chart paper or on an overhead transparency rather than on the board—you will revisit this dotplot in the Lesson 1.2.2. You can also have students submit their numerical values for you to enter into a spreadsheet so that you can recreate the graphs using technology.

Draw a scale labeled from 0 to 10 with tic marks at every tenth if possible. Add a title of “Best Guess.” Then have each student add his or her guess to the dotplot. You can now talk about the distribution of guesses. Was there general consensus around a particular value? Was there a lot of variability from student to student? Combining Questions 5 and 6 and constructing both dotplots at the same time may be a quicker approach.)
(6) Look at the average fish length you computed in Question 4 for your sample of five fish. Add your average to a dotplot that displays the averages for all students.

(Note: Again, use flip chart paper or an overhead transparency. Draw a scale that matches the one used for the best guess dotplot and add the title “Self-Selected Sample Average.” Have each student add his or her average to the dotplot.)

Part I Wrap-Up [Instructor-led discussion]

Begin the wrap-up by discussing the distribution of the sample averages.

- Was there general consensus around a particular value?
- Was there a lot of variability in the average from student to student?
- In what ways were the distributions of the averages and the best guesses similar? Were there any notable differences in the two distributions?

Now to the important point of this task and the lead-in to the rest of the lessons in this topic. Both methods for estimating average fish length (best guess and self-selected representative sample) will probably overestimate the actual average length for fish in the tank. This is because students often overlook the large number of small fish in the tank. At this point, you can reveal the actual average length of the fish in the tank, which is 3.1 centimeters (actually 3.103, but 3.1 is close enough for this task). Mark 3.1 on the two dotplots. If your class is typical, the two dotplots show distributions that tend to center above 3.1.

At this point, ask the class the following questions as part of the whole-group discussion.

- Looking back, do you think your representative sample was really representative of the population? If not, what went wrong?
- The judgments about average length or even about what a representative sample might look like are not very good! Why do you think the judgments were not very good?

Part II [Student Handout]

(Note: At this point, ask students to answer Questions 7 and 8. This could be done by having students work individually or in small groups.)

(7) What does it mean to say that a sample is a representative sample?

(Answer: You want the distribution of values [lengths] in the sample to be similar to the distribution of lengths in the population.)

(8) If you can only measure five fish, what might be a better way to select the fish than using personal judgment?

(Answer: Select the five fish at random [give every fish an equal chance of being selected]. The idea of random selection was introduced in Lesson 1.1.2, and random sampling is the topic of the next lesson.)
Initiating Lesson 1.2.1: Collecting Data by Sampling

Part II Wrap-Up [Instructor-led discussion]

In closing, emphasize that when you talk about a biased sampling method, you do not just mean the sample averages differ from the true value but that they differ from the true value in a systematic way (i.e., in the same direction). The idea of bias is introduced more formally in Lessons 1.2.2 and 1.2.4. Also keep in mind that you talk about the method being biased, not just one sample.

Homework [Student Handout]

(1) Recall that in Lesson 1.1.2, the terms population and sample were defined as follows:

A population is the entire group of individuals or objects that you are interested in learning about. Usually, it is not possible to study the entire population, so you collect data on just a part of the population. The part of the population for which data are collected is called a sample.

A census is a study that involves collecting data on every individual in the population.

(a) For the fish length example in this lesson, what is the population of interest?

(Answer: All 91 fish in the tank)

(b) What would taking a census of the population to determine average fish length involve?

(Answer: Measuring the length of each of the 91 fish)

(2) Another way of selecting five fish for a sample that could be used to estimate the average length of the fish in the tank is described below. Explain why this method of selecting a sample would produce sample averages that differ from the actual average fish length in a systematic way. Will this method tend to produce sample averages that are too big or sample averages that are too small? (Hint: Look at your picture and think about carrying out this process yourself.)

Proposed Selection Method: Hold a pencil about 18 inches above and centered over the picture of the fish. Drop the pencil onto the picture and then measure the fish that the pencil point lands closest to. Repeat this process four times to obtain a sample of five fish.

(Answer: This still tends to overestimate the length of fish because the larger fish have a higher probability of being landed on.)

(3) Fish are fed by sprinkling fish food on the surface of the water. Fish swim to the top to feed. Another way of selecting five fish for a sample is to measure the first five fish to swim to the top of the tank when fish food is added. Do you think this is a good way to select the five fish for the sample? Explain your reasoning.

(Note: Be flexible in the answers you accept. Students may think the smaller fish are faster and more likely to get to the top first. Students may think larger fish are hungrier and more aggressive and therefore more likely to get to the top first. As long as they suggest a connection between how the fish in the sample will tend to differ from the rest of the fish with respect to length, they have made a sound argument. Students also may not see a link between feeding and length. Encourage them to think about the desire to make sure every fish has an equal chance of being selected.)
Often organizations are interested in learning about some population, but it is not feasible to collect data on every individual in the population. For example, suppose you want to learn about the average number of units that students at your college are enrolled in this semester. The ideal way to proceed is to contact every student at your college and ask how many units he or she is enrolled in. Then you could compute the average and know the actual value.

- Is this practical for you to do? Why or why not?
- Suggest a different strategy that allows you to learn about the average number of units for students in this population.

You might be willing to settle for an estimate of the average, rather than needing to know the exact value. If you will accept a bit of inaccuracy, a reasonable approach is collecting data for only a sample of students and then using the data from the sample to calculate an estimate of the average.

### Part I: Something Fishy Going On

Suppose you have set up a large fish tank for 91 fish. With this many fish, you are worried that an even larger tank might be needed. You consult the First Tank Guide (www.firsttankguide.net/capacity.php) and see that it recommends about 1 gallon of water for each 2.5 centimeters of fish length. If you have an estimate of the average length of the fish in the tank and you know that there are a total of 91 fish, you could decide how many gallons of water is the least amount the tank should hold. For example, if your estimate of the average length of the fish is 5 centimeters, the total length of fish in the tank is approximately $91 \times 5 = 455$ centimeters.

Let’s start by thinking about the question you want to answer and decide what should be measured to answer this question. The question is, “What is the average length of the fish in the tank?”

To answer this question, you need to know how to measure the length of a fish. For this task, the length of a fish is defined as the length from the tip of the nose to the midpoint of the tail. For example, the length of the two fish below is shown by the line drawn from nose to tail.
Initiating Lesson 1.2.1: Collecting Data by Sampling

Your instructor will now provide you with a picture of the fish in the tank, but keep this picture face down and do not look at it until your instructor tells you to do so.

(1) Based on your look at the fish population, draw a line in the space below that represents your best guess of the average length of the fish in the tank.

(2) Use the ruler provided by your instructor to measure the length of the line that represents your best guess of the average fish length. Record your value to the nearest tenth of a centimeter.

(3) Look at the picture again and pick five fish that you think are representative of the fish in the tank. Circle these five fish (your sample).

(4) Using the key provided by your instructor, record the lengths of the five fish you selected. Then compute the average of these five fish lengths.

(5) Starting with your best guess of the average fish length from Question 2, add your guess to the dotplot that displays the guesses for all students in the class.

(6) Look at the average fish length you computed in Question 4 for your sample of five fish. Add your average to a dotplot that displays the averages for all students.

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanancenter.org/mathways.
Initiating Lesson 1.2.1: Collecting Data by Sampling

Part II

(7) What does it mean to say that a sample is a representative sample?

(8) If you can only measure five fish, what might be a better way to select the fish than using personal judgment?
Initiating Lesson 1.2.1: Collecting Data by Sampling

Homework

(1) Recall that in Lesson 1.1.2, the terms population and sample were defined as follows:

A population is the entire group of individuals or objects that you are interested in learning about. Usually, it is not possible to study the entire population, so you collect data on just a part of the population. The part of the population for which data are collected is called a sample.

A census is a study that involves collecting data on every individual in the population.

(a) For the fish length example in this lesson, what is the population of interest?

(b) What would taking a census of the population to determine average fish length involve?

(2) Another way of selecting five fish for a sample that could be used to estimate the average length of the fish in the tank is described below. Explain why this method of selecting a sample would produce sample averages that differ from the actual average fish length in a systematic way. Will this method tend to produce sample averages that are too big or sample averages that are too small? (Hint: Look at your picture and think about carrying out this process yourself.)

Proposed Selection Method: Hold a pencil about 18 inches above and centered over the picture of the fish. Drop the pencil onto the picture and then measure the fish that the pencil point lands closest to. Repeat this process four times to obtain a sample of five fish.
Initiating Lesson 1.2.1: Collecting Data by Sampling

(3) Fish are fed by sprinkling fish food on the surface of the water. Fish swim to the top to feed. Another way of selecting five fish for a sample is to measure the first five fish to swim to the top of the tank when fish food is added. Do you think this is a good way to select the five fish for the sample? Explain your reasoning.
Supporting Lesson 1.2.2: Random Sampling

Estimated 50-minute class sessions: 1

Materials Required
Each student needs a copy of the following:
- the numbered fish picture,
- the paper ruler, and
- either the fish picture with fish lengths added or the fish length key.

You need flip chart paper or overhead transparencies for the dotplot constructed in Task 1. You also need the two class dotplots constructed in Lesson 1.2.1.

Learning Goals
Students will begin to understand
- that the absence of bias is one characteristic of a good sampling plan.
- that bias can sometimes be introduced by the way the sample is selected.
- how random sampling protects against selection bias.
- the roles of sample size and population size in determining the accuracy of a sample.
- that a small sample can be considered representative of a large population if care is taken in the sample selection.

Students will begin to be able to
- given a finite population, describe a method for selecting a simple random sample.
- summarize a categorical data set using a bar chart.
- critically evaluate a sampling plan.
- describe the implications of a good sampling plan.

Introduction [Student Handout]

(Note: The introduction can be done as instructor-led discussion, or students can read it on their own before or in class. If students read the introduction before class, start the class period by asking some basic comprehension questions about the material to make sure they understood it. For example, you might ask the following:

- What do you remember about the fish activity?
- Why did you take a sample from the fish population (rather than measure the lengths of all fish in the population)?
- Why is it important for a sample to be representative of the population?
- What was the problem with letting you personally judge a representative sample?)

In Lesson 1.2.1, you saw that some methods of selecting a sample from a population can lead to misleading conclusions about a population. In that lesson, informal “first impression” estimates, where you guessed at the average fish length after looking at the fish population for a short time, tended to overestimate the actual average fish length. This may not be surprising as you only had a short time to look at the fish population. You also saw that when classmates attempted to pick a sample that was representative of the population and then use the fish in the sample to estimate the average length of fish in the population, the sample averages still consistently tended to be too large.
Supporting Lesson 1.2.2: Random Sampling

This happened because you overlooked the many small fish in the population—your attention was drawn to the large fish, causing them to be overrepresented in the sample. If some individuals in the population are more likely to be included in the sample than others, the sampling method results in selection bias. Even if you tried to pick one fish of each type, this ignores that there are more small fish than large fish in this tank. More generally, you have selection bias when some types of individuals in the population (like the small fish) are chosen in a manner that does not represent how often they occur in the population.

Instead of letting your own intuitions and biases influence selection of the sample, you need to use an impartial method for selecting the sample. This way the sample is likely to have characteristics that are similar to the population. In this lesson, you will learn about simple random sampling—a method for selecting a sample from a population that avoids selection bias.

Task 1: Simple Random Sampling [Student Handout, estimated time: 30 minutes]

To avoid selection bias, you want to select samples in a manner that ensures every individual in the population has the same chance of being selected. One way to do this is to select a simple random sample. A simple random sample of size \( n \) is a sample selected in a way that ensures every different possible sample of size \( n \) has the same chance of being chosen. For example, to select a sample of five fish from the fish population, you could number the fish in the tank from 1 to 91 and then take one of the following approaches to obtain a simple random sample of fish:

- Write the numbers from 1 to 91 on identical slips of paper, place the slips in a bag, mix well, and select 5 slips. These slips identify which fish will be included in the sample.
- Use a table of random numbers, a random number generator, or a calculator/computer to generate five random numbers between 1 and 91. These five numbers identify which fish will be included in the sample.

(Note: If you plan to use a random number table or technology to generate random numbers, take a few minutes to show students how to do this. If using a random number table, use two-digit numbers and make sure to include a way of choosing a starting point so that students do not all end up selecting the same five fish.)

1. Using an appropriate method, select a simple random sample of five fish. Indicate which fish (by number) will be included in your sample.

2. Using the fish picture provided by your instructor where the fish have been numbered from 1 to 91, find the five fish that correspond to the numbers selected for your sample. Determine the lengths of these five fish.

(Note: You can provide the second picture of the fish where the lengths are indicated or the key that gives the length of each type of fish. Another option is having students use the ruler to measure the five fish selected.

Ask students whether they think this is a practical plan for selecting a sample of fish. You can acknowledge that this is an artificial setting, but that it has been chosen because it illustrates some important principles of sampling. Point out that if these were real fish in a real tank, it might be impractical to put identification numbers on the fish and then select those particular identification numbers (though this is sometimes done in nature studies to track migration of large animals).
Supporting Lesson 1.2.2: Random Sampling

Remind them that this method ensures every different possible sample of five fish is equally likely to be selected. Students might also suggest taking a photograph of the tank and selecting fish from that, but this assumes the individual fish can be identified and tracked down from the photograph. Let students know that alternative, but still random, methods will be discussed in later lessons, such as taking every 10th fish as they are passed between two tanks.)

(3) Compute the average length of the five fish in your sample.

(4) Add your average to the dotplot that displays the averages for the random samples for all students in the class.

(Note: Construct the class dotplot on a piece of flip chart paper or on an overhead transparency, depending on how you made the dotplots in Lesson 1.2.1.

Draw a scale that matches the one used for the “best guess” dotplot and add the title “Random Sample Average.” Have each student add his or her average to the dotplot. You can now talk about the distribution of random sample averages. Was there general consensus around a particular value? Was there variability from student to student? Now show the two dotplots from Lesson 1.2.1. Have students directly compare the results to the guesses and averages from the nonrandom samples, specifically highlighting the change in center in comparison to the population value.)

Wrap-Up [Instructor-led discussion]

Begin the wrap-up by locating the population average length (3.1 centimeters) on the dotplot of random sample averages. Note that the random sample averages tend to center around 3.1 centimeters. Then show the two dotplots constructed in the previous lesson and compare the dotplot constructed using averages from random samples to the other two dotplots.

A key point to drive home here is that the simple random sample method eliminates selection bias. Even though you do not hit the population average head-on every time, you are just as likely to be too low as too high. There is no longer a consistent tendency to overestimate or underestimate the population average.

Next ask students what they think the dotplot would look like if they had taken a simple random sample of 10 fish instead of 5 fish. To help students answer this question, ask them whether they think larger samples tend to produce sample averages that are closer to or farther away from the actual population average. (You can refer to other contexts—like more questions on an exam or multiple attempts in sporting events, such as the Olympic long jump or rodeo events.) Then ask what this implies about the dotplot in terms of variability from sample to sample. See if you can lead the discussion to get students to say that they think the dotplot would still be centered around 3.1 centimeters, but there would probably be less variability from student to student (sample to sample)—the dots would cluster more tightly around 3.1 centimeters.

Finally, ask students whether they think starting with a sample size of 10 rather than 5 when they visually picked a representative sample would have fixed the problem observed when self-selected samples were used (especially if they had a larger population). Emphasize that random selection is actually more important than sample size in allowing you to be comfortable that the sample is representative of the population.

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™, Statway™, and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Supporting Lesson 1.2.2: Random Sampling

Task 2: Investigating Sampling Method and Sample Size [Student Handout, estimated time: 10–15 minutes]

(Note: This task can be done as a whole-class discussion or by having students work in pairs or small groups until the wrap-up, which is done as whole-class discussion.)

Phil is a compulsively neat book collector. He has 500 books arranged from shortest to tallest on shelves in his den. You can number the books from 1 to 500, starting with the shortest book, as shown below. (Only a few of the 500 books are shown, but imagine there are 500 books arranged from shortest to tallest on 25 shelves of 20 books each.)

(Note: Take a minute at this point to make sure students understand the picture of books on the shelves and how the books are arranged.)
Supporting Lesson 1.2.2: Random Sampling

(5) Phil is expanding his book collection and wants to build additional bookcases to hold the new books. He needs to decide how tall the shelves in the new bookcases should be, so he plans to estimate the average height for the population of 500 books that he already owns. Phil will select a sample of books and compute the average height of the books in the sample. Describe a method that he could use to select a simple random sample of 10 books from the population of 500 books.

(Answer: To obtain a simple random sample, Phil can number all the books (1–500) and then use a random number table or a calculator to select 10 identification numbers. The corresponding 10 books would be his sample.)

One hundred different simple random samples of 10 books were selected and the average height (in inches) for each sample was computed. A dotplot of the 100 sample averages is shown here:

[Dotplot image]

Following are two more dotplots. One was constructed by plotting the sample averages for 100 different simple random samples of 5 books and the other by plotting the sample averages for 100 different simple random samples of 20 books.

[Dotplot images]
Supporting Lesson 1.2.2: Random Sampling

(6) Which of the two dotplots is the plot of the sample averages for size 20? Why do you think this?

(Answer: Dotplot 2 shows averages from samples with the larger sample size. You expect there to be less sample-to-sample variation with larger samples. The sample averages cluster more closely to the population average with the larger sample size.)

(7) Consider the following proposed methods for selecting a sample from the population of books.

- Method A: Select the first 50 books (those numbered 1 to 50).
- Method B: Select a simple random sample of 20 books.

If the books in the sample will be used to estimate the average height of books in the population, which method is better (more likely to result in a sample average close to the actual population average) for selecting the sample? Explain why you chose this method.

(Answer: Method A results in the smallest possible sample mean. Larger books have no chance of being represented in this sample. Method B selects books from throughout the entire population and should result in a representative sample that includes larger books, medium-sized books, and smaller books. Even with the smaller sample size, these 20 books will more likely reflect the characteristics of the population.)

(8) Consider the following proposed methods for selecting a sample from the population of books.

- Method C: Select a simple random sample of 20 books.
- Method D: Select a simple random sample of 50 books.

Which method do you recommend for selecting a sample of books to estimate the average height of books in the population? Explain why you chose this method.

(Answer: Method D. If both methods involve random samples, the larger the random sample the more likely you are to obtain a sample average close to the population mean.)

(9) Is it a good idea for Phil to build the bookcase for his new books by making every shelf the same height and using the estimate of the average book height for the shelf height? Explain your reasoning.
Supporting Lesson 1.2.2: Random Sampling

(Answer: No, this does not take into account the variability of the heights around that mean. It is quite possible that many of the books are taller than the average book height and would not fit on his new bookcase if he builds all shelves with a height equal to the average height.

Note: Students may ask whether some of the old books can be moved to the new bookcase in order to find a home for taller new books. Tell them that this is not what the question is asking—it indicates that the new bookcase will hold the new books.)

Wrap-Up [Instructor-led discussion]

Key points for the wrap-up include the following:

- Simple random sampling eliminates selection bias.
- For simple random samples, larger samples tend to produce estimates that are more accurate.
- Increasing the sample size cannot make up for a poor sampling plan.
Supporting Lesson 1.2.2: Random Sampling

Task 3: Not What You Might Think... [Student Handout, estimated time: 15 minutes]

(Note: This task summarizes a rather counterintuitive idea and is probably best accomplished by instructor-led discussion. Some suggestions for how this discussion might be broken up are embedded in the task. Ask students to read the first two paragraphs, or you can incorporate them into a minilecture.)

When sampling from a population, a common misconception is that the larger the population, the larger the sample size needs to be in order to produce accurate estimates of population characteristics.

Consider a population of 5,000 students over age 18 who are enrolled at a particular college; you are interested in the proportion who are registered to vote. Suppose 3,500 of these students are registered to vote and 1,500 are not registered. Think of this population as consisting of 5,000 values:

\[
\begin{array}{c|c|c}
& \text{R} & \text{N} \\
3,500 & 1,500 \\
\end{array}
\]

where \( R \) stands for registered to vote and \( N \) stands for not registered.

(Note: Make sure students understand what the population is and the notation above. Then have them read the next paragraph and take a look at the bar chart.)

Notice that the values in this population are not numbers. One way to describe non-numerical data graphically is to use a bar chart. In a bar chart, each possible value in the non-numerical data set (like registered to vote and not registered) is represented by a bar. The height of the bar corresponds to the number of times that value appears in the data set.

For the student population previously described, you could represent the population using a bar chart with two bars—one for registered and one for not registered.

(Note: Make sure that students understand how the bar chart represents the population and what the height of each bar represents. You can then have students read the next section on converting counts to proportions, or you can incorporate this into a minilecture. This is also a place where you could add further discussion of percentages if you think it is appropriate for your students.)
Supporting Lesson 1.2.2: Random Sampling

It is also common to use the proportion corresponding to each value to determine the height of the bars in a bar chart. The proportion is a decimal between 0 and 1 that is computed by dividing a category count by the total number of data values in the data set. For example, the proportion for the Registered category in the population of 5,000 students is

\[
\frac{3,500}{5,000} = 0.70
\]

and the proportion for the Not Registered category is

\[
\frac{1,500}{5,000} = 0.30
\]

A bar chart using the category proportions is shown here:

(Note: You can also have students discuss the results and bar graph in terms of percentages and give them practice going between proportions and percentages. Again, you can either have students read this next small section or incorporate it into a minilecture.)

Now let's consider what happens when you take random samples from this population. Think of students in the population as being numbered from 1 to 5,000, with the first 3,500 (those with numbers 1 to 3,500) corresponding to students who are registered to vote and those numbered 3,501 to 5,000 corresponding to those who are not registered.

Random numbers were used to select 50 students, resulting in the following 50 data values:

R  N  R  N  N  N  R  N  N  R  R  
R  N  N  R  R  R  N  R  R  R  N  
R  R  R  N  R  N  R  R  N  R  
R  R  R  N  R  N  R  R  N  R  
R  R  N  N  R  R  R  R  N  R  
R  R  N  N  R  R  R  R  N  R
Supporting Lesson 1.2.2: Random Sampling

A bar chart for this simple random sample is shown here:

![Bar Chart for One Sample of 50 Students]

(10) Approximately what proportion of students in the sample of 50 is registered to vote? How does this compare to the actual population value of 0.70? Are these values equal? Does this surprise you?

(Answer: About 0.65 [or 65%]. This is less than the population proportion of 0.70. Whereas you expect the sample proportion to be somewhat close to the population proportion, because only a sample was selected you do not expect the sample proportion and population proportion to be exactly equal. So, it is not surprising that the sample proportion and population proportion are different.)

(11) If you selected another simple random sample of 50 students, do you think the sample proportion would be exactly equal to the sample proportion in Question 10? Why or why not?

(Answer: No, the sample proportions will vary from sample to sample. Some will be larger than 0.70, some smaller.)

One hundred simple random samples of size 50 were selected from the population of 5,000 students, and the proportion of students who were registered to vote was computed for each sample. These sample proportions were used to construct the following dotplot.
Supporting Lesson 1.2.2: Random Sampling

(12) Do the sample proportions tend to center around the actual population proportion of 0.70?

(Answer: Yes, the middle of this distribution appears to be around 0.70, with about half the sample proportions below 0.70 and about half above 0.70.)

(13) If samples of size 100 are used instead of samples of size 50, how does the dotplot of the sample proportions compare to the one for samples of size 50?

(Answer: The dotplot is still centered around 0.70 but is less spread out due to smaller sample-to-sample variability with the larger sample size.)

In general, for simple random sampling, the larger the sample size, the more closely sample estimates of a population average or a population proportion tend to cluster around the actual population value. This is true for any population size. However, what is surprising to most people is that unless the population size is quite small, the accuracy of the estimates depends on the sample size, not on the population size.

Using a sample of size 500 to estimate a population proportion produces estimates of about the same accuracy whether the population size is 10,000 or 100,000 or 1,000,000. To illustrate this, consider the following three populations:

- Population A: 10,000 people, 7,000 of whom are registered to vote
- Population B: 100,000 people, 70,000 of whom are registered to vote
- Population C: 1,000,000 people, 700,000 of whom are registered to vote

(Note: Ask students to calculate the proportion who are registered to vote for each population so that they can see that this proportion is the same for all three populations.)

One hundred different simple random samples of size 500 were selected from Population A and the 100 sample proportions were used to make the dotplot labeled “Population Size 10,000” in the following graph. This process was repeated (100 simple random samples each of size 500) for the other two populations.

(14) How do these three dotplots illustrate that the accuracy expected when using a simple random sample to estimate a population proportion does not depend on the population size?
Supporting Lesson 1.2.2: Random Sampling

(Answer: The shapes, centers, and spreads of these distributions appear similar. All three are mound-shaped, centered around 0.70, and have values between 0.65 and 0.75, with most of the values between about 0.68 and 0.73. The change in the population size does not appear to affect the pattern in the resulting sample proportions.)

Wrap-Up [Instructor-led discussion]

Key points for the wrap-up include the following:

- For simple random sampling, the larger sample size results in better accuracy (produces sample estimates that cluster more closely to the population value).
- For simple random sampling, accuracy depends on sample size but not population size (as long as the population size is large).
- Sample size and how the sample is selected are more important in determining how accurate estimates will be than population size (again assuming a pretty large population).

If students ask, large population means populations that are at least 10 times larger than the size of the sample.

Homework [Student Handout]

(1) Suppose you want to select a simple random sample of 75 students from the full-time students at your college in order to learn about the average number of hours students at the college spend online in a typical day. A list of all full-time students arranged in alphabetical order is available. Describe how you would go about selecting a simple random sample of 75 students from this population.

(Answer: Number all the students on the alphabetical list, use a random number table or calculator [or numbered slips of paper in a hat] to select 75 numbers, determine which students correspond to those numbers on the list, and select them for the sample. Then ask each of the 75 selected students how many hours he or she spends online in a typical day.)

(2) You want to estimate the average amount of time that students at a particular college spend studying in a typical week. Which of the following sampling methods do you recommend? For each method, explain why you did or did not select it as the recommended method.

- Method A: Select 50 students at random for the students at the college.
- Method B: Select 100 students at random from the students at the college.
- Method C: Select the 200 students enrolled in introductory statistics at the college this semester.
- Method D: Select the 300 students enrolled in English literature at the college this semester.

(Answers: Method A: This is reasonable because it selects a random sample from all students.  
Method B: This is reasonable like Method A, but is preferred because the sample size is larger.  
Method C: This is not reasonable because students in a statistics class probably differ from other students at the college with respect to how much time they spend studying in a typical week.)
Method D: This is not reasonable because students in an English literature class are probably not representative of all students at the college with respect to how much time they spend studying in a typical week.)

(3) Several years ago a group of people worked to get an initiative on the California state ballot. The initiative proposed adding “none of the above” to the list of candidates running for office. A spokesperson for the initiative was critical of a poll that showed the measure was losing by 10 percentage points (55% percent were against the measure and 45% were for the measure). He questioned the result of the poll because it was based on a random sample of 1,000 registered voters in California. He commented that because the population of California is so large (there are about 23.5 million registered voters in California), a sample of only 1,000 voters could not possibly provide an accurate estimate of the proportion of registered voters in California who support the ballot initiative (Associated Press, January 30, 2000).

(a) Is his criticism valid? Explain why or why not.

(Answer: This is not accurate because a random sampling method was used. This should produce a representative sample regardless of the population size as long as the sample was selected from a list of all registered voters in California.)

(b) Would the criticism be valid if this had been a national initiative and 1,000 people were randomly selected from all adult Americans who were registered to vote?

(Answer: Random sampling should produce a representative sample regardless of the population size. A sample of 1,000 Californian registered voters has the same level of accuracy as a random sample of all registered Americans.)
Supporting Lesson 1.2.2: Random Sampling

In Lesson 1.2.1, you saw that some methods of selecting a sample from a population can lead to misleading conclusions about a population. In that lesson, informal “first impression” estimates, where you guessed at the average fish length after looking at the fish population for a short time, tended to overestimate the actual average fish length. This may not be surprising as you only had a short time to look at the fish population. You also saw that when classmates attempted to pick a sample that was representative of the population and then use the fish in the sample to estimate the average length of fish in the population, the sample averages still consistently tended to be too large.

This happened because you overlooked the many small fish in the population—your attention was drawn to the large fish, causing them to be overrepresented in the sample. If some individuals in the population are more likely to be included in the sample than others, the sampling method results in selection bias. Even if you tried to pick one fish of each type, this ignores that there are more small fish than large fish in this tank. More generally, you have selection bias when some types of individuals in the population (like the small fish) are chosen in a manner that does not represent how often they occur in the population.

Instead of letting your own intuitions and biases influence selection of the sample, you need to use an impartial method for selecting the sample. This way the sample is likely to have characteristics that are similar to the population. In this lesson, you will learn about simple random sampling—a method for selecting a sample from a population that avoids selection bias.

Task 1: Simple Random Sampling

To avoid selection bias, you want to select samples in a manner that ensures every individual in the population has the same chance of being selected. One way to do this is to select a simple random sample. A simple random sample of size \( n \) is a sample selected in a way that ensures every different possible sample of size \( n \) has the same chance of being chosen. For example, to select a sample of five fish from the fish population, you could number the fish in the tank from 1 to 91 and then take one of the following approaches to obtain a simple random sample of fish:

- Write the numbers from 1 to 91 on identical slips of paper, place the slips in a bag, mix well, and select 5 slips. These slips identify which fish will be included in the sample.
- Use a table of random numbers, a random number generator, or a calculator/computer to generate five random numbers between 1 and 91. These five numbers identify which fish will be included in the sample.

(1) Using an appropriate method, select a simple random sample of five fish. Indicate which fish (by number) will be included in your sample.
(2) Using the fish picture provided by your instructor where the fish have been numbered from 1 to 91, find the five fish that correspond to the numbers selected for your sample. Determine the lengths of these five fish.

(3) Compute the average length of the five fish in your sample.

(4) Add your average to the dotplot that displays the averages for the random samples for all students in the class.
Supporting Lesson 1.2.2: Random Sampling

Task 2: Investigating Sampling Method and Sample Size

Phil is a compulsively neat book collector. He has 500 books arranged from shortest to tallest on shelves in his den. You can number the books from 1 to 500, starting with the shortest book, as shown below. (Only a few of the 500 books are shown, but imagine there are 500 books arranged from shortest to tallest on 25 shelves of 20 books each.)
(5) Phil is expanding his book collection and wants to build additional bookcases to hold the new books. He needs to decide how tall the shelves in the new bookcases should be, so he plans to estimate the average height for the population of 500 books that he already owns. Phil will select a sample of books and compute the average height of the books in the sample. Describe a method that he could use to select a simple random sample of 10 books from the population of 500 books.

One hundred different simple random samples of 10 books were selected and the average height (in inches) for each sample was computed. A dotplot of the 100 sample averages is shown here:

![Dotplot of sample averages](image)

Following are two more dotplots. One was constructed by plotting the sample averages for 100 different simple random samples of 5 books and the other by plotting the sample averages for 100 different simple random samples of 20 books.
Supporting Lesson 1.2.2: Random Sampling

(6) Which of the two dotplots is the plot of the sample averages for size 20? Why do you think this?

(7) Consider the following proposed methods for selecting a sample from the population of books.

- Method A: Select the first 50 books (those numbered 1 to 50).
- Method B: Select a simple random sample of 20 books.

If the books in the sample will be used to estimate the average height of books in the population, which method is better (more likely to result in a sample average close to the actual population average) for selecting the sample? Explain why you chose this method.
### Supporting Lesson 1.2.2: Random Sampling

(8) Consider the following proposed methods for selecting a sample from the population of books.

- Method C: Select a simple random sample of 20 books.
- Method D: Select a simple random sample of 50 books.

Which method do you recommend for selecting a sample of books to estimate the average height of books in the population? Explain why you chose this method.

(9) Is it a good idea for Phil to build the bookcase for his new books by making every shelf the same height and using the estimate of the average book height for the shelf height? Explain your reasoning.
Supporting Lesson 1.2.2: Random Sampling

Task 3: Not What You Might Think...

When sampling from a population, a common misconception is that the larger the population, the larger the sample size needs to be in order to produce accurate estimates of population characteristics.

Consider a population of 5,000 students over age 18 who are enrolled at a particular college; you are interested in the proportion who are registered to vote. Suppose 3,500 of these students are registered to vote and 1,500 are not registered. Think of this population as consisting of 5,000 values:

\[
\frac{R}{N} = \frac{3,500}{1,500}
\]

where \( R \) stands for registered to vote and \( N \) stands for not registered.

Notice that the values in this population are not numbers. One way to describe non-numerical data graphically is to use a bar chart. In a bar chart, each possible value in the non-numerical data set (like \textit{registered to vote} and \textit{not registered}) is represented by a bar. The height of the bar corresponds to the number of times that value appears in the data set.

For the student population previously described, you could represent the population using a bar chart with two bars—one for \textit{registered} and one for \textit{not registered}.

![Bar Chart]

It is also common to use the proportion corresponding to each value to determine the height of the bars in a bar chart. The proportion is a decimal between 0 and 1 that is computed by dividing a category count by the total number of data values in the data set. For example, the proportion for the Registered category in the population of 5,000 students is

\[
\frac{3,500}{5,000} = 0.70
\]

and the proportion for the Not Registered category is

\[
\frac{1,500}{5,000} = 0.30
\]
Supporting Lesson 1.2.2: Random Sampling

A bar chart using the category proportions is shown here:

![Bar chart showing population proportions registered and not registered.]

Now let’s consider what happens when you take random samples from this population. Think of students in the population as being numbered from 1 to 5,000, with the first 3,500 (those with numbers 1 to 3,500) corresponding to students who are registered to vote and those numbered 3,501 to 5,000 corresponding to those who are not registered.

Random numbers were used to select 50 students, resulting in the following 50 data values:

```
R  N  R  N  N  N  R  N  N  R  R
R  N  N  R  R  R  N  R  R  N  N
R  R  R  N  R  N  R  R  N  R
R  R  R  N  R  N  R  R  N  R
R  R  N  N  R  R  R  R  R  N  R
```

A bar chart for this simple random sample is shown here:

![Bar chart showing one sample of 50 students.]

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Supporting Lesson 1.2.2: Random Sampling

(10) Approximately what proportion of students in the sample of 50 is registered to vote? How does this compare to the actual population value of 0.70? Are these values equal? Does this surprise you?

(11) If you selected another simple random sample of 50 students, do you think the sample proportion would be exactly equal to the sample proportion in Question 10? Why or why not?

One hundred simple random samples of size 50 were selected from the population of 5,000 students, and the proportion of students who were registered to vote was computed for each sample. These sample proportions were used to construct the following dotplot.

(12) Do the sample proportions tend to center around the actual population proportion of 0.70?
Supporting Lesson 1.2.2: Random Sampling

(13) If samples of size 100 are used instead of samples of size 50, how does the dotplot of the sample proportions compare to the one for samples of size 50?

In general, for simple random sampling, the larger the sample size, the more closely sample estimates of a population average or a population proportion tend to cluster around the actual population value. This is true for any population size. However, what is surprising to most people is that unless the population size is quite small, the accuracy of the estimates depends on the sample size, not on the population size.

Using a sample of size 500 to estimate a population proportion produces estimates of about the same accuracy whether the population size is 10,000 or 100,000 or 1,000,000. To illustrate this, consider the following three populations:

- Population A: 10,000 people, 7,000 of whom are registered to vote
- Population B: 100,000 people, 70,000 of whom are registered to vote
- Population C: 1,000,000 people, 700,000 of whom are registered to vote

One hundred different simple random samples of size 500 were selected from Population A and the 100 sample proportions were used to make the dotplot labeled “Population Size 10,000” in the following graph. This process was repeated (100 simple random samples each of size 500) for the other two populations.

(14) How do these three dotplots illustrate that the accuracy expected when using a simple random sample to estimate a population proportion does not depend on the population size?
Supporting Lesson 1.2.2: Random Sampling

Homework

(1) Suppose you want to select a simple random sample of 75 students from the full-time students at your college in order to learn about the average number of hours students at the college spend online in a typical day. A list of all full-time students arranged in alphabetical order is available. Describe how you would go about selecting a simple random sample of 75 students from this population.

(2) You want to estimate the average amount of time that students at a particular college spend studying in a typical week. Which of the following sampling methods do you recommend? For each method, explain why you did or did not select it as the recommended method.

- Method A: Select 50 students at random for the students at the college.
- Method B: Select 100 students at random from the students at the college.
- Method C: Select the 200 students enrolled in introductory statistics at the college this semester.
- Method D: Select the 300 students enrolled in English literature at the college this semester.
Supporting Lesson 1.2.2: Random Sampling

(3) Several years ago a group of people worked to get an initiative on the California state ballot. The initiative proposed adding “none of the above” to the list of candidates running for office. A spokesperson for the initiative was critical of a poll that showed the measure was losing by 10 percentage points (55% percent were against the measure and 45% were for the measure). He questioned the result of the poll because it was based on a random sample of 1,000 registered voters in California. He commented that because the population of California is so large (there are about 23.5 million registered voters in California), a sample of only 1,000 voters could not possibly provide an accurate estimate of the proportion of registered voters in California who support the ballot initiative (Associated Press, January 30, 2000).

(a) Is his criticism valid? Explain why or why not.

(b) Would the criticism be valid if this had been a national initiative and 1,000 people were randomly selected from all adult Americans who were registered to vote?
Supporting Lesson 1.2.3: Other Sampling Strategies

Estimated number of 50-minute class sessions: 1

Learning Goals
Students will begin to understand
- that there are reasonable alternatives to simple random sampling.
- why it is best to avoid convenience samples and volunteer response samples.
- that it is not always feasible to select a simple random sample from a population of interest.
- how stratified random sampling can be preferable to simple random sampling if there is an effective way to stratify.

Students will begin to be able to
- distinguish between simple random sampling, stratified random sampling, and systematic sampling.
- explain why it is desirable to create homogeneous subgroups when considering stratified random sampling.
- identify characteristics of a good sampling plan.

Introduction [Student Handout]
Even though simple random sampling is a way to select a sample that is objective and avoids selection bias, it is not always easy to select a simple random sample and sometimes it can be impractical. In such situations, you want to consider other sampling strategies that produce a sample that is representative of the population. In this lesson, systematic sampling and stratified random sampling are considered.

Task 1: Systematic Sampling [Student Handout, estimated time: 10 minutes]
Suppose you are interested in learning how far the people at a football game traveled to see the game. There is a sellout crowd of 20,000 people. For simplicity, assume that the seats in the stadium are numbered from 1 to 20,000.

1. Describe a method for selecting a simple random sample of 200 people from the population of people attending the game.

   (Answer: Use a random number table or a random number generator to select the seat numbers for 200 seats, find the 200 people sitting in those seats, and survey them.

   Note: Make sure students focus on the list that identifies the people by seat number, selecting the seat numbers, and then finding the people in those seats. Students could suggest using a list of all ticket buyers, but that is even less realistic and more impractical.)

2. Is the plan you described in Question 1 easy to implement? What are some difficulties you would face in actually collecting data on how far the individuals in the sample had traveled to attend the game?

   (Answer: This does not seem practical. It would be difficult and time consuming to track down the 200 people that were selected. People may not be in the seat when you get there, and it could be difficult to get around to all of the selected seats in a big stadium—especially if there are only a small number of people available to survey the selected people.)
Supporting Lesson 1.2.3: Other Sampling Strategies

(3) Suppose someone proposes posting two people at each of the four entrances to the stadium. At each entrance, one person is in charge of counting the people passing through the entrance in order to identify every 100th person entering. The second person asks each person identified in this way how far he or she traveled to attend the game. This should result in a sample of about 200 people (assuming roughly 5,000 people go through each entrance and selecting every 100th person will identify about 50 people at each entrance). Do you think this strategy for selecting the sample is reasonable? Why do you think so?

(Answer: There should not be any selection bias with this method, and it should produce a representative sample. There is no reason to suspect that those traveling longer or shorter distances are more likely to be selected.

Note: If the two people working at each entrance begin by selecting a random number between 1 and 100 to identify the first person to be included in the sample and then select every 100th person thereafter, all individuals in the population have an equal chance of being selected.)

The sampling strategy described in Question 3 is an example of systematic sampling. In systematic sampling, the individuals in the population are arranged in some order (e.g., in a list of the members of the population or in the previous example by the order they enter the stadium). Then the sample is selected in a “systematic” way by choosing (usually randomly) a number \( k \), such as 100 in the previous example, and then choosing every \( k \)th individual. It is also a good idea to start the process by selecting a random number between 1 and \( k \) to identify the first person to be included in the sample. For example, if the selected number was 38, the 38th person through the gate is included in the sample and then you select the 138th person, the 238th, and so on.

(4) Suppose an amusement park is open from 8 a.m. to 10 p.m. every day of the year. You are interested in learning about the average daily attendance at the park. One way to learn about this average is to choose a sample of days and record the attendance for those days. Is it a good idea to select a systematic sample of days by recording attendance every 14th day for one year? Why or why not?

(Answer: The problem here is with the number 14. This results in always selecting the same day of the week. Attendance patterns are quite possibly related to day of the week. For example, if you always select Saturdays, then attendance is probably higher than other days and so the sample average is probably greater than the actual average attendance. On the other hand, if you always select Wednesdays, then attendance values probably tend to be smaller than other days, and the sample average is probably smaller than the actual average.)

In general, systematic sampling works well when there are no repeating patterns in the population list or ordered arrangement of the population that might be related to the variable you are interested in (e.g., park attendance).
Supporting Lesson 1.2.3: Other Sampling Strategies

Task 2: Exploring Sampling Variability [Student Handout, estimated time: 10 minutes]

(Note: This task works well as a quick whole-class discussion.)

Consider a small population consisting of 20 students. Each student was asked the number of units he or she was enrolled in this semester, resulting in the following data:

15 15 15 15 15 15 15 15 15 15
15 15 15 15 15 15 15 15 15 15

(5) What is the average number of units for this population?

(Answer: Summing up all the values and dividing by 20 gives 15 units as the average.)

(6) If you select a sample of two students from this population at random, how would the value of the sample average compare to the population average? If you select a different sample of size 2, would you get the same value for the sample average?

(Answer: The sample average will always be 15, regardless of the sample size. There will be no sample-to-sample variability in this case.)

(7) What aspect of the population provides an explanation for your answers in Question 6?

(Answer: Because there is no variability in the population, every sample consists of the same two values. Therefore, the sample average is always the same [15]. There is no sample-to-sample variability.)

Now consider a different population of 20 students, with the following values for number of units:

14 16 18 16 12 16 13 17 13 12
14 18 17 15 18 16 16 12 14 13

(8) What is the average number of units for this population?

[Answer: \((14 + 16 + \ldots + 14 + 13)/20 = 15\) units]

(9) If you select a sample of two students from this population at random, would the value of the sample average be equal to the population average? If you select a different sample of size 2, would you get the same value for the sample average?

(Answer: The sample average may or may not equal the population average. You also will not get the same value for the sample average each time you select a sample from this population. For example, you could end up with a average as small as 12 or as large as 18.)

(10) Are your answers to Questions 6 and 9 different? If so, why are they different?

(Answer: The answers differ; in the second case, there is variability in the population. Because of this, you can potentially obtain different values for the sample average across the different samples.)
Supporting Lesson 1.2.3: Other Sampling Strategies

(Note: If you have time, give students another population [with mean 15] that has even more sample variability and ask them to conjecture how that affects the amount of sample to sample variation.

| 17 | 17 | 19 | 14 | 24 | 20 | 13 | 22 | 15 | 15 |
| 19 | 13 | 10 | 13 | 14 | 15 | 10 | 11 | 9  | 10 |

Point out that a sample of 2 may potentially result in even smaller sample averages [e.g., 9.5] or even larger sample averages [e.g., 23].)

Task 3: Stratified Random Sampling [Student Handout, estimated time: 20 minutes]

Consider the following two populations:

Population 1: All students who received an associate’s degree from your college in the past academic year.

Population 2: All students who received an associate’s degree from your college in the past academic year and graduated with a grade point average (GPA) of 3.5 or higher.

(11) If you are interested in the GPAs of students in the population, which population would have more variability? In other words, for which population would GPA tend to differ more from student to student? Explain why you chose this population.

(Answer: The GPAs for Population 1 will be more variable. GPAs in Population 1 range from 2.0 to 4.0, whereas in Population 2, they only range from 3.5 to 4.0.)

(12) Suppose you take a random sample of size 50 from each of these two populations and compute the sample averages. Which population do you think the sample average is closer to the actual population average? Why do you think this? (Hint: Think about your answers in Task 2.)

(Answer: Although you do not know for sure, you suspect that the sample average from Population 2 will be closer to the population average than the sample average from Population 1. This is because the sample sizes are equal, but Population 1 has more variability in the GPA values. This additional variability in the population leads to more variability in the sample averages from sample to sample as well [as discussed in Task 2]. It is more likely that you get a sample average far from the population average with Population 1 than Population 2.)

When there is not much variability in a population, even a small sample can yield a fairly accurate estimate of the population average. For example, because the GPAs of students who graduate with a GPA of 3.5 or higher are quite similar, even a sample of 5 or 10 students from this group would tend to produce an average GPA that is a pretty accurate estimate of the population average. On the other hand, if the individuals in a population tend to be more variable, it is not as easy to produce an accurate estimate. For example, for the population of all students graduating with an associate’s degree (a much more diverse group with respect to GPA than the population of students graduating with a GPA of 3.5 or higher), a larger sample is needed to produce an accurate estimate of the population average.

(Note: Remind students that there is less sample-to-sample variability in sample averages with larger samples. So even with a diverse population, with large samples there may not be much sample-to-sample variability, and any one sample should provide a reasonably accurate of the population mean.)
Supporting Lesson 1.2.3: Other Sampling Strategies

Therefore, you are leading students to keep in mind two possible strategies for obtaining an accurate estimate—more homogenous populations or larger sample sizes.)

**Stratified random sampling** is a sampling method that takes advantage of the fact that you can get good information from small samples if you are sampling from a group that does not have much variability. In stratified random sampling, the population of interest is divided up into nonoverlapping subgroups, and then a simple random sample is taken from each subgroup. The results from the subgroups are then pooled together to produce one estimate for the overall population. If the subgroups are less variable than the population as a whole, then a stratified sample can produce more accurate estimates of population characteristics than a simple random sample of the same size.

(Comment: Pooling together refers to a weighted average of the results, with the weights being proportional to the strata sizes. This is not included in the Student Handout since the calculation of such averages falls out of the scope of the Statway curriculum.)

(13) Suppose you are interested in estimating the average amount of money spent on textbooks this semester by students at your college. You plan to use stratified random sampling to select a sample from the population of students at the college. Which of the following is a better way to create the subgroups? Explain your reasoning.

**Option A:** Use subgroups based on field of study by dividing students into three groups: humanities, liberal arts, and science and mathematics.

**Option B:** Use subgroups based on the first letter of last name by dividing students into three groups—A to H, I to P, and Q to Z.

(Note: Remind students that the goal is to create subgroups that are reasonably homogenous particularly with respect to the variable of interest. In this case, you expect textbook expenses to be more similar for students within the same discipline. On the other hand, you do not expect textbook prices to be related to last name groupings. For example, textbook prices for students in the humanities should be pretty similar. However, there is still lots of variability in textbook prices among students with last names starting with A to H. Therefore, to create more similar groups, the first strategy is preferred.)

**Task 4: One Method to Avoid!** [Student Handout, estimated time: 10 minutes]

When evaluating a sampling plan, it is important to think carefully about whether the plan is likely to produce a sample that is representative of the population. One strategy that is almost never a good idea is to use a *convenience sample*. This method does not use random selection and involves using an easily available or “convenient” group to form a sample. Data from a convenience sample should not be used to generalize to any larger population.

One common form of convenience sampling is to rely on volunteers who respond to an advertisement or who must call in to a telephone number or log on to the Internet to participate in a survey.

(14) Science textbooks are among the most expensive books sold in the campus bookstore. Suppose you are interested in selecting a sample of students at your college to estimate the average amount of money spent on textbooks in a particular semester. Explain why it would not be a good idea to use the 100 students enrolled in a Chemistry course as the sample.

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™, Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.aftdanacenter.org/mathways.
Supporting Lesson 1.2.3: Other Sampling Strategies

(Answer: The data from these students represent what is paid for textbooks by students taking Chemistry that semester, a course that probably has higher textbook costs than other courses. Many students at your college are not taking science courses that semester [presumably having lower textbook costs], and they are not available for your sample. This will probably lead to a sample average that is higher than the population value.)

(15) Suppose you are interested in estimating the average distance from the primary residence to campus for students at your college. Because students who are enrolled in an online course communicate with the college by email, it would be easy to contact these students by email to ask them how far away from campus they live. Is this a good way of selecting the sample? Explain your reasoning.

(Answer: This is not appropriate because those who choose to take an online course probably do so because coming to campus is more difficult for them. This will probably overrepresent students who live further from campus.)

Wrap-Up [Instructor-led discussion]

The following are main points to include:

- There are many different methods for selecting a sample from a population. Some methods are better than others!
- Methods that use randomness to choose the sample from the entire population without selection bias allow you to believe the sample is representative of the population.
- You can further increase the accuracy of sample estimates (reduce the sample-to-sample variability) if you use stratified random sampling with subgroups that are homogeneous with respect to the variable of interest.

Homework [Student Handout]

(1) The article, “I’d Like to Buy a Vowel, Drivers Say” (USA Today, August 7, 2001) speculates that young people prefer automobile names consisting of just numbers and/or letters that do not form a word (such as Hyundai’s XG300, Mazda’s 626, and BMW’s 325i). The article says that Hyundai had planned to identify the car that was eventually marketed as the XG300 with the name Concerto until it determined that consumers thought XG300 sounded more “technical” and deserving of a higher price. How much more would students at your college be willing to pay for an XG300 than a comparable car called Concerto?

Suppose a list of all the students at your college is available. For the purpose of this exercise, assume that there are 15,000 students at the college.

(a) Describe a method for selecting a simple random sample of 300 students from this population.

(Answer: Number everyone on the list from 00001 to 15000. Use a random number table or a random number generator to select 300 five-digit numbers. Find those identification numbers on the list and the contact the 300 people corresponding to those identification numbers.)

(b) Describe a method for selecting a systematic sample of about 300 students from this population.
Supporting Lesson 1.2.3: Other Sampling Strategies

(Answer: 15,000/300 = 50, so you want to select every 50th person from the list to end up with a sample of 300 students. It is a good idea to use a random number between 1 and 50 to determine the starting point.)

(c) Suppose the list of students also includes students’ ages. Describe a method for selecting a stratified random sample that is based on dividing the population into four subgroups based on student age.

(Answer: Suppose you think students range in age from 18 to 50. Create a list of all students ages 18 to 19, a second for those ages 20 to 21, a third for those ages 22 to 30, and a fourth for those ages 31 to 50. Then take a random sample of 75 students from each group. The age ranges are not critical but are useful to attempt to get roughly equally sized groups.)

(d) Do you think it is better to use two subgroups based on sex to select a stratified random sample than to use the subgroups described in Question 1c? Explain your reasoning?

(Answer: Students can probably build an argument either way, as long as the argument focuses on which variable [age or gender] is more likely related to how much more students would be willing to pay for an XG300. For example, males may be more willing to pay more for a car with a “technical” name than females.)

(2) The authors of the paper, “Illicit Use of Psychostimulants among College Students” (Psychology, Health & Medicine, 2002) surveyed college students about their use of legal and illegal stimulants. The sample of students surveyed consisted of students enrolled in a psychology class at a small, competitive college in the United States.

(a) Was this sample a simple random sample, stratified sample, systematic sample, or convenience sample? Explain your reasoning.

(Answer: This is a convenience sample because the sample was taken from just one class of students and was not a random sample from the population of all college students.)

(b) Can conclusions based on data from this survey be generalized to all U.S. college students? Why or why not?

(Answer: No. You cannot assume that this class of students is representative of all college students because this was a convenience sample. The use of stimulants may be more or less common at this particular university [especially given that it is more of an elite institution]. Students in a psychology class may not be representative of all students even at this school.)
Supporting Lesson 1.2.3: Other Sampling Strategies

Even though simple random sampling is a way to select a sample that is objective and avoids selection bias, it is not always easy to select a simple random sample and sometimes it can be impractical. In such situations, you want to consider other sampling strategies that produce a sample that is representative of the population. In this lesson, *systematic sampling* and *stratified random sampling* are considered.

**Task 1: Systematic Sampling**

Suppose you are interested in learning how far the people at a football game traveled to see the game. There is a sellout crowd of 20,000 people. For simplicity, assume that the seats in the stadium are numbered from 1 to 20,000.

(1) Describe a method for selecting a simple random sample of 200 people from the population of people attending the game.

(2) Is the plan you described in Question 1 easy to implement? What are some difficulties you would face in actually collecting data on how far the individuals in the sample had traveled to attend the game?

(3) Suppose someone proposes posting two people at each of the four entrances to the stadium. At each entrance, one person is in charge of counting the people passing through the entrance in order to identify every 100th person entering. The second person asks each person identified in this way how far he or she traveled to attend the game. This should result in a sample of about 200 people (assuming roughly 5,000 people go through each entrance and selecting every 100th person will identify about 50 people at each entrance). Do you think this strategy for selecting the sample is reasonable? Why do you think so?
Supporting Lesson 1.2.3: Other Sampling Strategies

The sampling strategy described in Question 3 is an example of systematic sampling. In systematic sampling, the individuals in the population are arranged in some order (e.g., in a list of the members of the population or in the previous example by the order they enter the stadium). Then the sample is selected in a “systematic” way by choosing (usually randomly) a number \( k \), such as 100 in the previous example, and then choosing every \( k \)th individual. It is also a good idea to start the process by selecting a random number between 1 and \( k \) to identify the first person to be included in the sample. For example, if the selected number was 38, the 38th person through the gate is included in the sample and then you select the 138th person, the 238th, and so on.

(4) Suppose an amusement park is open from 8 a.m. to 10 p.m. every day of the year. You are interested in learning about the average daily attendance at the park. One way to learn about this average is to choose a sample of days and record the attendance for those days. Is it a good idea to select a systematic sample of days by recording attendance every 14th day for one year? Why or why not?

In general, systematic sampling works well when there are no repeating patterns in the population list or ordered arrangement of the population that might be related to the variable you are interested in (e.g., park attendance).
Supporting Lesson 1.2.3: Other Sampling Strategies

Task 2: Exploring Sampling Variability

Consider a small population consisting of 20 students. Each student was asked the number of units he or she was enrolled in this semester, resulting in the following data:

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

(5) What is the average number of units for this population?

(6) If you select a sample of two students from this population at random, how would the value of the sample average compare to the population average? If you select a different sample of size 2, would you get the same value for the sample average?

(7) What aspect of the population provides an explanation for your answers in Question 6?
Supporting Lesson 1.2.3: Other Sampling Strategies

Now consider a different population of 20 students, with the following values for number of units:

<table>
<thead>
<tr>
<th>14</th>
<th>16</th>
<th>18</th>
<th>16</th>
<th>12</th>
<th>16</th>
<th>13</th>
<th>17</th>
<th>13</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>18</td>
<td>17</td>
<td>15</td>
<td>18</td>
<td>16</td>
<td>16</td>
<td>12</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

(8) What is the average number of units for this population?

(9) If you select a sample of two students from this population at random, would the value of the sample average be equal to the population average? If you select a different sample of size 2, would you get the same value for the sample average?

(10) Are your answers to Questions 6 and 9 different? If so, why are they different?
Supporting Lesson 1.2.3: Other Sampling Strategies

Task 3: Stratified Random Sampling
Consider the following two populations:

Population 1: All students who received an associate’s degree from your college in the past academic year.

Population 2: All students who received an associate’s degree from your college in the past academic year and graduated with a grade point average (GPA) of 3.5 or higher.

(11) If you are interested in the GPAs of students in the population, which population would have more variability? In other words, for which population would GPA tend to differ more from student to student? Explain why you chose this population.

(12) Suppose you take a random sample of size 50 from each of these two populations and compute the sample averages. Which population do you think the sample average is closer to the actual population average? Why do you think this? (Hint: Think about your answers in Task 2.)

When there is not much variability in a population, even a small sample can yield a fairly accurate estimate of the population average. For example, because the GPAs of students who graduate with a GPA of 3.5 or higher are quite similar, even a sample of 5 or 10 students from this group would tend to produce an average GPA that is a pretty accurate estimate of the population average. On the other hand, if the individuals in a population tend to be more variable, it is not as easy to produce an accurate estimate. For example, for the population of all students graduating with an associate’s degree (a much more diverse group with respect to GPA than the population of students graduating with a GPA of 3.5 or higher), a larger sample is needed to produce an accurate estimate of the population average.
Supporting Lesson 1.2.3: Other Sampling Strategies

**Stratified random sampling** is a sampling method that takes advantage of the fact that you can get good information from small samples if you are sampling from a group that does not have much variability. In stratified random sampling, the population of interest is divided up into nonoverlapping subgroups, and then a simple random sample is taken from each subgroup. The results from the subgroups are then pooled together to produce one estimate for the overall population. If the subgroups are less variable than the population as a whole, then a stratified sample can produce more accurate estimates of population characteristics than a simple random sample of the same size.

(13) Suppose you are interested in estimating the average amount of money spent on textbooks this semester by students at your college. You plan to use stratified random sampling to select a sample from the population of students at the college. Which of the following is a better way to create the subgroups? Explain your reasoning.

**Option A:** Use subgroups based on field of study by dividing students into three groups: humanities, liberal arts, and science and mathematics.

**Option B:** Use subgroups based on the first letter of last name by dividing students into three groups—A to H, I to P, and Q to Z.
Supporting Lesson 1.2.3: Other Sampling Strategies

Task 4: One Method to Avoid!

When evaluating a sampling plan, it is important to think carefully about whether the plan is likely to produce a sample that is representative of the population. One strategy that is almost never a good idea is to use a convenience sample. This method does not use random selection and involves using an easily available or “convenient” group to form a sample. Data from a convenience sample should not be used to generalize to any larger population.

One common form of convenience sampling is to rely on volunteers who respond to an advertisement or who must call in to a telephone number or log on to the Internet to participate in a survey.

(14) Science textbooks are among the most expensive books sold in the campus bookstore. Suppose you are interested in selecting a sample of students at your college to estimate the average amount of money spent on textbooks in a particular semester. Explain why it would not be a good idea to use the 100 students enrolled in a Chemistry course as the sample.

(15) Suppose you are interested in estimating the average distance from the primary residence to campus for students at your college. Because students who are enrolled in an online course communicate with the college by email, it would be easy to contact these students by email to ask them how far away from campus they live. Is this a good way of selecting the sample? Explain your reasoning.
Homework

(1) The article, “I’d Like to Buy a Vowel, Drivers Say” (USA Today, August 7, 2001) speculates that young people prefer automobile names consisting of just numbers and/or letters that do not form a word (such as Hyundai’s XG300, Mazda’s 626, and BMW’s 325i). The article says that Hyundai had planned to identify the car that was eventually marketed as the XG300 with the name Concerto until it determined that consumers thought XG300 sounded more “technical” and deserving of a higher price. How much more would students at your college be willing to pay for an XG300 than a comparable car called Concerto?

Suppose a list of all the students at your college is available. For the purpose of this exercise, assume that there are 15,000 students at the college.

(a) Describe a method for selecting a simple random sample of 300 students from this population.

(b) Describe a method for selecting a systematic sample of about 300 students from this population.

(c) Suppose the list of students also includes students’ ages. Describe a method for selecting a stratified random sample that is based on dividing the population into four subgroups based on student age.

(d) Do you think it is better to use two subgroups based on sex to select a stratified random sample than to use the subgroups described in Question 1c? Explain your reasoning?
Supporting Lesson 1.2.3: Other Sampling Strategies

(2) The authors of the paper, “Illicit Use of Psychostimulants among College Students” \((Psychology, Health & Medicine, 2002)\) surveyed college students about their use of legal and illegal stimulants. The sample of students surveyed consisted of students enrolled in a psychology class at a small, competitive college in the United States.

(a) Was this sample a simple random sample, stratified sample, systematic sample, or convenience sample? Explain your reasoning.

(b) Can conclusions based on data from this survey be generalized to all U.S. college students? Why or why not?
Supporting Lesson 1.2.4: Sources of Bias in Sampling

Estimated number of 50-minute class sessions: 1

Materials Required

There are two versions of the survey questions for Task 1. Distribute the Student Handout containing Version 1 of the questions to half of the class and the Student Handout with Version 2 to the remainder of the class.

Learning Goals

Students will begin to understand that bias can also be introduced by characteristics of the study design other than the way in which the sample is selected.

Students will begin to be able to

- explain how the wording of survey questions has the potential to introduce bias.
- give an example of a situation where a sensitive question or the desire to provide a socially acceptable response may introduce bias.

Introduction [Student Handout]

Bias in sampling occurs when there is a tendency for samples to differ from the population in some systematic way. Recall from Lesson 1.2.2 that one type of bias is selection bias, which is a tendency for samples to differ from the population because some individuals in the population are more likely to be selected for the sample. You saw that one way to avoid selection bias is to select a simple random sample. In this lesson, you will see that nonsampling bias can also be introduced by the way in which data are obtained from the selected sample.

Task 1: It Can Make a Difference... [Student Handout, estimated time: 20 minutes]

(1) Following is a brief survey consisting of two questions. Circle your answers to each question.

(Note: In this task, students are asked two questions. There are actually two versions of the survey questions [see the next page], but do not tell students this until after they have answered the questions.)
Supporting Lesson 1.2.4: Sources of Bias in Sampling

### Survey Questions for Task 1 [Version 1]

A. Do you think that your college should forbid public speeches that encourage illegal behavior?

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

B. Suppose a small village of 750 people is facing an outbreak of a rare contagious disease. The disease is fatal if untreated. Two medications (described below) that treat the disease are available. You can only choose one medication. Which medication would you select for the village?

<table>
<thead>
<tr>
<th>Medication 1</th>
<th>Medication 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication 1 always works, but there is only enough to treat 500 people. If Medication 1 is chosen, 250 people will die.</td>
<td></td>
</tr>
<tr>
<td>Medication 2 is unpredictable. If Medication 2 is chosen, there is a 2/3 chance that no one will die and a 1/3 chance that everyone will die.</td>
<td></td>
</tr>
</tbody>
</table>

### Survey Questions for Task 1 [Version 2]

1. Do you think that your college should allow public speeches that encourage illegal behavior?

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

2. Suppose a small village of 750 people is facing an outbreak of a rare contagious disease. The disease is fatal if untreated. Two medications (described below) that treat the disease are available. You can only choose one medication. Which medication would you select for the village?

<table>
<thead>
<tr>
<th>Medication 1</th>
<th>Medication 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication 1 always works, but there is only enough to treat 500 people. If Medication 1 is chosen, 500 people will be saved.</td>
<td></td>
</tr>
<tr>
<td>Medication 2 is unpredictable. If Medication 2 is chosen, there is a 1/3 chance that no one will be saved and a 2/3 chance that everyone will be saved.</td>
<td></td>
</tr>
</tbody>
</table>
(Note: After students have answered the survey questions, ask a student to read Question A out loud to the class. At this point, students will see that there is a slight difference in wording. For students that received the version that uses the word forbid, determine how many answered this question and what fraction of them responded yes to the question. Then for students that received the version that used the word allow, determine how many answered this question and what fraction responded no to this question. When questions like these were asked about speeches against democracy in 1941 [during World War II], 45% said that they would forbid speeches against democracy, but 75% said that they would not allow speeches against democracy. It appears that people reacted negatively to the word forbid.

Repeat a similar process with the second question. Here some students received a question where the options were worded in terms of how many people would die. The second version of this question describes the same outcome, but is worded in terms of how many people would survive. In answering questions similar to these, people seem more likely to choose the certain outcome when the choices are worded in terms of saving lives but the uncertain outcome when the choices are worded in terms of deaths.)

The type of bias explored in the question wording activity is called response bias, which occurs when questions are worded in a way that might distort or influence the response.

Keep in mind the following when evaluating survey questions or writing your own questions:

- Questions should be easy to understand. The sentence structure should be simple and the wording should be at an appropriate level for the given population of interest.
- Avoid emotionally charged words (e.g., forbid) and include background information if it is needed for people to make an informed response, but do not include information that is clearly one-sided or could influence the response.
- Questions should not make the responder feel threatened or embarrassed. Some people are hesitant to express what they see as an unpopular opinion, especially if they think that the person conducting the survey would disagree.

(2) Now consider the following question that was part of a Gallup survey sponsored by the American Paper Institute (Wall Street Journal; May 17, 1994):

It is estimated that disposable diapers account for less than 2 percent of the trash in today’s landfills. In contrast, beverage containers, third-class mail and yard waste are estimated to account for 21% of trash in landfills. Given this, in your opinion, would it be fair to tax or ban disposable diapers?

Do you think that the wording of this question may have influenced the way people responded to the question? If so, do you think that this was intentional?

(Answer: This question will probably lead people to say taxing or banning disposable diapers is not fair as they have been cast as not being the larger problem. Because this survey was sponsored by the American Paper Institute, it may have a vested interest in people buying disposable diapers, which are made from paper products.)
Supporting Lesson 1.2.4: Sources of Bias in Sampling

(3) The wording of the survey question in Question 2 may have influenced the way people responded. Suppose each person in a random sample of 200 students from your college was asked this survey question and that 80 of them responded that it would be fair to tax or ban disposable diapers. The proportion in the sample that responded this way is then \( \frac{80}{200} = 0.40 \). Equivalently, 40% of the sample responded that it would be fair. Do you think that 40% is an overestimate or underestimate of the percentage of all students at your college who think that a tax or ban of disposable diapers is fair? Why do you think so?

(Answer: Because the survey question’s wording is leading respondents to say that they do not think a ban or tax is fair, the sample percentage is probably an underestimate of the actual percentage of students who think it is fair.)

(4) Do you think selecting a larger random sample—for example, 500 students—would fix the problem identified in Question 3? Why do you think so?

(Answer: No. A larger sample size cannot correct the leading nature of the survey question’s wording.)

(5) Working with a partner, write a better survey question to replace the survey question in Question 2.

(Answer: Many revisions are possible, and it is good to discuss them as a class. In particular, students should try to come up with questions that address both pros and cons for the tax or do not provide background information beyond definitions at all.)

Wrap-Up

Make the point that survey question wording can make a difference in how people respond. Care needs to be taken in writing questions, and it is always a good idea to pretest survey questions by asking people to “think out loud” as they are answering the questions.

Task 2: Nonresponse Bias [Student Handout, estimated time: 15 minutes]

*Nonresponse bias* occurs when it is not possible to collect data from every individual in the sample. For example, you select a random sample of students at your college, but some refuse to participate or it is impossible to contact some of the selected students. If the people who respond differ in some systematic way from those who do not respond, even if the original sample was selected at random, data obtained only from responders may not be representative of the population.

The nonresponse rate (the proportion of the originally selected sample that do not respond) for surveys and opinion polls can differ from study to study depending on the survey topic, the length of the survey, and the way in which the survey is conducted.

Discuss the following questions with a partner.

(6) Do you think the nonresponse rate would be higher for a 20-question survey or a 50-question survey? Why do you think so?
Supporting Lesson 1.2.4: Sources of Bias in Sampling

(Answer: You expect the nonresponse rate to be higher for a longer survey because respondents will tire, become less interested in participating, or simply run out of time.)

(7) Do you think the nonresponse rate for a survey of college students would be higher for a survey about student loans or for a survey about saving for retirement? Why do you think so?

(Answer: The nonresponse rate will probably be higher for a survey about saving for retirement because the topic is not as relevant and interesting to college students.)

(8) Suppose you wanted to survey students at your college to learn about how they use various student services (e.g., health center, library, financial aid office, campus recreation facilities). The college has provided a list of all students, and you have selected a random sample of 250 students.

Consider the survey administration methods below. Arrange these methods in order from the one that you think would have the fewest nonresponders to the one that you think would have the largest number of nonresponders. Provide a brief justification for your ordering.

Method 1: Mail a survey to each student in the sample, using the student’s permanent home address. Students are asked to complete the survey and return it by mail.

Method 2: Email each student in the sample, asking the student to click on a link that takes him or her to the survey. Students then complete the survey online.

Method 3: Attempt to contact each student in the sample by telephone and ask him or her to answer the survey questions over the phone. If the student is not at home at the time of the call, a second attempt to contact the student is made on a different day and time.

Method 4: Go to the location listed as the current address of each student in the sample and attempt to interview the student in person to obtain answers to the survey questions. If the student is not at home or refuses to participate, they are counted as a nonresponder.

(Answers: Fewest nonresponders—Method 3; second fewest nonresponders—Method 4; second most nonresponders—Method 2; most nonresponders—Method 1.

Methods 3 and 4 are similar, but Method 3 is the only one that mentions a follow-up. Many polling organizations require at least three follow-ups before a person is counted as a nonresponder. Students may also argue for Method 4 because in-person interviews should be more effective than phone interviews. Methods 1 and 2 are highly susceptible to high nonresponse rates because people are notorious for ignoring requests for responses by mail. They may be more likely to respond to electronic mail than slower mail methods that require postage.

Students may order these differently, which is okay as long as they can explain their reasoning about how the method affects the response rate.)

Wrap-Up [Instructor-led discussion]

The following are main points to include:

- There are reasons a set of responses may not be representative of the larger population even if a random sample was selected.
Supporting Lesson 1.2.4: Sources of Bias in Sampling

- There are strategies for lowering nonresponse rates such as shorter, easier-to-understand questionnaires; use of balanced questions without emotionally laden wordings; and follow-ups for subjects that cannot initially be contacted.
- It is important to consider how a survey is administered and the exact wording of the questions used.

Task 3: Evaluating a Survey [Student Handout, estimated time: 10 minutes]

Look at the following study description from Introduction to Statistics and Data Analysis (Cengage Learning, 2012):

The article “What People Buy from Fast Food Restaurants: Caloric Content and Menu Item Selection” (Obesity [2009]: 1369 – 1374) reported that the average number of calories consumed at lunch in New York City fast food restaurants was 827. The researchers selected 267 of the 1625 fast food locations in New York City at random. The paper states that at each of these locations “adult customers were approached as they entered the restaurant and asked to provide their food receipt when exiting and to complete a brief survey.”

(9) Do you think that approaching customers as they entered the restaurant and before they ordered might have influenced what they ordered? Is this a problem? Why do you think so?

(Answer: If the participants knew the survey was going to be about how healthy their food choices were, they may be more likely to order healthier foods than they had originally planned. This means the responses received will not be representative of normal consumer behavior.)

(10) Is the potential bias identified in Question 9 response bias or nonresponse bias?

(Answer: Response bias)

(11) Some people who were approached chose not to participate in the study. Is this a problem? Why do you think so?

(Answer: This could be a problem if those who chose not to respond differ systematically in some way from those who participate. For example, those with less healthy habits may be less willing to have someone investigate their eating habits.)

(12) Is the potential bias identified Question 11 response bias or nonresponse bias?

(Answer: Nonresponse bias)

Wrap-Up [Instructor-led discussion]

An important point to make here is that bias in surveys can be introduced by how the sample is selected (sampling bias) or by how the data are collected once the sample is selected (nonsampling bias). A common misconception is that increasing the sample size is a way to deal with bias. However, taking a larger sample will not help if the way the sample is selected or the
Supporting Lesson 1.2.4: Sources of Bias in Sampling

The way in which information is obtained from those in the sample is flawed. Remind students that they do not usually know for sure whether bias is present, but by taking appropriate precautions, they have a stronger argument that bias will not be present.

Homework [Student Handout]

(1) Researchers were interested in learning whether people with higher levels of education use the Internet in different ways than those who do not have as much formal education. To answer this question, they used data from a national telephone survey. Approximately 1,300 households were selected for the survey, and 270 of them completed the interview. What type of bias should the researchers be concerned about and why?

(Answer: There is a high nonresponse rate here [1,030/1,300 = 79% of those selected did not complete the interview!]. The researchers should be concerned that those who responded differ on this issue from those who did not [nonresponse bias].)

(2) The report, “Undergraduate Students and Credit Cards in 2004: An Analysis of Usage Rates and Trends” (Nellie Mae, May 2005) describes the results of a survey of college students. Based on the data collected, the report concluded that about 21% of undergraduates with credit cards pay them off each month and that the average outstanding balance on undergraduates’ credit cards is about $2,169. The survey data was from an online survey that was emailed to 1,260 students. Responses were received from 132 of these students. What concerns do you have about generalizing these conclusions to the population of all undergraduate students?

(Answer: There is a high nonresponse rate and probably nonresponse bias. The students responded to the online survey may differ in some systematic ways from those choosing not to respond. For example, students who do not regularly pay off their cards or who have a high balance may be embarrassed by this and may choose not to participate in the survey.)

---

Supporting Lesson 1.2.4: Sources of Bias in Sampling

*Bias* in sampling occurs when there is a tendency for samples to differ from the population in some systematic way. Recall from Lesson 1.2.2 that one type of bias is *selection bias*, which is a tendency for samples to differ from the population because some individuals in the population are more likely to be selected for the sample. You saw that one way to avoid selection bias is to select a simple random sample. In this lesson, you will see that nonsampling bias can also be introduced by the way in which data are obtained from the selected sample.

**Task 1: It Can Make a Difference…**

(1) Following is a brief survey consisting of two questions. Circle your answers to each question.

<table>
<thead>
<tr>
<th>Survey Questions for Task 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Do you think that your college should forbid public speeches that encourage illegal behavior?</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>B. Suppose a small village of 750 people is facing an outbreak of a rare contagious disease. The disease is fatal if untreated. Two medications (described below) that treat the disease are available. You can only choose one medication. Which medication would you select for the village?</td>
</tr>
<tr>
<td>Medication 1</td>
</tr>
<tr>
<td>Medication 1 always works, but there is only enough to treat 500 people. If Medication 1 is chosen, 250 people will die.</td>
</tr>
<tr>
<td>Medication 2 is unpredictable. If Medication 2 is chosen, there is a 2/3 chance that no one will die and a 1/3 chance that everyone will die.</td>
</tr>
</tbody>
</table>
Supporting Lesson 1.2.4: Sources of Bias in Sampling

The type of bias explored in the question wording activity is called response bias, which occurs when questions are worded in a way that might distort or influence the response.

Keep in mind the following when evaluating survey questions or writing your own questions:

- Questions should be easy to understand. The sentence structure should be simple and the wording should be at an appropriate level for the given population of interest.
- Avoid emotionally charged words (e.g., forbid) and include background information if it is needed for people to make an informed response, but do not include information that is clearly one-sided or could influence the response.
- Questions should not make the responder feel threatened or embarrassed. Some people are hesitant to express what they see as an unpopular opinion, especially if they think that the person conducting the survey would disagree.

(2) Now consider the following question that was part of a Gallup survey sponsored by the American Paper Institute (Wall Street Journal; May 17, 1994):

It is estimated that disposable diapers account for less than 2 percent of the trash in today’s landfills. In contrast, beverage containers, third-class mail and yard waste are estimated to account for 21% of trash in landfills. Given this, in your opinion, would it be fair to tax or ban disposable diapers?

Do you think that the wording of this question may have influenced the way people responded to the question? If so, do you think that this was intentional?

(3) The wording of the survey question in Question 2 may have influenced the way people responded. Suppose each person in a random sample of 200 students from your college was asked this survey question and that 80 of them responded that it would be fair to tax or ban disposable diapers. The proportion in the sample that responded this way is then $\frac{80}{200} = 0.40$. Equivalently, 40% of the sample responded that it would be fair. Do you think that 40% is an overestimate or underestimate of the percentage of all students at your college who think that a tax or ban of disposable diapers is fair? Why do you think so?
(4) Do you think selecting a larger random sample—for example, 500 students—would fix the problem identified in Question 3? Why do you think so?

(5) Working with a partner, write a better survey question to replace the survey question in Question 2.
Supporting Lesson 1.2.4: Sources of Bias in Sampling

Task 2: Nonresponse Bias

Nonresponse bias occurs when it is not possible to collect data from every individual in the sample. For example, you select a random sample of students at your college, but some refuse to participate or it is impossible to contact some of the selected students. If the people who respond differ in some systematic way from those who do not respond, even if the original sample was selected at random, data obtained only from responders may not be representative of the population.

The nonresponse rate (the proportion of the originally selected sample that do not respond) for surveys and opinion polls can differ from study to study depending on the survey topic, the length of the survey, and the way in which the survey is conducted.

Discuss the following questions with a partner.

(6) Do you think the nonresponse rate would be higher for a 20-question survey or a 50-question survey? Why do you think so?

(7) Do you think the nonresponse rate for a survey of college students would be higher for a survey about student loans or for a survey about saving for retirement? Why do you think so?
Suppose you wanted to survey students at your college to learn about how they use various student services (e.g., health center, library, financial aid office, campus recreation facilities). The college has provided a list of all students, and you have selected a random sample of 250 students.

Consider the survey administration methods below. Arrange these methods in order from the one that you think would have the fewest nonresponders to the one that you think would have the largest number of nonresponders. Provide a brief justification for your ordering.

**Method 1:** Mail a survey to each student in the sample, using the student’s permanent home address. Students are asked to complete the survey and return it by mail.

**Method 2:** Email each student in the sample, asking the student to click on a link that takes him or her to the survey. Students then complete the survey online.

**Method 3:** Attempt to contact each student in the sample by telephone and ask him or her to answer the survey questions over the phone. If the student is not at home at the time of the call, a second attempt to contact the student is made on a different day and time.

**Method 4:** Go to the location listed as the current address of each student in the sample and attempt to interview the student in person to obtain answers to the survey questions. If the student is not at home or refuses to participate, they are counted as a nonresponder.
Supporting Lesson 1.2.4: Sources of Bias in Sampling

Task 3: Evaluating a Survey

Look at the following study description from Introduction to Statistics and Data Analysis (Cengage Learning, 2012):

The article “What People Buy from Fast Food Restaurants: Caloric Content and Menu Item Selection” (Obesity [2009]: 1369 – 1374) reported that the average number of calories consumed at lunch in New York City fast food restaurants was 827. The researchers selected 267 of the 1625 fast food locations in New York City at random. The paper states that at each of these locations “adult customers were approached as they entered the restaurant and asked to provide their food receipt when exiting and to complete a brief survey.”

(9) Do you think that approaching customers as they entered the restaurant and before they ordered might have influenced what they ordered? Is this a problem? Why do you think so?

(10) Is the potential bias identified in Question 9 response bias or nonresponse bias?

(11) Some people who were approached chose not to participate in the study. Is this a problem? Why do you think so?

(12) Is the potential bias identified Question 11 response bias or nonresponse bias?
Supporting Lesson 1.2.4: Sources of Bias in Sampling

Homework

(1) Researchers were interested in learning whether people with higher levels of education use the Internet in different ways than those who do not have as much formal education.\(^1\) To answer this question, they used data from a national telephone survey. Approximately 1,300 households were selected for the survey, and 270 of them completed the interview. What type of bias should the researchers be concerned about and why?

(2) The report, “Undergraduate Students and Credit Cards in 2004: An Analysis of Usage Rates and Trends” (Nellie Mae, May 2005) describes the results of a survey of college students. Based on the data collected, the report concluded that about 21% of undergraduates with credit cards pay them off each month and that the average outstanding balance on undergraduates’ credit cards is about $2,169. The survey data was from an online survey that was emailed to 1,260 students. Responses were received from 132 of these students. What concerns do you have about generalizing these conclusions to the population of all undergraduate students?

Supporting Lesson 1.2.4: Sources of Bias in Sampling

Bias in sampling occurs when there is a tendency for samples to differ from the population in some systematic way. Recall from Lesson 1.2.2 that one type of bias is selection bias, which is a tendency for samples to differ from the population because some individuals in the population are more likely to be selected for the sample. You saw that one way to avoid selection bias is to select a simple random sample. In this lesson, you will see that nonsampling bias can also be introduced by the way in which data are obtained from the selected sample.

Task 1: It Can Make a Difference...

(1) Following is a brief survey consisting of two questions. Circle your answers to each question.

**Survey Questions for Task 1**

1. Do you think that your college should allow public speeches that encourage illegal behavior?

   Yes  No

2. Suppose a small village of 750 people is facing an outbreak of a rare contagious disease. The disease is fatal if untreated. Two medications (described below) that treat the disease are available. You can only choose one medication. Which medication would you select for the village?

   Medication 1	 Medication 2

   Medication 1 always works, but there is only enough to treat 500 people. If Medication 1 is chosen, 500 people will be saved.

   Medication 2 is unpredictable. If Medication 2 is chosen, there is a 1/3 chance that no one will be saved and a 2/3 chance that everyone will be saved.
The type of bias explored in the question wording activity is called response bias, which occurs when questions are worded in a way that might distort or influence the response.

Keep in mind the following when evaluating survey questions or writing your own questions:

- Questions should be easy to understand. The sentence structure should be simple and the wording should be at an appropriate level for the given population of interest.
- Avoid emotionally charged words (e.g., forbid) and include background information if it is needed for people to make an informed response, but do not include information that is clearly one-sided or could influence the response.
- Questions should not make the responder feel threatened or embarrassed. Some people are hesitant to express what they see as an unpopular opinion, especially if they think that the person conducting the survey would disagree.

(2) Now consider the following question that was part of a Gallup survey sponsored by the American Paper Institute (Wall Street Journal; May 17, 1994):

> It is estimated that disposable diapers account for less than 2 percent of the trash in today’s landfills. In contrast, beverage containers, third-class mail and yard waste are estimated to account for 21% of trash in landfills. Given this, in your opinion, would it be fair to tax or ban disposable diapers?

Do you think that the wording of this question may have influenced the way people responded to the question? If so, do you think that this was intentional?

(3) The wording of the survey question in Question 2 may have influenced the way people responded. Suppose each person in a random sample of 200 students from your college was asked this survey question and that 80 of them responded that it would be fair to tax or ban disposable diapers. The proportion in the sample that responded this way is then \[
\frac{80}{200} = 0.40.
\]
Equivalently, 40% of the sample responded that it would be fair. Do you think that 40% is an overestimate or underestimate of the percentage of all students at your college who think that a tax or ban of disposable diapers is fair? Why do you think so?
Supporting Lesson 1.2.4: Sources of Bias in Sampling

(4) Do you think selecting a larger random sample—for example, 500 students—would fix the problem identified in Question 3? Why do you think so?

(5) Working with a partner, write a better survey question to replace the survey question in Question 2.
Supporting Lesson 1.2.4: Sources of Bias in Sampling

Task 2: Nonresponse Bias

Nonresponse bias occurs when it is not possible to collect data from every individual in the sample. For example, you select a random sample of students at your college, but some refuse to participate or it is impossible to contact some of the selected students. If the people who respond differ in some systematic way from those who do not respond, even if the original sample was selected at random, data obtained only from responders may not be representative of the population.

The nonresponse rate (the proportion of the originally selected sample that do not respond) for surveys and opinion polls can differ from study to study depending on the survey topic, the length of the survey, and the way in which the survey is conducted.

Discuss the following questions with a partner.

(6) Do you think the nonresponse rate would be higher for a 20-question survey or a 50-question survey? Why do you think so?

(7) Do you think the nonresponse rate for a survey of college students would be higher for a survey about student loans or for a survey about saving for retirement? Why do you think so?
Supporting Lesson 1.2.4: Sources of Bias in Sampling

(8) Suppose you wanted to survey students at your college to learn about how they use various student services (e.g., health center, library, financial aid office, campus recreation facilities). The college has provided a list of all students, and you have selected a random sample of 250 students.

Consider the survey administration methods below. Arrange these methods in order from the one that you think would have the fewest nonresponders to the one that you think would have the largest number of nonresponders. Provide a brief justification for your ordering.

**Method 1:** Mail a survey to each student in the sample, using the student’s permanent home address. Students are asked to complete the survey and return it by mail.

**Method 2:** Email each student in the sample, asking the student to click on a link that takes him or her to the survey. Students then complete the survey online.

**Method 3:** Attempt to contact each student in the sample by telephone and ask him or her to answer the survey questions over the phone. If the student is not at home at the time of the call, a second attempt to contact the student is made on a different day and time.

**Method 4:** Go to the location listed as the current address of each student in the sample and attempt to interview the student in person to obtain answers to the survey questions. If the student is not at home or refuses to participate, they are counted as a nonresponder.
Supporting Lesson 1.2.4: Sources of Bias in Sampling

Task 3: Evaluating a Survey

Look at the following study description from *Introduction to Statistics and Data Analysis* (Cengage Learning, 2012):

The article “What People Buy from Fast Food Restaurants: Caloric Content and Menu Item Selection” (*Obesity* [2009]: 1369 – 1374) reported that the average number of calories consumed at lunch in New York City fast food restaurants was 827. The researchers selected 267 of the 1625 fast food locations in New York City at random. The paper states that at each of these locations “adult customers were approached as they entered the restaurant and asked to provide their food receipt when exiting and to complete a brief survey.”

(9) Do you think that approaching customers as they entered the restaurant and before they ordered might have influenced what they ordered? Is this a problem? Why do you think so?

(10) Is the potential bias identified in Question 9 response bias or nonresponse bias?

(11) Some people who were approached chose not to participate in the study. Is this a problem? Why do you think so?

(12) Is the potential bias identified Question 11 response bias or nonresponse bias?
Supporting Lesson 1.2.4: Sources of Bias in Sampling

Homework

(1) Researchers were interested in learning whether people with higher levels of education use the Internet in different ways than those who do not have as much formal education. To answer this question, they used data from a national telephone survey. Approximately 1,300 households were selected for the survey, and 270 of them completed the interview. What type of bias should the researchers be concerned about and why?

(2) The report, “Undergraduate Students and Credit Cards in 2004: An Analysis of Usage Rates and Trends” (Nellie Mae, May 2005) describes the results of a survey of college students. Based on the data collected, the report concluded that about 21% of undergraduates with credit cards pay them off each month and that the average outstanding balance on undergraduates’ credit cards is about $2,169. The survey data was from an online survey that was emailed to 1,260 students. Responses were received from 132 of these students. What concerns do you have about generalizing these conclusions to the population of all undergraduate students?

---

Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

Estimated number of 50-minute class sessions: 2

Task 1 and A Bit of Terminology are covered in the first session and Tasks 2 and 3 and Lesson Wrap-Up in the second session. It may be possible to start on Task 2 in the first session.

Materials Required

For Task 3, each student or pair of students needs a coin, a marker, and a copy of the page with common scales for the dotplots. You also need masking tape for displaying the dotplots created by the different groups on the board.

Learning Goals

Students will begin to understand the purpose of random assignment in the design of an experiment.

Students will begin to be able to

- given the description of an experiment, identify the response variable, treatment variable, experimental conditions, and experimental units.
- given the description of an experiment, describe a process for randomly assigning experimental units to experimental conditions.
- explain the benefits of random assignment in the design of an experiment.
- given the description of an experiment, identify which of the following was part of the experimental design: random assignment and direct control.
- describe characteristics of a well-designed experiment.

Introduction [Student Handout]

Recall from Lesson 1.1.2 that an experiment is a statistical study that involves observing how a response variable behaves under different experimental conditions. If you see a difference in the responses between the experimental conditions, you would like to decide whether the observed difference is a result of the experimental conditions. To reach this conclusion, you need to rule out other possible explanations for any differences in response observed between the experimental conditions. To accomplish this, the experiment needs to be planned carefully.

Task 1: Planning an Experiment [Student Handout; approximately 35 minutes]

(Note: Have students work in small groups to complete Questions 1–5 of this task. Then bring the class together for a whole-class discussion that wraps up this task.)

In this task, consider an experiment to investigate the relationship between the color of a soda (soft drink) and how people rate the taste of the soda. You want to collect data that allows you to answer the question, “Does coloring a soda blue affect the way people rate the taste of the soda?”

PepsiCo might have liked to know the answer to this question before it introduced the unconventional and short-lived Pepsi Blue in 2002!

To investigate this question, plan to ask some people taste a clear soda and rate the taste on a scale of 0–10. Other people will be asked to taste the same soda, but with blue food coloring added. (Food coloring is tasteless, so it changes the color of a food or beverage without changing its flavor.) The people who taste the blue soda will also be asked to rate the taste on a scale of 0–10.
The idea is to compare the ratings from people tasting the clear soda with the ratings from people tasting the blue soda to see whether the ratings for the blue soda tend to be lower than, about the same as, or higher than the ratings for the clear soda.

(1) Suppose 50 students from your college have volunteered to participate in the experiment. Work with your group to come up with a plan for carrying out the experiment. Suppose you are telling others how to carry out the experiment according to your plan. Create a list of steps they could use to implement your plan. Keep in mind that the goal is to isolate the effect of color on the taste of soda.

When you think you have a good plan, answer Questions 2–5.

(Note: Give students about 10 minutes to brainstorm in groups. If a group stalls, prompt it with some issues brought up on the next page. Remind groups that they are to put together a list of instructions that others could use to carry out the experiment on their own. The groups will probably make different decisions. After the groups have brainstormed and written their plans, discuss together some of the issues, what different groups decided, and perhaps arrive at a consensus plan.

Keep track of the issues identified by the groups. Help them see which issues are more critical than others [e.g., the amount of lighting in the room: something that might not affect taste or is already constant between the two groups]. If a group wants to let everyone taste both sodas, that is a reasonable plan. Just make sure they randomize which soda is tasted first.

If students are not considering the following issues, you can discuss the questions together.)

(2) Did you create two experimental groups (a blue group and a clear group)? If so, how did you decide which participants were assigned to which group?

(Note: At this stage, students may only say “randomly.” Push them to explain what that means and how they could carry the random assignment out in detail [e.g., flipping a coin to decide]. Make sure they are focusing on the random assignment to groups and not how the 50 volunteers are selected to participate. Some students will also use the word randomly but not have a probability method in mind [e.g., randomly let the boys drink blue soda and the girls drink clear soda].)

(3) Did you consider the temperature of the sodas? Could a soda’s temperature affect people’s ratings of its taste? How could you make sure that temperature does not provide an alternative explanation if one color tends to receive a higher rating than the other color?

(Note: This question should prompt students to think about ensuring the temperature is the same for all soda samples. It would be problematic if the blue soda is allowed to get warm while the clear soda is being tasted, as that provides an explanation other than color for the blue soda receiving lower scores. The instructions can include a step to chill the sodas for a set amount of time and take them out of the refrigerator at the same time.)

(4) Did you consider the age and carbonation of the sodas? How so?
Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

(Note: The instructions can include purchasing the sodas at the same time and opening them at the same time.)

(5) Did you consider the amount of soda tasted?

(Note: The instructions can include measuring the amount of soda to make sure it is consistent. Groups can also decide whether subjects are allowed to take more than one sip of the soda as they decide on their rating.)

Wrap-Up
At the end of the discussion, remind students that the goal is to eliminate any alternative explanations for differences in taste ratings between the two groups. They have now dealt with two main strategies for this: creating similar groups and holding variables constant. Highlight these strategies as they appear in the list of issues written on the board. The next two tasks formalize these strategies in more detail.)

A Bit of Terminology [Student Handout, estimated time: 5–10 minutes]
(Note: This can be read by students individually or incorporated into a brief minilecture.)

Before you move on to Task 2, let’s take a look at some terminology used when describing experiments.

When planning an experiment, it is helpful to think of the research question as being in the form of:

What is the effect of ___1___ on ___2___?

For example, for the soda taste experiment described in Task 1, you could say:

What is the effect of color on taste rating?

The numbered parts of the question correspond to an important aspect of the plan for the experiment.

1—“the effect of ___”

This part of the question identifies the treatment variable for the experiment. The observed values of this variable that will be used in the experiment determine the experimental conditions that will be compared in the experiment. For the soda taste experiment, the treatment variable is color and the experimental conditions to be compared are clear and blue.

2—“on ___”

This part of the question identifies the response variable in the experiment. For the soda taste experiment, the response variable is taste rating.

You also need to identify who the participants in the experiment will be. Sometimes the participants are people, as in the soda taste experiment. However, participants are not always people.

For example, you could conduct an experiment to investigate the effect of a gasoline additive (the treatment variable) on fuel efficiency (the response variable). In this experiment, you might compare fuel efficiencies for two groups of cars, where one group of cars used gas without the additive and the other group of cars used gas with the additive. Here, the participants in the experiment are the cars. In general, participants in an experiment are referred to as experimental units.
Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

Task 2: Direct Control [Student Handout, estimated time: 10 minutes]

(Note: Have students read the introduction or incorporate it into minilecture.)

The goal of an experiment is to determine the effects of different experimental conditions on some response variable. To do this, you want to rule out other possible explanations for why the responses to the different experimental conditions might differ. There are two common strategies to help with this: direct control and random assignment.

The first strategy (direct control) is simple. If there are variables other than the treatment variable in the experiment that you think might affect the response, try to control those variables. For example, in the soda taste experiment, the temperature of the soda might make a difference in how people rate the taste. If the soda is served warm, its flavor might tend to be rated lower than if it is served cold. This could be a problem if the blue soda’s serving temperatures differed from the clear soda’s serving temperatures. The easy way to fix this is to make sure that all soda samples are kept chilled and are served at the same temperature. Then, if there is a difference in the taste ratings for the clear and blue groups, it cannot be explained by differences in temperature.

Consider the following description of an experiment to investigate whether the way people dry their hands after washing them has an effect on how clean their hands are:

An experiment was conducted to compare bacteria reduction for three different hand drying methods. In this experiment, subjects handled uncooked chicken for 45 seconds, then washed their hands with a single squirt of soap for 60 seconds, and then used one of three hand drying methods. After completely drying their hands, the bacteria count on their hands was measured.

(6) What is the treatment variable in this experiment? How many experimental conditions were there in this experiment?

(Answer: Drying method; three experimental conditions)

(7) What is the response variable in this experiment?

(Answer: Bacteria count)

(8) One variable that might affect the response is the length of time that people handled the raw chicken. The researchers controlled this by having everyone handle the chicken for the same amount of time—45 seconds. Two other variables were controlled in this experiment. What are they?

(Answer: The amount of soap [single squirt] and the amount of time washing hands [60 seconds])
Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

Wrap-Up [Instructor-led discussion]

First, make sure students are comfortable with the terminology. You may want to return to the soda experiment: the response variable is the taste rating, the treatment variable is the soda color, and the participants are the 50 student volunteers. Second, make sure students understand that in experiments, they can often control variables other than the treatment variable that may affect the response. Return to the instructions they provided for the soda experiment and circle those that correspond to applying direct control (e.g., the temperature of soda). Additional controls could be asking the students not to eat or brush their teeth immediately prior to participating in the experiment.

In addition, warn students that sometimes there are variables they can think of in advance that might affect the response but are not easily controlled. In the soda study, they can purchase the sodas on the same day, but do not really know exactly how long ago it was bottled. They also may not be able to control the temperature of the classroom that day. They also cannot control whether someone is a frequent soda drinker. These kinds of variables are hopefully addressed through random assignment, which is discussed in Task 3.

Task 3: Random Assignment—Creating Comparable Groups [Student Handout, estimated time: 40 minutes]

(Note: The first parts of this task [Questions 9–11] are probably best accomplished as a whole-group discussion. Have the students read the introductory text that includes the description of the experiment and then spend a few minutes making sure everyone understands the experiment description before beginning Question 9.)

An experiment was carried out to compare recovery time for two different types of surgery proposed to treat children with hernias.²

Method 1: laparoscopic repair (a surgery that uses only three small incisions)

Method 2: open repair (a surgery that involves making one large incision)

Eighty-nine children with hernias participated in the experiment.

To compare the two experimental conditions (surgery methods), you want to create two groups of children that you think are similar with respect to any variables that might affect the response variable (recovery time).

One variable that the researchers thought might affect recovery time was the age of the child. One way to deal with this would be to only use 10-year-old children in the study (direct control). However, it might be difficult to find enough 10-year-old children with hernias to serve as subjects for the study. So how do you create treatment groups that are similar with respect to age? One way to assign children at random to one of the two experimental conditions (Method 1 and Method 2) is to flip a coin for each child to determine which group they are assigned to. For example, if the coin lands H (heads up) the child is assigned to the Method 1 group, and if it lands T (tails up) the child is assigned to the Method 2 group.

Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

Does this really work? To investigate this, consider a somewhat smaller problem. Instead of 89 children, suppose there are 30 children with hernias who volunteered to participate in the experiment. The ages of these 30 children are given in the following table.

<table>
<thead>
<tr>
<th>Child</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
</tr>
</tbody>
</table>

(9) The random assignment was started by flipping a coin to decide whether Child 1 would be assigned to the Method 1 group or the Method 2 group. The toss resulted in a T, so Child 1 was assigned to Method 2, as indicated in the table. For each of the remaining children, flip a coin to determine which group the child should be assigned to and place the child in the appropriate group in the table. Continue until you have placed all of the children in one of the two groups. (Using this method for random assignment, you will probably not end up with exactly 15 children in each group, but that is okay.)
## Method 1 Group

<table>
<thead>
<tr>
<th>Child</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Method 2 Group

<table>
<thead>
<tr>
<th>Child</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12</td>
</tr>
</tbody>
</table>

(10) Using the blank dotplots at the end of this lesson, construct a dotplot of the ages of children assigned to the Method 1 group and a dotplot of the ages of the children assigned to the Method 2 group.

(11) Calculate the average age for the children assigned to the Method 1 group and the average age for the children assigned to the Method 2 group. Write these values on the dotplots you created in Question 10. Which group has the largest average age? Is it substantially larger than the average age for the other group, or are the two averages about the same?

(Answer: Because random assignment was used to create the groups, the two group averages are expected to be about the same.)
Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

(Note: Now have students tape the dotplots around the room and ask them to look at the different results.)

(12) Considering the dotplots and the two averages, do you think the random assignment was successful in creating groups that were similar with respect to the ages of the children in the groups?

(Note: With random assignment, the two groups are expected to be similar with respect to age. Students should notice that none of the pairs of dotplots had large differences between the treatment groups. About half the time the mean may be larger for Group 1, and about half the time the mean may be larger for Group 2.)

(13) Other variables that might affect recovery time are weight and fitness level. Do you think that the random assignment to experimental conditions would create groups that are similar on these two variables as well as age? Why do you think so?

(Note: There was nothing special in the random assignment about age. Children of different weights and fitness levels are equally likely to end up in either treatment group, and overall should balance out between the two groups. This should help convince students that random assignment should also work for creating groups that are similar on other variables as well, even those they may not know about in advance.)

(Note: In summarizing the exploration of the different dotplots that have been created, make the following key points:

- Not every group ends up with the same results. Not every group ends up with the same averages in the two groups. There is variability across the random assignments.
- Still, the age distributions in the two groups should look pretty similar. It is very unlikely that random assignments lead to one group having much higher ages than the other.)

Wrap-Up [Instructor-led discussion]

The following are key points to make:

- Random assignment tends to create comparable groups. It does not tend to favor one condition over the other in terms of age.
- This will be true for any variable that could potentially affect recovery time, such as severity of hernia.
- Therefore, random assignment allows students to rule out these other variables as possible alternative explanations for an observed difference between the experimental groups. As long as there are a reasonable number of experimental units (participants) in the experiment, random assignment tends to create groups that are similar on variables that they know may have an effect on the response and also on variables that they do not know about. (Students still need to worry about the possible unlucky random assignment; this is addressed in Lesson 1.4.1.)
Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

Lesson Wrap-Up [Instructor-led discussion]

The primary goal in designing an experiment is to eliminate other possible explanations for any differences you find in the response variable between the treatment groups. There are two primary methods for doing this: direct control and random assignment. In particular, random assignment takes care of the things that you cannot control or do not know about in advance. The key advantage is if students believe there are no other explanations, they can draw a cause-and-effect conclusion about the effect of the treatment variable on the response variable. Remind students that this is different than using random sampling to generalize results from a sample to a larger population.

Further Notes

The lesson introduction said that to be able to conclude the experimental conditions have an effect on the response variable, you need to do two things:

- Rule out chance as a possible explanation for the observed differences.
- Rule out other possible explanations for why the differences occurred.

Direct control and random assignment are two strategies that allow students to rule out other possible explanations for why the differences occurred.

Lesson 1.4.1 (Drawing Conclusions from Statistical Studies) looks at how to also rule out chance as a possible explanation for any observed differences between experimental groups.

If students can rule out alternate explanations and rule out chance as a plausible explanation, then they can conclude that the difference in response is due to the experimental conditions.

Homework [Student Handout]

(1) A study done by researchers at Kings College London found that infomania has a temporary, but significant, negative effect on intelligence quotient (IQ). In the experiment, the researchers divided volunteers into two groups. Each subject took an IQ test. One group had to check e-mail and respond to instant messages while taking the test, and the other group took the IQ test without any distractions. Researchers found that the average score on the IQ test for the distracted group was 10 points lower than the average for the group that was not distracted.  

(a) What is the response variable in this experiment?  
(Answer: IQ score)

(b) What is the treatment variable and what are the experimental conditions?  
(Answer: The treatment variable is whether they have to respond to messages, so there are two experimental conditions.)

(c) Explain why the researchers should use random assignment to put each volunteer in one of the experimental groups rather than letting the volunteers decide which group they wanted to be in.

---

(Answer: This tends to balance out any variables other than the treatment variable between the two groups.)

(2) Do ethnic group and gender influence the type of care that a heart patient receives? The following passage is from a 1999 USA Today article.\(^4\)

Previous research suggested blacks and women were less likely than whites and men to get cardiac catheterization or coronary bypass surgery for chest pain or a heart attack. Scientists blamed differences in illness severity, insurance coverage, patient preference, and health care access. The researchers eliminated those differences by videotaping actors—two black men, two black women, two white men, and two white women—describing chest pain from identical scripts. They wore identical gowns, used identical gestures, and were taped from the same position. Researchers asked 720 primary care doctors at meetings of the American College of Physicians or the American Academy of Family Physicians to watch a tape and recommend care. The doctors thought the study focused on clinical decision making.

The video a particular doctor watched was determined by the roll of a four-sided die.

(a) In the experiment described, the researchers chose to control a number of variables other than ethnic group and gender that might have influenced the recommended treatment. How did they do this?

(Answer: The actors wore the same gowns, followed the same scripts, and used the same hand gestures in describing the same pain.)

(b) Did the researchers use random assignment to create the experimental groups? If so, how did they do this?

(Answer: Random assignment was implemented by using a four-sided die to determine which videotape the physician viewed.)
Dotplots of Ages for the Two Experimental Groups (Task 3)

Average Age:

Average Age:
Recall from Lesson 1.1.2 that an experiment is a statistical study that involves observing how a response variable behaves under different experimental conditions. If you see a difference in the responses between the experimental conditions, you would like to decide whether the observed difference is a result of the experimental conditions. To reach this conclusion, you need to rule out other possible explanations for any differences in response observed between the experimental conditions. To accomplish this, the experiment needs to be planned carefully.

**Task 1: Planning an Experiment**

In this task, consider an experiment to investigate the relationship between the color of a soda (soft drink) and how people rate the taste of the soda. You want to collect data that allows you to answer the question, “Does coloring a soda blue affect the way people rate the taste of the soda?”

PepsiCo might have liked to know the answer to this question before it introduced the unconventional and short-lived Pepsi Blue in 2002!

To investigate this question, plan to ask some people taste a clear soda and rate the taste on a scale of 0–10. Other people will be asked to taste the same soda, but with blue food coloring added. (Food coloring is tasteless, so it changes the color of a food or beverage without changing its flavor.) The people who taste the blue soda will also be asked to rate the taste on a scale of 0–10.

The idea is to compare the ratings from people tasting the clear soda with the ratings from people tasting the blue soda to see whether the ratings for the blue soda tend to be lower than, about the same as, or higher than the ratings for the clear soda.

1. Suppose 50 students from your college have volunteered to participate in the experiment. Work with your group to come up with a plan for carrying out the experiment. Suppose you are telling others how to carry out the experiment according to your plan. Create a list of steps they could use to implement your plan. Keep in mind that the goal is to isolate the effect of color on the taste of soda.

    When you think you have a good plan, answer Questions 2–5.
Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

(2) Did you create two experimental groups (a blue group and a clear group)? If so, how did you decide which participants were assigned to which group?

(3) Did you consider the temperature of the sodas? Could a soda’s temperature affect people’s ratings of its taste? How could you make sure that temperature does not provide an alternative explanation if one color tends to receive a higher rating than the other color?

(4) Did you consider the age and carbonation of the sodas? How so?

(5) Did you consider the amount of soda tasted?
A Bit of Terminology

Before you move on to Task 2, let’s take a look at some terminology used when describing experiments.

When planning an experiment, it is helpful to think of the research question as being in the form of:

What is the effect of ____ 1 ____ on ____ 2 ____?

For example, for the soda taste experiment described in Task 1, you could say:

What is the effect of color on taste rating?

The numbered parts of the question correspond to an important aspect of the plan for the experiment.

1—“the effect of ___”

This part of the question identifies the treatment variable for the experiment. The observed values of this variable that will be used in the experiment determine the experimental conditions that will be compared in the experiment. For the soda taste experiment, the treatment variable is color and the experimental conditions to be compared are clear and blue.

2—“on ___”

This part of the question identifies the response variable in the experiment. For the soda taste experiment, the response variable is taste rating.

You also need to identify who the participants in the experiment will be. Sometimes the participants are people, as in the soda taste experiment. However, participants are not always people.

For example, you could conduct an experiment to investigate the effect of a gasoline additive (the treatment variable) on fuel efficiency (the response variable). In this experiment, you might compare fuel efficiencies for two groups of cars, where one group of cars used gas without the additive and the other group of cars used gas with the additive. Here, the participants in the experiment are the cars. In general, participants in an experiment are referred to as experimental units.
Task 2: Direct Control

The goal of an experiment is to determine the effects of different experimental conditions on some response variable. To do this, you want to rule out other possible explanations for why the responses to the different experimental conditions might differ. There are two common strategies to help with this: direct control and random assignment.

The first strategy (direct control) is simple. If there are variables other than the treatment variable in the experiment that you think might affect the response, try to control those variables. For example, in the soda taste experiment, the temperature of the soda might make a difference in how people rate the taste. If the soda is served warm, its flavor might tend to be rated lower than if it is served cold. This could be a problem if the blue soda’s serving temperatures differed from the clear soda’s serving temperatures. The easy way to fix this is to make sure that all soda samples are kept chilled and are served at the same temperature. Then, if there is a difference in the taste ratings for the clear and blue groups, it cannot be explained by differences in temperature.

Consider the following description of an experiment to investigate whether the way people dry their hands after washing them has an effect on how clean their hands are:

An experiment was conducted to compare bacteria reduction for three different hand drying methods. In this experiment, subjects handled uncooked chicken for 45 seconds, then washed their hands with a single squirt of soap for 60 seconds, and then used one of three hand drying methods. After completely drying their hands, the bacteria count on their hands was measured.\(^1\)

(6) What is the treatment variable in this experiment? How many experimental conditions were there in this experiment?

(7) What is the response variable in this experiment?

(8) One variable that might affect the response is the length of time that people handled the raw chicken. The researchers controlled this by having everyone handle the chicken for the same amount of time—45 seconds. Two other variables were controlled in this experiment. What are they?

---

\(^1\)Infectious Disease News. (September 2010).
Task 3: Random Assignment—Creating Comparable Groups

An experiment was carried out to compare recovery time for two different types of surgery proposed to treat children with hernias.²

Method 1: laparoscopic repair (a surgery that uses only three small incisions)
Method 2: open repair (a surgery that involves making one large incision)

Eighty-nine children with hernias participated in the experiment.

To compare the two experimental conditions (surgery methods), you want to create two groups of children that you think are similar with respect to any variables that might affect the response variable (recovery time).

One variable that the researchers thought might affect recovery time was the age of the child. One way to deal with this would be to only use 10-year-old children in the study (direct control). However, it might be difficult to find enough 10-year-old children with hernias to serve as subjects for the study. So how do you create treatment groups that are similar with respect to age? One way to assign children at random to one of the two experimental conditions (Method 1 and Method 2) is to flip a coin for each child to determine which group they are assigned to. For example, if the coin lands H (heads up) the child is assigned to the Method 1 group, and if it lands T (tails up) the child is assigned to the Method 2 group.

Does this really work? To investigate this, consider a somewhat smaller problem. Instead of 89 children, suppose there are 30 children with hernias who volunteered to participate in the experiment. The ages of these 30 children are given in the following table.

<table>
<thead>
<tr>
<th>Child</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td>23</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>26</td>
<td>12</td>
</tr>
<tr>
<td>27</td>
<td>12</td>
</tr>
<tr>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
</tr>
</tbody>
</table>


The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™, Statway™, and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

(9) The random assignment was started by flipping a coin to decide whether Child 1 would be assigned to the Method 1 group or the Method 2 group. The toss resulted in a T, so Child 1 was assigned to Method 2, as indicated in the table. For each of the remaining children, flip a coin to determine which group the child should be assigned to and place the child in the appropriate group in the table. Continue until you have placed all of the children in one of the two groups. (Using this method for random assignment, you will probably not end up with exactly 15 children in each group, but that is okay.)

<table>
<thead>
<tr>
<th>Method 1 Group</th>
<th>Method 2 Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(10) Using the blank dotplots at the end of this lesson, construct a dotplot of the ages of children assigned to the Method 1 group and a dotplot of the ages of the children assigned to the Method 2 group.
Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

(11) Calculate the average age for the children assigned to the Method 1 group and the average age for the children assigned to the Method 2 group. Write these values on the dotplots you created in Question 10. Which group has the largest average age? Is it substantially larger than the average age for the other group, or are the two averages about the same?

(12) Considering the dotplots and the two averages, do you think the random assignment was successful in creating groups that were similar with respect to the ages of the children in the groups?

(13) Other variables that might affect recovery time are weight and fitness level. Do you think that the random assignment to experimental conditions would create groups that are similar on these two variables as well as age? Why do you think so?
Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

Homework

(1) A study done by researchers at Kings College London found that *infomania* has a temporary, but significant, negative effect on intelligence quotient (IQ). In the experiment, the researchers divided volunteers into two groups. Each subject took an IQ test. One group had to check e-mail and respond to instant messages while taking the test, and the other group took the IQ test without any distractions. Researchers found that the average score on the IQ test for the distracted group was 10 points lower than the average for the group that was not distracted.³

(a) What is the response variable in this experiment?

(b) What is the treatment variable and what are the experimental conditions?

(c) Explain why the researchers should use random assignment to put each volunteer in one of the experimental groups rather than letting the volunteers decide which group they wanted to be in.

Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

(2) Do ethnic group and gender influence the type of care that a heart patient receives? The following passage is from a 1999 USA Today article:\(^4\)

Previous research suggested blacks and women were less likely than whites and men to get cardiac catheterization or coronary bypass surgery for chest pain or a heart attack. Scientists blamed differences in illness severity, insurance coverage, patient preference, and health care access. The researchers eliminated those differences by videotaping actors—two black men, two black women, two white men, and two white women—describing chest pain from identical scripts. They wore identical gowns, used identical gestures, and were taped from the same position. Researchers asked 720 primary care doctors at meetings of the American College of Physicians or the American Academy of Family Physicians to watch a tape and recommend care. The doctors thought the study focused on clinical decision making.

The video a particular doctor watched was determined by the roll of a four-sided die.

(a) In the experiment described, the researchers chose to control a number of variables other than ethnic group and gender that might have influenced the recommended treatment. How did they do this?

(b) Did the researchers use random assignment to create the experimental groups? If so, how did they do this?

---

Initiating Lesson 1.3.1: Collecting Data by Conducting an Experiment

Dotplots of Ages for the Two Experimental Groups (Task 3)

Average Age:

Average Age:
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

Estimated number of 50-minute class sessions: 1

Learning Goals
Students will begin to understand
- the purpose of including a control group in an experiment.
- the purpose of blinding in an experiment.

Students will begin to be able to
- given a description of an experiment, identify which of the following was part of the experimental design: use of a control group, use of a placebo, and blinding.
- describe characteristics of a well-designed experiment.

Introduction [Student Handout]
In Lesson 1.3.1, direct control and the use of random assignment to create experimental groups were introduced as important considerations in planning an experiment. This lesson looks at a few additional things you may want to consider when planning an experiment or evaluating someone else’s plan.

Task 1: Blinding [Student Handout, estimated time: 15 minutes]
Sometimes people have set ideas about the effectiveness of various experimental conditions in an experiment, and these prior beliefs might influence the response.

(1) Suppose you are planning an experiment to compare the effect of two different doses of a medication on the pain experienced after knee surgery. Patients are assigned at random to either a low-dose or high-dose group. They are given the medication after surgery and then asked to rate their level of pain one hour later. If patients know they get the low dose or high dose of the medication, do you think it might influence how they rate the pain level after one hour? Why do you think so?

(Answer: If you tell patients they are receiving a higher dosage of pain medication, this will probably affect their rating of pain. They might become convinced that the pain is less simply from the mere suggestion that it should be less.)

(2) Given your answer to Question 1, can you suggest a strategy that eliminates any concerns you might have about prior beliefs influencing the response?

(Answer: Do not tell the patients which dose they are receiving. Note: For ethical reasons, when patients agree to participate in the experiment, they should be informed that there are two dosage levels and they could receive either one.)

When participants in an experiment do not know which experimental condition they are assigned to, they are said to be blinded. By making sure participants do not know what experimental condition they are assigned to, you can prevent prior beliefs about the experimental conditions from influencing the response.

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™, Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

(3) While blinding participants is often a good idea, it is not always possible. Give an example of an experiment where it is not possible for the subjects to be blind to the experimental condition.

(Answer: Many answers are possible. Example: Does eating breakfast in the morning reduce the number of calories consumed at lunch on average versus skipping breakfast? Participants are fully aware of whether they ate breakfast. Because of this, it is also a good idea that they not know the exact response under investigation or the overall goals of the study so as to not influence their behavior.)

Another form of blinding in an experiment is to plan the experiment in a way that ensures that whoever measures the response does not know which experimental group a participant is assigned to. For example, when student essays on an exam like the Scholastic Aptitude Test (SAT) are scored, the grader often uses judgment to decide between adjacent scores (a paper might fall between a 4 and 5 on the grading scale and the grader must decide between these two scores).

Suppose you want to conduct an experiment to compare essay scores for students participating in one of two different review courses: an 8-hour course and a 40-hour course. A grader deciding between two adjacent scores might unconsciously be influenced in one direction if he or she knows that the essay was written by a student who had completed the 40-hour review course. This problem could be eliminated by blinding the person measuring the response.

When possible, it is usually a good idea to consider blinding both the participants and the person measuring the response.

(4) In the experiment to assess the effect of review course length on exam scores—an 8-hour course versus a 40-hour course to prepare for the Critical Reading portion of the SAT exam—is it possible to blind the students participating in the experiment? Explain your reasoning.

(Answer: No, students know how long the course was, though they may not know the length of the other review course.)

(5) In an experiment to compare two different products for waterproofing hiking boots, participants are given a pair of hiking boots that were waterproofed using one of the two methods. Participants then wear the boots on a hike along a mountain stream that involves crossing the stream in several places. After the hike, the boots are left to dry and then sent to a lab where a technician evaluates the boots for water damage. Do you recommend blinding only the participants, only the lab technician, or both the participants and lab technician? Explain your reasoning.

(Answer: Recommend blinding both the participants and the lab technician evaluating the boots for water damage. The participants could behave differently if they believed the boots are less waterproof. The technician also uses some subjective judgment in rating the water damage and therefore should not be subconsciously influenced by knowing which boots are supposed to show less damage.)
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

6. In this experiment, the waterproofing product is applied to the boots before the boots are given to participants. Is it important that the person applying the waterproofing be blinded? Why do you think so?

(App: This person will probably not come into contact with participants and plays no role in evaluating the performance of the boots, so it is okay if this person is not blinded. Note: Students may answer yes, with a justification. Just make sure they focus on creating any possible biases or influences on the boot’s performance, like applying double coats for one product and not the other.)

Wrap-Up [Instructor-led discussion]

The following are key points to make:

- Additional steps can sometimes be taken to further ensure there are no differences between the treatment groups
- Blinding in particular is generally a good idea as there is a fair bit of evidence of the power of suggestion.
- Blinding is not always possible (e.g., you cannot really blind subjects to color of soda if that is the treatment variable under investigation!), but should be used when it makes sense.

Task 2: Control Groups and Placebos [Student Handout, estimated time: 15 minutes]

Consider an experiment to investigate whether a regular exercise program can reduce resting heart rate. The researchers conducting the experiment have proposed the following experimental groups:

- **Group 1**: Participants swim for 30 minutes each day for a month.
- **Group 2**: Participants walk for 30 minutes each day for a month.

There are 120 people who volunteered to participate in the experiment, and the researchers plan to use random assignment to assign each volunteer to one of the two experimental groups. Resting heart rate will be measured at the beginning of the experiment and again at the end of the experiment. The change in resting heart rate will then be computed for each participant.

7. Suppose at the end of the month, a dotplot of the changes in resting heart rate for those in Group 1 looks very similar to the dotplot of the changes in resting heart rate for those in Group 2. One possible explanation for seeing this is that swimming and walking have no effect on resting heart rate. What is a different explanation for seeing similar dotplots?

(App: Examples include the following:

- Both treatments have an effect, but the effect is similar. [Note: Not thinking the groups were different to begin with is a good explanation because random assignment was used to create the groups.]
- Random chance—There may be a difference, and it just was not seen this time.
- The time period was not long enough to see an effect.)
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

(8) Can you think of a way to modify the experiment so that the change in resting heart rate data allows you to see whether there is a difference between swimming and walking on resting heart rate and also whether there was any effect at all on change in resting heart rate?

(Answer: Students need to know about changes in resting heart rates for individuals who do not exercise at all. This will help determine whether swimming or walking makes a difference (versus just natural changes in resting heart rate over time).)

If one goal of an experiment is to determine whether some experimental condition has an effect, it is important to make sure that the experiment includes a group that corresponds to “no treatment.” This group is called a control group. Including a control group in an experiment provides a baseline for comparison.

Consider the following description of an experiment:

Researchers studied 208 infants whose brains were temporarily deprived of oxygen due to complications at birth. Once the oxygen deprivation was detected, these babies were then randomly assigned to either usual care (the control group) or to a whole-body cooling group. The goal was to see whether reducing body temperature for three days after birth increased the chance of survival without brain damage.\(^1\)

(9) Why was it important for researchers to include a control group in this experiment?

(Answer: To have a comparison group. The researchers would not be able to determine whether the body cooling worked without knowing the survival rate of infants not receiving body cooling.)

In experiments that use human subjects, use of a control group may not be enough to establish whether a treatment really has an effect. Studies have shown that people sometimes respond in a positive way to treatments that have no active ingredients, such as colored water or sugar pills (called placebos). People often report that these nontreatments relieve pain or reduce other symptoms such as dizziness.

Because of this placebo effect, experiments often include a control group and a placebo group. Those in the placebo group get a placebo that is identical in appearance (and taste, etc.) to what the people in other experimental groups receive, but has no active ingredients. In this situation, you also want the participants blinded—you do not want to tell people they are receiving a placebo!

(10) How does including a control group and a placebo group in an experiment allow you to decide whether a placebo effect existed?

(Answer: Because the placebo is an empty treatment, you should not expect any difference in response from those in the control group who receive no treatment. If researchers find a difference in the response between the control group and the placebo group, this indicates a placebo effect existed.)

Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

(11) How does including a placebo group in an experiment allow you to decide whether a particular experimental condition has a real effect on the response?

(Answer: If the experimental group does not outperform the placebo group, this indicates any effect arises from the placebo effect rather than the treatment condition.)

(12) The body cooling experiment described earlier included a control group but not a placebo group. Is it a good idea to also include a placebo group? If so, explain how you might do this. If not, explain why you do not think a placebo group is necessary.

(Answer: Researchers are observing physiological reactions in infants, which are probably not influenced by the placebo effect. There is less subjective judgment in evaluating the response variable in this study, so blinding may not be necessary. Also, for ethical reasons, researchers probably do not want a condition equivalent to a fake treatment.)

Wrap-Up [Instructor-led discussion]

The following are key points to make:

- Using a control group in an experimental design allows for comparison. This helps you decide whether an observed improvement may have occurred naturally without any intervention.
- Using a placebo group allows you to see whether the treatment has an impact over and above placebo effect.

Task 3: Topic Wrap-Up2 [Student Handout, estimated time: 15 minutes]

(Note: This topic wrap-up can either be done as a whole-group discussion or by having students talk about it in small groups and then bring the class back together after 7–8 minutes to talk about the questions the groups came up with.)

An advertisement for a sweatshirt that appeared in SkyMall Magazine (a catalog distributed by some airlines) stated the following:

This is not your ordinary hoody! Why? Fact: Research shows that written words on containers of water can influence the water’s structure for better or worse depending on the nature and intent of the word. Fact: The human body is 70% water. What if positive words were printed on the inside of your clothing?

For only $79, you could purchase a hooded sweatshirt that had more than 200 positive words (such as hope, gratitude, courage, and love) in 15 different languages printed on the inside of the sweatshirt so you could benefit by being surrounded by these positive words. The reference to the “fact” that written words on containers of water can influence the water’s structure appears to be based on the work of Dr. Masaru Emoto, who typed words on paper, pasted the words on bottles of water, and observed how the water reacted to the words by seeing what kind of crystals were formed in the water. He describes several of the experiments in his self-published book, The Message from Water.

---

2 Adapted from Introduction to Statistics and Data Analysis, Cengage Learning, 2012.

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.aidanacenter.org/mathways.
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

If you interviewed Dr. Emoto, what questions would you want to ask about this experiment? How would knowing the answers to these questions help you evaluate his experiments?

(Note: Possible questions include the following:

- How many different bottles of water were used?
- How many different experimental conditions were there and what were they?
- What exactly was the response variable and how was it measured?
- Was the person measuring the response variable blind to the words used on the bottle?
- How were the bottles of water assigned to treatment conditions?
- What characterized words as positive?
- Were the words always the same size? [e.g., Did they block the same amount of light?]
- What other types of controls were used? [e.g., Were the bottles kept in the same room, at the same temperature, same amount of time, etc.?])

Homework [Student Handout]

(1) An experiment to investigate whether dogs can be trained to recognize cancer by smell was described in a newspaper article\(^3\) with the headline, “Doctor Dogs Diagnose Cancer by Sniffing It Out.” In the experiment, dogs were trained to distinguish between people with breast and lung cancer and people without cancer by sniffing exhaled breath. Dogs were trained to lay down if they detected cancer in a breath sample. After training, dogs’ ability to detect cancer was tested using breath samples from people whose breath had not been used in training the dogs. The paper states, “The researchers blinded both the dog handlers and the experimental observers to the identity of the breath samples.” Explain why this blinding is important.

(Answer: If the dog handler or the observers knew which was the correct breath sample they could inadvertently pass some signals on to the dogs and influence the dogs’ behavior.)

(2) Swedish researchers concluded that viewing and discussing art can lead to improvement in medical conditions such as high blood pressure. This conclusion was based on an experiment in which 20 elderly women gathered once a week to discuss different works of art. The study also included a group of 20 elderly women who met once a week to discuss their hobbies and interests. At the end of four months, the art discussion group was found to have lower blood pressure than the hobbies discussion group.

(a) Why is it important to determine whether the researchers randomly assigned the women participating in the study to one of the two groups?

(Answer: To ensure that there was no real difference between the groups—especially with respect to initial blood pressure—prior to the imposition of the treatment variable. If the participants self-selected which group they were in, those who chose the art group may have had lower blood pressure to begin with.)

---

\(^3\) Doctor dogs diagnose cancer by sniffing it out. (2006, January 9). Knight Ridder Newspapers.
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

(b) Why do you think the researchers included the hobbies discussion group in this study?

(Answer: This provides a comparison group to make sure it was not just meeting together that lowered blood pressure and allowed the researchers to assess the effect of the type of activity during the meeting.)

c) Suppose the experiment had also included a third group of women who did not meet at all, and the researchers randomly assigned participants to one of the three groups. Is including this third group an improvement over what was done in the actual study? Why do you think so?

(Answer: This helps determine whether the blood pressure levels were influenced by the women meeting together, regardless of what they talk about.)
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

In Lesson 1.3.1, direct control and the use of random assignment to create experimental groups were introduced as important considerations in planning an experiment. This lesson looks at a few additional things you may want to consider when planning an experiment or evaluating someone else’s plan.

Task 1: Blinding

Sometimes people have set ideas about the effectiveness of various experimental conditions in an experiment, and these prior beliefs might influence the response.

(1) Suppose you are planning an experiment to compare the effect of two different doses of a medication on the pain experienced after knee surgery. Patients are assigned at random to either a low-dose or high-dose group. They are given the medication after surgery and then asked to rate their level of pain one hour later. If patients know they get the low dose or high dose of the medication, do you think it might influence how they rate the pain level after one hour? Why do you think so?

(2) Given your answer to Question 1, can you suggest a strategy that eliminates any concerns you might have about prior beliefs influencing the response?
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

When participants in an experiment do not know which experimental condition they are assigned to, they are said to be blinded. By making sure participants do not know what experimental condition they are assigned to, you can prevent prior beliefs about the experimental conditions from influencing the response.

(3) While blinding participants is often a good idea, it is not always possible. Give an example of an experiment where it is not possible for the subjects to be blind to the experimental condition.

Another form of blinding in an experiment is to plan the experiment in a way that ensures that whoever measures the response does not know which experimental group a participant is assigned to. For example, when student essays on an exam like the Scholastic Aptitude Test (SAT) are scored, the grader often uses judgment to decide between adjacent scores (a paper might fall between a 4 and 5 on the grading scale and the grader must decide between these two scores).

Suppose you want to conduct an experiment to compare essay scores for students participating in one of two different review courses: an 8-hour course and a 40-hour course. A grader deciding between two adjacent scores might unconsciously be influenced in one direction if he or she knows that the essay was written by a student who had completed the 40-hour review course. This problem could be eliminated by blinding the person measuring the response.

When possible, it is usually a good idea to consider blinding both the participants and the person measuring the response.

(4) In the experiment to assess the effect of review course length on exam scores—an 8-hour course versus a 40-hour course to prepare for the Critical Reading portion of the SAT exam—is it possible to blind the students participating in the experiment? Explain your reasoning.
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

(5) In an experiment to compare two different products for waterproofing hiking boots, participants are given a pair of hiking boots that were waterproofed using one of the two methods. Participants then wear the boots on a hike along a mountain stream that involves crossing the stream in several places. After the hike, the boots are left to dry and then sent to a lab where a technician evaluates the boots for water damage. Do you recommend blinding only the participants, only the lab technician, or both the participants and lab technician? Explain your reasoning.

(6) In this experiment, the waterproofing product is applied to the boots before the boots are given to participants. Is it important that the person applying the waterproofing be blinded? Why do you think so?
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

Task 2: Control Groups and Placebos

Consider an experiment to investigate whether a regular exercise program can reduce resting heart rate. The researchers conducting the experiment have proposed the following experimental groups:

Group 1: Participants swim for 30 minutes each day for a month.
Group 2: Participants walk for 30 minutes each day for a month.

There are 120 people who volunteered to participate in the experiment, and the researchers plan to use random assignment to assign each volunteer to one of the two experimental groups. Resting heart rate will be measured at the beginning of the experiment and again at the end of the experiment. The change in resting heart rate will then be computed for each participant.

(7) Suppose at the end of the month, a dotplot of the changes in resting heart rate for those in Group 1 looks very similar to the dotplot of the changes in resting heart rate for those in Group 2. One possible explanation for seeing this is that swimming and walking have no effect on resting heart rate. What is a different explanation for seeing similar dotplots?

(8) Can you think of a way to modify the experiment so that the change in resting heart rate data allows you to see whether there is a difference between swimming and walking on resting heart rate and also whether there was any effect at all on change in resting heart rate?
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

If one goal of an experiment is to determine whether some experimental condition has an effect, it is important to make sure that the experiment includes a group that corresponds to “no treatment.” This group is called a control group. Including a control group in an experiment provides a baseline for comparison.

Consider the following description of an experiment:

Researchers studied 208 infants whose brains were temporarily deprived of oxygen due to complications at birth. Once the oxygen deprivation was detected, these babies were then randomly assigned to either usual care (the control group) or to a whole-body cooling group. The goal was to see whether reducing body temperature for three days after birth increased the chance of survival without brain damage.¹

(9) Why was it important for researchers to include a control group in this experiment?

In experiments that use human subjects, use of a control group may not be enough to establish whether a treatment really has an effect. Studies have shown that people sometimes respond in a positive way to treatments that have no active ingredients, such as colored water or sugar pills (called placebos). People often report that these nontreatments relieve pain or reduce other symptoms such as dizziness.

Because of this placebo effect, experiments often include a control group and a placebo group. Those in the placebo group get a placebo that is identical in appearance (and taste, etc.) to what the people in other experimental groups receive, but has no active ingredients. In this situation, you also want the participants blinded—you do not want to tell people they are receiving a placebo!

(10) How does including a control group and a placebo group in an experiment allow you to decide whether a placebo effect existed?

Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

(11) How does including a placebo group in an experiment allow you to decide whether a particular experimental condition has a real effect on the response?

(12) The body cooling experiment described earlier included a control group but not a placebo group. Is it a good idea to also include a placebo group? If so, explain how you might do this. If not, explain why you do not think a placebo group is necessary.
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

Task 3: Topic Wrap-Up

An advertisement for a sweatshirt that appeared in *SkyMall Magazine* (a catalog distributed by some airlines) stated the following:

This is not your ordinary hoody! Why? Fact: Research shows that written words on containers of water can influence the water’s structure for better or worse depending on the nature and intent of the word. Fact: The human body is 70% water. What if positive words were printed on the inside of your clothing?

For only $79, you could purchase a hooded sweatshirt that had more than 200 positive words (such as *hope, gratitude, courage,* and *love*) in 15 different languages printed on the inside of the sweatshirt so you could benefit by being surrounded by these positive words. The reference to the “fact” that written words on containers of water can influence the water’s structure appears to be based on the work of Dr. Masaru Emoto, who typed words on paper, pasted the words on bottles of water, and observed how the water reacted to the words by seeing what kind of crystals were formed in the water. He describes several of the experiments in his self-published book, *The Message from Water.*

If you interviewed Dr. Emoto, what questions would you want to ask about this experiment? How would knowing the answers to these questions help you evaluate his experiments?

---

2 Adapted from Introduction to Statistics and Data Analysis, *Cengage Learning*, 2012.

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Supporting Lesson 1.3.2: Other Design Considerations—Blinding, Control Groups, and Placebos

Homework

(1) An experiment to investigate whether dogs can be trained to recognize cancer by smell was described in a newspaper article\(^3\) with the headline, “Doctor Dogs Diagnose Cancer by Sniffing It Out.” In the experiment, dogs were trained to distinguish between people with breast and lung cancer and people without cancer by sniffing exhaled breath. Dogs were trained to lay down if they detected cancer in a breath sample. After training, dogs’ ability to detect cancer was tested using breath samples from people whose breath had not been used in training the dogs. The paper states, “The researchers blinded both the dog handlers and the experimental observers to the identity of the breath samples.” Explain why this blinding is important.

\(^{3}\) Doctor dogs diagnose cancer by sniffing it out. (2006, January 9). Knight Ridder Newspapers.
(2) Swedish researchers concluded that viewing and discussing art can lead to improvement in medical conditions such as high blood pressure. This conclusion was based on an experiment in which 20 elderly women gathered once a week to discuss different works of art. The study also included a group of 20 elderly women who met once a week to discuss their hobbies and interests. At the end of four months, the art discussion group was found to have lower blood pressure than the hobbies discussion group.

(a) Why is it important to determine whether the researchers randomly assigned the women participating in the study to one of the two groups?

(b) Why do you think the researchers included the hobbies discussion group in this study?

(c) Suppose the experiment had also included a third group of women who did not meet at all, and the researchers randomly assigned participants to one of the three groups. Is including this third group an improvement over what was done in the actual study? Why do you think so?
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

Estimated number of 50-minute class sessions: 2
Each task in this lesson should take about one 50-minute class session to complete.

Materials Required
The 20 slips of paper cut from the page at the end of this lesson or 20 index cards for each pair of students for Task 2.

Learning Goals
Students will begin to understand the role that chance variability plays in the statistical decision making process.

Students will begin to be able to
- distinguish between a population and a sample in a statistical setting.
- distinguish between categorical and numerical data.
- evaluate the strength of evidence against a claim about a population or against a claim of no effect in an experiment.

Introduction and Review [Student Handout]
(Note: The introduction can read by students or done as a minilecture.)

In this lesson, you will see examples of the two common types of statistical studies introduced in this module: observational studies and experiments.

An observational study involves using data to learn about characteristics of some population. For example, you might want to learn about the proportion of students at your college who work more than 20 hours per week or the average amount of money students at your college spend each month on housing. In either case, the population of interest is all students at the college. It is usually not possible to study the entire population—imagine how much work it would be to collect information on housing cost for every student at your college. When this is the case, the usual solution is to study just part of the population, which is called a sample. You saw in Lessons 1.2.1 and 1.2.2 how important it is to carefully select the sample from the population. Task 1 of this lesson explores how you can use information from a sample to evaluate claims about a population.

An experiment involves using data to answer questions of the type, “What is the effect of ...?” In a statistical experiment, the value of some response is measured under different experimental conditions to evaluate the effect of the different conditions on the response. For example, you might want to learn whether the color of a food product affects how people evaluate its taste. You saw in Lesson 1.3.1 that random assignment to experimental groups allows you to evaluate the effect of the experimental conditions. Task 2 of this lesson explores how you can evaluate whether random chance might still be an explanation for any observed differences in the response variable.
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

Task 1: Drawing Conclusions from an Observational Study [Student Handout, estimated time: 50 minutes]

(Note: The first part of Task 1 [Questions 1–6] is probably best accomplished by having students work in pairs or small groups.)

Begin this task by reviewing part of the description of Study 1 from Lesson 1.1.1:

Researchers at the Center for Reproductive Medicine at Brigham and Women’s Hospital wondered what proportion of women who visit a fertility clinic would want the opportunity to choose the sex of a future child. They also wondered if those who would like to choose the sex were more likely to want a boy or girl. The researchers mailed a survey containing 19 questions to women who had visited the Center. One question on the survey asked women whether they would like the option of choosing the sex of a future child. If the response to that question was yes, a follow-up question asked whether they would choose a boy or girl. Of the 229 women who wanted to choose, 89 (38.9%) said they would choose a boy and 140 (61.1%) said they would choose a girl.

This research study was the topic of an article that appeared on the website LiveScience (March 23, 2005). The following statements are taken from the LiveScience summary:

“When given the chance to choose the sex of their baby, women are just as likely to choose pink socks as blue, a new study shows.”

“Of the 561 women who participated in the study, 229 said they would like to choose the sex of a future child. Among these 229 there was no greater demand for boys or girls.”

The researchers thought it was reasonable to regard the 229 women who participated in the study as representative of the population of women who would like to choose the sex of a future child.

(1) The LiveScience summary of the study indicates that there is no preference for girls over boys for women who would like to choose the sex of a future child. If there really is no preference, what does this imply about the proportion of women in the population who would choose a girl? What do you expect to see for the proportion of 229 women in the sample who would choose a girl?

(Answer: The proportion of women in the population who would choose a girl is 0.5 if there is no preference in the population. Therefore, you should expect to see about half of the 229 women in the sample to specify they would choose a girl.)

(2) The actual study found that 140 of the women in the sample of 229 women chose a girl. What proportion of women in the sample chose a girl?

(Answer: 140/229 = 0.61. Note: You may need to remind students how to compute the proportion as a decimal from the fraction 140/229.)

(3) How does the observed sample proportion who chose a girl in the study compare to what you expect if there is no preference for one sex over the other in the population of women who would like to choose the sex of a future child?

(Answer: The observed sample proportion of 0.61 is larger than the expected proportion of 0.5.)
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(4) What are two possible explanations for why you might observe a sample proportion that was larger than 0.5?

(Answer: Explanation 1—The population proportion is larger than 0.5. Explanation 2—The population proportion is equal to 0.5, but just by chance the researchers happened to select a sample with a larger proportion choosing a girl.)

One possible explanation for why you might observe a sample proportion as large as 0.61, even if the population proportion equals 0.5, is random chance. You do not really expect the proportion observed in any one sample to be exactly equal to 0.5. By random chance alone, you may get more or fewer women picking a girl in different samples. The key question is whether a sample proportion as large as 0.61 is consistent with the type of chance variation you expect to see.

(Note: Remind students at this point about the activities in Lesson 1.1.1, where this kind of question was also considered. In Lesson 1.1.1, they looked at the chance of seeing as many correct responses assuming people were picking answers strictly at random. Here, the focus is not on the randomness in the picking process but on the randomness in the sampling process.)

(5) Suppose you have a large population of 10,000 women who want to choose the sex of a future child and that this population is evenly divided between those who would choose a boy and those who would choose a girl. How might you collect data from this population to help you decide what values are likely to be observed for the sample proportion when a sample of 229 women is randomly selected from a population in which there is no preference for one sex over the other?

(Answer: You need to model the random selection process to assess the amount of sample-to-sample variation by chance alone. In this case, select a random sample of 229 women from this population of 10,000 women and compute the sample proportion that chose a female child for that sample. Then repeat this process a large number of times.)

(6) How does doing what you suggest in Question 5 help you decide whether it is reasonable to think that the proportion of women in the population who would choose a girl could be 0.5?

(Answer: The process outlined in Question 5 tells you about typical values for the sample proportion when the population proportion equals 0.5. By itself, it does not really tell you much. However, if you then look at the actual observed sample proportion, you can see whether it falls among the typical values. If it does, then you have no reason to think the sample did not come from a population where half wanted to choose a girl. On the other hand, if the actual observed sample proportion is not similar to those generated in Question 5, you have evidence that the researchers’ sample came from a population that was not 50/50.)

(Note: At this point it is probably a good idea to pull the class back together and complete the remaining parts of this task as an instructor-led, whole-group discussion.)
For this sample, 122 of the 229 women chose a girl. The sample proportion is $122/229 = 0.53$. So, for this sample, even though the proportion that would choose a girl in the population was 0.5, the sample proportion was greater than 0.5 just by chance, as a result of which women happened to be included in the sample.

(7) Will this process always yield a sample proportion above 0.5? If not, how often do you expect to get a sample proportion above 0.5 and how often below 0.5?

(Answer: This process will not always yield a sample proportion above 0.5. In fact, you should expect the sample proportion to be above 0.5 about half the time and below 0.5 about half the time.)
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(8) Use the dotplot to answer the following questions.

(a) What was the smallest sample proportion observed?
   (Answer: 0.41)

(b) What was the largest sample proportion observed?
   (Answer: 0.59)
   (Note: You may want to note that the sample proportions are centered around 0.5, with about half above 0.5 and about half below 0.5.)

(c) How many of the 250 samples from the population of 10,000 women where half would choose a girl and half would choose a boy resulted in a sample proportion as large as 0.61?
   (Answer: None of the 250 samples had a sample proportion as large as 0.61.)

(9) For the actual sample of 229 women from the study, the proportion that would choose a girl was 0.61. Do you think this is convincing evidence that the proportion of women in the population of all women who would like to choose the sex of a future child is 0.5? Why or why not?
   (Answer: This provides evidence that the actual sample did not come from a population where the proportion that would choose a girl is 0.5. The dotplot shows that it is very unlikely to get a sample proportion as large as 0.61 when the population proportion is equal to 0.50, just by chance. So the fact that the researchers got such a large sample proportion is convincing evidence that the observed proportion of 0.61 is not higher than 0.50 just by chance [random sampling] alone.)
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(10) Do you agree with the conclusion in the *LiveScience* summary given at the beginning of this task? If not, suggest a rewording of the summary that you think is a better description of the study results. Explain your reasoning.

(Answer: The conclusion does not seem to agree with analysis, which indicates that there is evidence that women who want to choose the sex of the child are not equally likely to choose a girl or boy. The proportion was not extremely large \([0.61]\) but was large enough to be convincing that chance alone is not a reasonable explanation for the sample proportion the researchers observed.)

Wrap-Up of Task 1 [Instructor-led discussion]

The following are points to emphasize:

- Sample proportions are expected to differ from the population proportion by chance alone. However, how they differ is not completely haphazard. Some sample proportions are considered typical (those near the population proportion) and some are considered surprising (those far from the population proportion).
- By replicating the random sampling process from a specific population, you can determine whether the observed sample proportion is consistent with the chance variation you expect to see in this random sampling process from that population.
- If the sample proportion value that was actually observed in the research study is unlikely to occur by chance alone, you have evidence that the research data did not come from the specified population.

Task 2: Drawing Conclusions from an Experiment [Student Handout, estimated time: 50 minutes]

(Note: The first part of Task 2 [Questions 11–25] is probably best accomplished by having students work in pairs or small groups. You may want to ask them to read the Study 2 description from Lesson 1.1.1 and answer Questions 11–14 prior to coming to class.)

Statistical experiments are usually conducted to collect data that allow you to answer questions like “What happens when ...?” or “What is the effect of ...?” In a statistical experiment, the value of some response is measured under different experimental conditions.

Begin this task by taking a look at the part of the description of Study 2 from Lesson 1.1.1:

Psychologists believe that people are less likely to do something if they think it will require a lot of effort. But how do people decide what things they think will be hard and what things they think will be easy? Researchers at the University of Michigan wondered whether how difficult it was to read the instructions for how to perform a task would affect how hard people thought the task would be. To investigate this, they performed an experiment. Twenty students were divided at random into two groups of 10 students each. One group received instructions for an exercise routine printed in a font that was easy to read, and the other group received the same set of instructions printed in a font that was difficult to read. A sample of each font appears below. Each student read the instructions, and then they were asked how many minutes they thought the exercise routine would take. For the group that read the instructions printed in an easy-to-read...
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

font, the average for the number of minutes they thought the routine would take was 8.23. For the group that read the same instructions printed in the font that was hard to read, the average was 15.1 minutes.

This is the easy-to-read font that was used in the study.
This is the hard-to-read font that was used in the study.

This study is a statistical experiment.

(11) What research question was this experiment trying to answer?
(Answer: Does the type of font affect how challenging people judge a task to be?)

(12) What is the response measured in this experiment?
(Answer: How many minutes people thought the task would take)

(13) What are the two experimental conditions for this experiment?
(Answer: Easy-to-read font and hard-to-read font)

(14) The 20 students participating in the experiment were divided into two experimental groups at random. Why do you think the researchers did this?
(Answer: The random assignment was to create groups that are similar with respect to other variables that might affect the response.)

Suppose the following data are the estimated time to complete the exercise routine (in minutes) given by the 20 students who participated in the study. (These data are consistent with summary values given in the paper.)

<table>
<thead>
<tr>
<th>Easy-to-Read Group</th>
<th>9</th>
<th>14</th>
<th>16</th>
<th>13</th>
<th>5</th>
<th>6</th>
<th>2</th>
<th>6</th>
<th>3</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard-to-Read Group</td>
<td>19</td>
<td>17</td>
<td>12</td>
<td>17</td>
<td>13</td>
<td>9</td>
<td>12</td>
<td>24</td>
<td>19</td>
<td>9</td>
</tr>
</tbody>
</table>

Notice that the observations here are numerical. This is different from the study described in Task 1, where the observations were girl or boy.
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(15) Plot these data on the axes shown below to create two dotplots.

(Note: You may need to remind students how to plot points. In addition, make sure they realize that now the data are quantitative and that is why dotplots are the appropriate graphical summary.)

![Dotplot for easy-to-read group](image1)

![Dotplot for hard-to-read group](image2)

(Answers:

Easy Font: ● ● ● ● ● ● ● ● ● ●

Hard Font: ● ● ● ● ● ● ● ● ● ●

Estimated time to complete task (minutes)

(16) What is one way the estimated times are different for the two groups? What is one way the estimated times are similar for the two groups?

(Answer: The estimated times for the hard-to-read font tend to be larger than the estimated item for the easy-to-read font. The variability—how spread out the values are—in the two distributions is similar.)

(17) Use the data given in the table to compute the average time estimate for completing the exercise routine for the easy-to-read instructions group. Does this confirm the value of 8.2 minutes given for the average in the study description?

(Answer: 9 + 14 + 16 + 13 + 5 + 6 + 2 + 6 + 3 + 8 = 82

82/10 = 8.2 minutes

This does match the study description.)
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(18) Use the data given in the table to compute the average time estimate for the hard-to-read instructions group. Does this confirm the value of 15.1 minutes given for the average in the study description?

(Answer: $19 + 17 + 12 + 17 + 13 + 9 + 12 + 24 + 19 + 9 = 151$

$151/10 = 15.1$

It matches.)

(19) How much greater is the average for the hard-to-read instructions group than the average for the easy-to-read instructions group? (Calculate as hard-to-read average – easy-to-read average.)

(Answer: $15.1 – 8.2 = 6.9$ minutes)

(20) One possible explanation for why the average time estimate for the hard-to-read instructions group is greater than the average for the easy-to-read instructions group is that the font the directions were printed in has an effect, and the difference in the averages reflects this. What is another possible explanation for the difference?

(Answer: Another explanation is that there really is no difference, but the researchers just happened by luck of the draw to get lower estimate givers in the easy-to-read group and higher estimate givers in the hard-to-read group. Random assignment is supposed to create equivalent groups, but perhaps they were not exactly equivalent.)

Suppose the font instructions were printed in really made no difference in how long a student estimates the exercise routine will take to complete. In other words, the first student would have said 9 minutes, whether the instructions were in the easy-to-read or hard-to-read font.

This means the difference in the group averages observed in the study is really just due to chance. In other words, it is just due to the random division of the 20 students into two groups of 10. (Possibly by luck of the draw alone, more pessimistic people ended up in the hard-to-read group.)

If you can rule out this explanation by showing that you are unlikely to observe a difference in the group averages as large as what occurred in the study just by chance, you can conclude that the data from the experiment provide convincing evidence in support of the theory that people think the exercise routine will take longer when the instructions are in a hard-to-read font.

(Note: A page that can be cut up into data slips appears at end of this lesson. The coding [boldface or italics] corresponds to the original group assignments.)

Your instructor will provide you with 20 slips of paper or index cards. Each slip of paper has one of the 20 estimated times written on it. To explore what values for the difference in group averages you might expect to see just by chance (assuming font size does not make a difference), start by doing the following:

- Mix the slips of paper well.
- Randomly select 10 to represent the easy-to-read group.
- The remaining 10 represent the hard-to-read group.
(21) Calculate the average of the 10 values you selected for the easy-to-read group. Calculate the average of the 10 values that you selected for the hard-to-read group. Finally, calculate the difference: hard-to-read average – easy-to-read average.

(Answer: Results will vary, but make sure everyone subtracts in the same direction. Some students will get positive differences and some negative differences.)

(22) What does a negative value for the difference imply in terms of the two averages?

(Answer: A negative value for the difference implies the easy-to-read average is larger than the hard-to-read average. Note: You may need to do a bit of instruction on negative numbers at this point if students are not comfortable with this.)

(23) Is the actual difference observed in the study (6.9 minutes) greater than, equal to, or smaller than the chance difference that you computed in Question 21?

(Note: Results will vary, but 6.9 is probably greater than what students computed.)

(24) Repeat the process of mixing the slips and randomly selecting 10 to represent the easy-to-read group. Calculate the two group averages and the difference in the averages.

(Note: Results will vary. Again make sure everyone is subtracting in the same order. Ask whether the same difference in averages was found both times they randomized. In discussing this, you can emphasize the random assignment to random assignment variability, similar to what they saw earlier with sampling. They are now modeling the randomness in the assignment process.)

(25) Can these two differences alone tell you much about how unusual a difference of 6.9 minutes is? Explain your reasoning.

(Answer: No, you need to see many more values to help judge the pattern in the results, and in particular whether 6.9 is typical or unusual.)

(Note: Have each student or group come up to the board and add the two chance differences they obtained to a class dotplot of the differences. Talk about this being a start, but that they would need many more trials to really get a sense of what the chance distribution of differences in averages would look like.

This is a good point to bring the class back together so you can complete the rest of this task as an instructor-led, whole-group discussion.)

Below is a dotplot constructed from repeating the process of mixing and randomly dividing the given times estimates into two groups of 10 and then computing the difference in group average time estimates. The dotplot is based on 200 repetitions.

(Note: Spend some time talking about this dotplot—make sure students see that this shows values of the difference in group averages that could be expected to happen just by chance when there is really no effect of font.)

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(26) How many dots are in the dotplot? What does each dot represent? Where is the distribution centered? Did you expect this?

(Answer: There are 200 dots. Each dot represents one random assignment and shows the difference in the group averages for that random assignment. The distribution is centered around zero. You expect this because these are the values of the differences when you know there is no genuine font effect.)

(27) Use the dotplot to answer the following questions.

(a) What was the smallest difference in group averages that occurred in the 200 repetitions?
   (Hint: the smallest difference is negative.)

(Answer: About –7)

(b) What was the largest difference in group averages that occurred in the 200 repetitions?

(Answer: About 7)

(c) How many of the 200 chance differences were as large or larger than the difference of 6.9 that was observed in the study?

(Answer: Two of the 200 were 6.9 or larger.)

(d) Does the data from the experiment provide convincing evidence in support of the theory that people tend to think an exercise routine will be more time consuming when the instructions are printed in a font that is hard to read? Explain your reasoning.

(Answer: Yes, this shows that it is rather unusual—though not impossible—to see a difference in the group averages as large as 6.9 when the font does not have an effect on people’s estimates of the requirement time. Because the researchers found a difference of 6.9 in the actual study, this is evidence that these results did not happen by chance alone but instead reflect a genuine font effect.)
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

Wrap-Up [Instructor-led discussion]

The following are key points to make:

- Although the goal of random assignment is to create equivalent groups, by chance you may end up with small differences in the groups.
- By modeling this random assignment process, you can assess how large a difference needs to be between the groups before you think it is larger than expected from unlucky random assignment alone.
- For this analysis, you assumed everyone’s estimate would be what it was, regardless of which treatment group they were assigned to.

You can show the parallels in this analysis to what was done in Task 1. Remind students that these conclusions focus on whether they have strong evidence—an unlucky random sample or random assignment is always possible, but the chances of observing what they did in the study by chance alone may be so small that they should instead say the evidence points strongly to a different explanation.

Homework [Student Handout]

(1) A college administrator is considering two possible ways to reduce cost. One possibility is to reduce the number of hours per week that the student union building is open, and the other possibility is to reduce the number of hours per week that the library stays open. She asked each student in a random sample of 100 students which of the two options they preferred. She found that 68 of the 100 students surveyed chose the student union option.

The administrator wonders whether it is safe to assume, based on this sample proportion, that more than half of the students at the college as a whole would choose the student union option. She knows that one possible explanation for why the proportion of students in the sample who chose this option is greater than 0.5 is that the students at the college are really evenly divided between the two options, but the sample proportion from the survey is greater than 0.5 just by chance.

Describe a process to allow the administrator to decide whether it is reasonable to rule out this explanation and to conclude that there is convincing evidence that more than half of the students at the college would choose the student union option. (Hint: It may be helpful to review Task 1 from this lesson if you need help getting started on this problem.)

(Answer: You can set up a population of the same size as the student population at this college, but with half of the population preferring the student union option and half preferring the library option. Then select a sample of 100 students from this population at random. Compute the proportion in the sample that prefers the student union option. Repeat this process a large number of times and see how often a sample proportion as large as 0.68 is obtained. If 0.68 is an unusual occurrence, this provides evidence that the population at this school is not 50/50.)
In an experiment to investigate the effect of noise on test performance, 30 students were randomly divided into two groups of 15. On test day, one group took the test in a quiet room and the other group took the test in a room while a recording of loud construction noises played. The exam scores for the two groups are shown below.

<table>
<thead>
<tr>
<th>Exam Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiet</td>
</tr>
<tr>
<td>Noisy</td>
</tr>
</tbody>
</table>

Describe a process that would enable you to use the information from this experiment to decide whether the observed difference in average exam score is large enough to convince you that the noisy environment really did have an effect on exam performance (i.e., the difference did not arise solely by random chance.) You do not need to actually carry out the process—just describe the steps you would follow. (Hint: You may want to review Task 2 from this lesson and then think about how you can apply what you learned in that task.)

(Answer: First, calculate the observed difference in group averages for the data:

\[83.4 - 72.7 = 10.7 \text{ points}\]

Put all the scores on slips of paper, then mix them up and divide them into a quiet group and a noisy group. Calculate the difference in the group averages as above \(\text{[quiet minus noisy]}\). Repeat this process a large number of times. Determine how often these random assignments show a difference in group averages as large as 10.7. If it is rare to generate a difference as large as 10.7 by chance \(\text{[random assignment]}\) alone, this provides evidence of a genuine noise effect.)
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

Data Values for Task 2

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>14</td>
<td>16</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>19</td>
<td>17</td>
<td>12</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>24</td>
<td>19</td>
<td>9</td>
</tr>
</tbody>
</table>
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

In this lesson, you will see examples of the two common types of statistical studies introduced in this module: observational studies and experiments.

An observational study involves using data to learn about characteristics of some population. For example, you might want to learn about the proportion of students at your college who work more than 20 hours per week or the average amount of money students at your college spend each month on housing. In either case, the population of interest is all students at the college. It is usually not possible to study the entire population—imagine how much work it would be to collect information on housing cost for every student at your college. When this is the case, the usual solution is to study just part of the population, which is called a sample. You saw in Lessons 1.2.1 and 1.2.2 how important it is to carefully select the sample from the population. Task 1 of this lesson explores how you can use information from a sample to evaluate claims about a population.

An experiment involves using data to answer questions of the type, “What is the effect of …?” In a statistical experiment, the value of some response is measured under different experimental conditions to evaluate the effect of the different conditions on the response. For example, you might want to learn whether the color of a food product affects how people evaluate its taste. You saw in Lesson 1.3.1 that random assignment to experimental groups allows you to evaluate the effect of the experimental conditions. Task 2 of this lesson explores how you can evaluate whether random chance might still be an explanation for any observed differences in the response variable.

Task 1: Drawing Conclusions from an Observational Study

Begin this task by reviewing part of the description of Study 1 from Lesson 1.1.1:

Researchers at the Center for Reproductive Medicine at Brigham and Women’s Hospital wondered what proportion of women who visit a fertility clinic would want the opportunity to choose the sex of a future child. They also wondered if those who would like to choose the sex were more likely to want a boy or girl. The researchers mailed a survey containing 19 questions to women who had visited the Center. One question on the survey asked women whether they would like the option of choosing the sex of a future child. If the response to that question was yes, a follow-up question asked whether they would choose a boy or girl. Of the 229 women who wanted to choose, 89 (38.9%) said they would choose a boy and 140 (61.1%) said they would choose a girl.

This research study was the topic of an article that appeared on the website LiveScience (March 23, 2005). The following statements are taken from the LiveScience summary:

“When given the chance to choose the sex of their baby, women are just as likely to choose pink socks as blue, a new study shows.”

“Of the 561 women who participated in the study, 229 said they would like to choose the sex of a future child. Among these 229 there was no greater demand for boys or girls.”

The researchers thought it was reasonable to regard the 229 women who participated in the study as representative of the population of women who would like to choose the sex of a future child.
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(1) The LiveScience summary of the study indicates that there is no preference for girls over boys for women who would like to choose the sex of a future child. If there really is no preference, what does this imply about the proportion of women in the population who would choose a girl? What do you expect to see for the proportion of 229 women in the sample who would choose a girl?

(2) The actual study found that 140 of the women in the sample of 229 women chose a girl. What proportion of women in the sample chose a girl?

(3) How does the observed sample proportion who chose a girl in the study compare to what you expect if there is no preference for one sex over the other in the population of women who would like to choose the sex of a future child?

(4) What are two possible explanations for why you might observe a sample proportion that was larger than 0.5?
One possible explanation for why you might observe a sample proportion as large as 0.61, even if the population proportion equals 0.5, is random chance. You do not really expect the proportion observed in any one sample to be exactly equal to 0.5. By random chance alone, you may get more or fewer women picking a girl in different samples. The key question is whether a sample proportion as large as 0.61 is consistent with the type of chance variation you expect to see.

(5) Suppose you have a large population of 10,000 women who want to choose the sex of a future child and that this population is evenly divided between those who would choose a boy and those who would choose a girl. How might you collect data from this population to help you decide what values are likely to be observed for the sample proportion when a sample of 229 women is randomly selected from a population in which there is no preference for one sex over the other?

(6) How does doing what you suggest in Question 5 help you decide whether it is reasonable to think that the proportion of women in the population who would choose a girl could be 0.5?

The preferences for one sample of 229 women selected at random from the population described in Question 5 is shown here:

| G | G | B | G | G | G | G | G | B | G | G | B | B | B | B |
| G | B | B | G | B | G | B | G | B | G | G | G | G | G | G |
| G | B | G | B | G | G | B | G | G | G | G | B | G | B | G |
| B | B | B | B | G | G | G | G | G | G | B | B | G | G | B |
| G | B | G | G | B | B | G | B | G | B | G | B | B | B | B |
| B | G | G | G | B | B | G | B | B | B | G | G | G | G | B |
| G | G | B | B | B | G | G | G | B | G | G | G | G | B | G |
| G | G | B | B | B | G | G | G | B | G | G | G | B | G | B |
| B | B | G | B | G | B | G | B | B | B | G | G | B | G | B |
| B | B | G | B | G | B | B | G | G | G | B | G | B | G | B |
| G | G | B | G | B | B | B | B | B | B | G | G | B | B | G |
| B | G | B | B | G | G | B | B | G | B | G | G | G | B | G |
| G | G | B | G | B | B | B | B | B | B | G | G | B | B | G |
| B | G | B | B | G | B | G | B | B | G | G | B | G | G | B |
| G | G | B | G | B | B | B | B | B | B | G | G | B | B | G |

The original versions of the Statway™ and Quantway™ courses were created by The Charles A. Dana Center at The University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching, and are copyright © 2011 by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at The University of Texas at Austin. STATWAY™/Statway™ and Quantway™ are trademarks of the Carnegie Foundation for the Advancement of Teaching. The Dana Center’s frontmatter for Statway™ and Quantway™ is available at www.utdanacenter.org/mathways.
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

For this sample, 122 of the 229 women chose a girl. The sample proportion is $\frac{122}{229} = 0.53$. So, for this sample, even though the proportion that would choose a girl in the population was 0.5, the sample proportion was greater than 0.5 just by chance, as a result of which women happened to be included in the sample.

(7) Will this process always yield a sample proportion above 0.5? If not, how often do you expect to get a sample proportion above 0.5 and how often below 0.5?

This process was repeated 250 times and a dotplot of the resulting 250 sample proportions is shown as follows:

(8) Use the dotplot to answer the following questions.

(a) What was the smallest sample proportion observed?
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(b) What was the largest sample proportion observed?

(c) How many of the 250 samples from the population of 10,000 women where half would choose a girl and half would choose a boy resulted in a sample proportion as large as 0.61?

(9) For the actual sample of 229 women from the study, the proportion that would choose a girl was 0.61. Do you think this is convincing evidence that the proportion of women in the population of all women who would like to choose the sex of a future child is 0.5? Why or why not?

(10) Do you agree with the conclusion in the LiveScience summary given at the beginning of this task? If not, suggest a rewording of the summary that you think is a better description of the study results. Explain your reasoning.
Task 2: Drawing Conclusions from an Experiment

Statistical experiments are usually conducted to collect data that allow you to answer questions like “What happens when ...?” or “What is the effect of ...?” In a statistical experiment, the value of some response is measured under different experimental conditions.

Begin this task by taking a look at the part of the description of Study 2 from Lesson 1.1.1:

Psychologists believe that people are less likely to do something if they think it will require a lot of effort. But how do people decide what things they think will be hard and what things they think will be easy? Researchers at the University of Michigan wondered whether how difficult it was to read the instructions for how to perform a task would affect how hard people thought the task would be. To investigate this, they performed an experiment. Twenty students were divided at random into two groups of 10 students each. One group received instructions for an exercise routine printed in a font that was easy to read, and the other group received the same set of instructions printed in a font that was difficult to read. A sample of each font appears below. Each student read the instructions, and then they were asked how many minutes they thought the exercise routine would take. For the group that read the instructions printed in an easy-to-read font, the average for the number of minutes they thought the routine would take was 8.23. For the group that read the same instructions printed in the font that was hard to read, the average was 15.1 minutes.

This study is a statistical experiment.

(11) What research question was this experiment trying to answer?

(12) What is the response measured in this experiment?

(13) What are the two experimental conditions for this experiment?
### Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(14) The 20 students participating in the experiment were divided into two experimental groups at random. Why do you think the researchers did this?

Suppose the following data are the estimated time to complete the exercise routine (in minutes) given by the 20 students who participated in the study. (These data are consistent with summary values given in the paper.)

<table>
<thead>
<tr>
<th>Easy-to-Read Group</th>
<th>9</th>
<th>14</th>
<th>16</th>
<th>13</th>
<th>5</th>
<th>6</th>
<th>2</th>
<th>6</th>
<th>3</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard-to-Read Group</td>
<td>19</td>
<td>17</td>
<td>12</td>
<td>17</td>
<td>13</td>
<td>9</td>
<td>12</td>
<td>24</td>
<td>19</td>
<td>9</td>
</tr>
</tbody>
</table>

Notice that the observations here are numerical. This is different from the study described in Task 1, where the observations were *girl* or *boy*.

(15) Plot these data on the axes shown below to create two dotplots.

(16) What is one way the estimated times are different for the two groups? What is one way the estimated times are similar for the two groups?
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(17) Use the data given in the table to compute the average time estimate for completing the exercise routine for the easy-to-read instructions group. Does this confirm the value of 8.2 minutes given for the average in the study description?

(18) Use the data given in the table to compute the average time estimate for the hard-to-read instructions group. Does this confirm the value of 15.1 minutes given for the average in the study description?

(19) How much greater is the average for the hard-to-read instructions group than the average for the easy-to-read instructions group? (Calculate as hard-to-read average – easy-to-read average.)

(20) One possible explanation for why the average time estimate for the hard-to-read instructions group is greater than the average for the easy-to-read instructions group is that the font the directions were printed in has an effect, and the difference in the averages reflects this. What is another possible explanation for the difference?
Suppose the font instructions were printed in really made no difference in how long a student estimates the exercise routine will take to complete. In other words, the first student would have said 9 minutes whether the instructions were in the easy-to-read or hard-to-read font.

This means the difference in the group averages observed in the study is really just due to chance. In other words, it is just due to the random division of the 20 students into two groups of 10. (Possibly by luck of the draw alone, more pessimistic people ended up in the hard-to-read group.)

If you can rule out this explanation by showing that you are unlikely to observe a difference in the group averages as large as what occurred in the study just by chance, you can conclude that the data from the experiment provide convincing evidence in support of the theory that people think the exercise routine will take longer when the instructions are in a hard-to-read font.

Your instructor will provide you with 20 slips of paper or index cards. Each slip of paper has one of the 20 estimated times written on it. To explore what values for the difference in group averages you might expect to see just by chance (assuming font size does not make a difference), start by doing the following:

- Mix the slips of paper well.
- Randomly select 10 to represent the easy-to-read group.
- The remaining 10 represent the hard-to-read group.

(21) Calculate the average of the 10 values you selected for the easy-to-read group. Calculate the average of the 10 values that you selected for the hard-to-read group. Finally, calculate the difference: hard-to-read average – easy-to-read average.

(22) What does a negative value for the difference imply in terms of the two averages?
(23) Is the actual difference observed in the study (6.9 minutes) greater than, equal to, or smaller than the chance difference that you computed in Question 21?

(24) Repeat the process of mixing the slips and randomly selecting 10 to represent the easy-to-read group. Calculate the two group averages and the difference in the averages.

(25) Can these two differences alone tell you much about how unusual a difference of 6.9 minutes is? Explain your reasoning.
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

Below is a dotplot constructed from repeating the process of mixing and randomly dividing the given times estimates into two groups of 10 and then computing the difference in group average time estimates. The dotplot is based on 200 repetitions.

![Dotplot Image]

(26) How many dots are in the dotplot? What does each dot represent? Where is the distribution centered? Did you expect this?

(27) Use the dotplot to answer the following questions.

(a) What was the smallest difference in group averages that occurred in the 200 repetitions?  
(Hint: the smallest difference is negative.)

(b) What was the largest difference in group averages that occurred in the 200 repetitions?
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(c) How many of the 200 chance differences were as large or larger than the difference of 6.9 that was observed in the study?

(d) Does the data from the experiment provide convincing evidence in support of the theory that people tend to think an exercise routine will be more time consuming when the instructions are printed in a font that is hard to read? Explain your reasoning.
Homework

(1) A college administrator is considering two possible ways to reduce cost. One possibility is to reduce the number of hours per week that the student union building is open, and the other possibility is to reduce the number of hours per week that the library stays open. She asked each student in a random sample of 100 students which of the two options they preferred. She found that 68 of the 100 students surveyed chose the student union option.

The administrator wonders whether it is safe to assume, based on this sample proportion, that more than half of the students at the college as a whole would choose the student union option. She knows that one possible explanation for why the proportion of students in the sample who chose this option is greater than 0.5 is that the students at the college are really evenly divided between the two options, but the sample proportion from the survey is greater than 0.5 just by chance.

Describe a process to allow the administrator to decide whether it is reasonable to rule out this explanation and to conclude that there is convincing evidence that more than half of the students at the college would choose the student union option. (Hint: It may be helpful to review Task 1 from this lesson if you need help getting started on this problem.)
Initiating Lesson 1.4.1: Drawing Conclusions from Statistical Studies

(2) In an experiment to investigate the effect of noise on test performance, 30 students were randomly divided into two groups of 15. On test day, one group took the test in a quiet room and the other group took the test in a room while a recording of loud construction noises played. The exam scores for the two groups are shown below.

<table>
<thead>
<tr>
<th>Exam Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiet</td>
</tr>
<tr>
<td>86</td>
</tr>
<tr>
<td>Noisy</td>
</tr>
<tr>
<td>77</td>
</tr>
</tbody>
</table>

Describe a process that would enable you to use the information from this experiment to decide whether the observed difference in average exam score is large enough to convince you that the noisy environment really did have an effect on exam performance (i.e., the difference did not arise solely by random chance.) You do not need to actually carry out the process—just describe the steps you would follow. (Hint: You may want to review Task 2 from this lesson and then think about how you can apply what you learned in that task.)