STATISTICS

Informed Decisions Using Data

Michael Sullivan, III

Statistics

Informed Decisions Using Data

Michael Sullivan, III
Joliet Junior College

Prentice Hall
PEARSON EDUCATION, INC.
Upper Saddle River, New Jersey 07458

Library of Congress Cataloging-in-Publication Data

Sullivan, Michael (date)
Statistics: informed decisions using data / Michael Sullivan III.
p. cm.
Includes index.
ISBN 0-13-061864-0
1. Statistics. I. Title.
QA276.12.S85 2004
001.4′22--dc21 2002192695

Editor in Chief/Acquisitions Editor: Sally Yagan
Assistant Editor: Vince Jansen
Associate Editor: Dawn Murrin
Project Manager: Jacquelyn Riotto
Vice President/Director of Production and Manufacturing: David W. Riccardi
Executive Managing Editor: Kathleen Schiaparelli
Senior Managing Editor: Linda Mihatov Behrens
Production Editor: Barbara Mack
Assistant Managing Editor, Math Media Production: John Matthews
Media Production Editor: Donna Crilly
Manufacturing Buyer: Michael Bell
Manufacturing Manager: Trudy Pisciotti
Marketing Manager: Krista M. Bettino
Marketing Assistant: Christine Bayeux
Development Editors: Anne Scanlan-Rohrer and Frank Purcell
Editor in Chief, Development: Carol Trueheart
Editorial Assistant/Print Supplements Editor: Joanne Wendelken
Art Director: Jonathan Boylan
Interior Designers: Judy Matz-Coniglio, Joseph Sengotta, and Geoffrey Cassar
Cover Designer: John Christiana
Art Editor: Thomas Benfatti
Creative Director: Carole Anson
Director of Creative Services: Paul Belfanti
Manager of Formatting: Jim Sullivan
Electronic Production Specialist: Joanne Del Ben
Electronic Page Makeup: Joanne Del Ben, Donna Marie Paukovits, Jackie Ambrosius
Director, Image Resource Center: Melinda Reo
Manager, Rights and Permissions: Zina Arabia
Interior Image Specialist: Beth Boyd-Brenzel
Cover Image Specialist: Karen Sanatar
Image Permission Coordinator: Michelina Viscusi
Photo Researcher: Melinda Alexander
Cover Photos: (From top to bottom) Antonio Mo/Getty Images, Inc.; Yellow Dog Productions/Getty Images, Inc.; Photodisc/Getty Images, Inc.; Michael Rosenfeld/Getty Images, Inc.
Art Studio: Artworks
Senior Manager: Patricia Burns
Production Manager: Ronda Whitson
Manager, Production Technologies: Matt Haas
Project Coordinator: Jessica Einsig
Illustrators: Jackie Ambrosius, Kathryn Anderson, Christopher Kayes, Mark Landis, Bill Neff, Audrey Simonetti, Stacy Smith
Quality Assurance: Pamela Taylor, Kenneth Mooney, Timothy Nguyen

Prentice Hall

Pearson Education, Inc.
Upper Saddle River, New Jersey 07458

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

ISBN 0-13-061864-0

Pearson Education LTD., London
Pearson Education Australia PTY, Limited, Sydney
Pearson Education Singapore, Pte. Ltd
Pearson Education North Asia Ltd, Hong Kong
Pearson Education Canada, Ltd., Toronto
Pearson Educación de Mexico, S.A. de C.V.
Pearson Education—Japan, Tokyo
Pearson Education Malaysia, Pte. Ltd

Statistics

INFORMED DECISIONS USING DATA

To My Wife
Yolanda

and

My Children
Michael, Kevin, and Marissa

PREFACE TO THE INSTRUCTOR

Having taught introductory statistics for the past 10 years at both 2-year and 4-year institutions has brought to my attention some basic needs and issues of the Introductory Statistics course. The diversity of both the students taking the course and the instructors teaching the course drives many of the challenges in presenting statistics at this level. Statistics is a powerful subject and one that I hold great passion for. It is the coupling of my passion for the subject with the desire for a text that would work for me, my students, and my school that led me to decide to pursue writing this textbook.

Before getting started on this text, I spent many hours reflecting on the goals of the project. Overall, I wanted a textbook that instilled in students my enthusiasm and appreciation for the subject and supported the variety of approaches to teaching the subject. In particular, the following were my goals for the text:

- Students see and appreciate the usefulness of the subject
- Students can be more successful in the course
- Students are provided ample practice
- Instructors can present the material using multiple philosophies

Keeping the Students' Interest

An immediate challenge for any instructor teaching this course is the motivation of the students. I always ask my students to pass in their class notes at the end of the semester. By studying these notes, I have been able to determine areas where my students lost interest and where they became excited about the subject. In addition, I was able to determine topics where the concept was lost and topics where the concept was completely understood. I also noticed that my students appreciated when I took a formal definition or theorem and expressed it in everyday language that they could understand. This view into how students read texts, attend lectures, and take notes was invaluable to me in writing this text.

Today's students demand that the material they study have relevance in their everyday lives. One of the joys of teaching a statistics course is that the question, "When will I ever use this stuff?" rarely comes up. However, students still say, "Okay, I can see the usefulness of this if I was a medical researcher, but that is not my goal in life!" Introductory statistics texts need to have interesting, real data sets that students can relate to. They need to see themselves analyzing the data and using the conclusions to make informed decisions that impact their lives. For example, what type of car should I buy? Where should I invest my retirement money? Where should I live? What major should I choose?

Student Success

It is absolutely true that if students experience success early in a course, they will be inclined to work harder as the course progresses. However, if students struggle early, it will lead to disenchanted students who are more willing to "give up" when a difficult concept is introduced.

How can success be experienced while still maintaining the integrity of the course? After all, statistics is a discipline whose theory is deep and mathematically intensive. In addition, understanding the concepts of statistics can be elusive. This does not mean, however, that an introductory textbook should "spare students details" of statistical thought. It is possible, and indeed, desirable, to expose students to the intricacies of statistics. This can be done without delving deep into mathematical thought and losing most of your students.

The text has been written using a conversational tone that "speaks" to the student. This is accomplished through explanations that don't overly complicate the material. Because the text was written using the class notes obtained from my students, the presentation flows like a lecture. The pedagogy that is found in the text mirrors the presentation that I use in my lectures. When my lecture begins, I tie concepts learned earlier in the course to concepts that I am about to present. This is the "Putting It All Together" feature found at the beginning of each chapter. Before tackling a new section, I review topics that are going to be needed in order to succeed in the section. This is the "Preparing for This Section" feature. I then list the objectives that we are going to cover within the section. As I present the material, I provide students with statistically accurate definitions/theorems. I then restate the definitions/theorems using everyday language that doesn't compromise the definition or theorem. This is the "In Your Own Words" feature. I also mention some of the pitfalls in statistical analysis. These pitfalls are displayed in the "Caution" feature. After going over an example in class, I like to have my students attempt to solve a similar problem in class so that I am confident that they understand the concept before continuing. This is the "Now Work Problem xx" feature.

In addition, the text has been written to appeal to a variety of learning styles. There are many, many graphs and figures that allow students to visualize results. For students who learn best through discovery, there are Explorations that guide students through a series of questions that develop statistical concepts. For auditory learners there is a lecture series in which I present all the examples in the text on compact disk or video.

Practice, Practice, Practice

The only way that students will learn statistics is by doing statistics. The exercise sets of a strong statistics text will help develop a student's confidence in, and understanding of, the material. This is accomplished in a twofold manner: (1) through graduated exercise sets, and (2) through problems that ask the student to think about concepts and encourage statistical thinking.

All the exercise sets begin with Concepts and Vocabulary, which are open-ended questions that ask students to define words and explain concepts using their own words. The next portion of the exercise sets is Skill Building. These are drill and practice type problems that develop computational skills that increase student understanding of formulas and concepts. These problems also serve as confidence building problems so that students experience early success. Finally, the Applying the Concepts problems are real data-based problems that ask a variety of questions that help develop solid statistical analysis. Not only do these problems ask the standard questions such as find the mean or compute a 95% confidence interval, they also ask for students to explain the results. In addition, there are questions that ask students to consider some additional questions. For example, what is the impact of an outlier on the arithmetic mean? Is the linear correlation coefficient resistant? What happens if we have outliers when constructing confidence intervals from small data sets? How do I know that the relation between two variables really is linear? These "higher level of thinking" problems truly develop a student's understanding of statistical thinking.

Flexibility

Just as there are many different learning styles, there are many different teaching styles. One of the challenges in writing a text is to create a product that "appeals to the masses." As I survey the halls of the Mathematics Department at Joliet Junior College, I see many high-quality instructors who present the same material in many different ways. Some like to incorporate as much technology as possible into their classroom, some prefer to minimize the technology. Some instructors prefer to use collaborative learning in order to present material, while others utilize lecture. For the Introductory Statistics course, some of the instructors are trained in the discipline of statistics, while others are trained in mathematics, but teach statistics. With these varied backgrounds, it is clear that a text needs to meet the needs of all these backgrounds and teaching philosophies.

Let's consider how technology is presented in the course. Every example in the text is presented using the "by hand" approach. The reason for this is twofold. First, and probably most importantly, by presenting a solution by hand, the student's ability to understand the concept is enhanced. How else can a student understand the concept of linear correlation, except by seeing the product of z-scores? Second, it allows for flexibility in philosophies. If your particular philosophy is to present statistics by utilizing "by hand" solutions, it is there for you. If you are more apt to use technology, then you can utilize the "Using Technology" feature. Following virtually every example is a gray "Using Technology" screen, which provides the output from a TI83+, Minitab, or Excel. In addition, problems that have 15 or more observations in the data set have a CD icon, which indicates to the instructor that the data set is available on the data CD.

Using the Text Efficiently with Your Syllabus

To meet the varied needs of diverse syllabi, this book has been organized with flexibility of use in mind. When structuring your syllabus, notice the topics listed in the "Preparing for This Section" material at the beginning of the section, which will tip you off to dependencies within the course.

The two most common variations within an Introductory Statistics course are the treatment of Regression Analysis and the treatment of Probability.

- **Coverage of Correlation and Regression** The text was written with the descriptive portion of bivariate data (Chapter 4) presented after the descriptive portion of univariate data (Chapter 3). For instructors who prefer to postpone the discussion of bivariate data until later in the course, simply skip Chapter 4 and return to it prior to covering Sections 12.1 and 12.2. Within Chapter 4, an instructor may skip Sections 4.3 and/or 4.4 without loss of continuity.
- **Coverage of Probability** The text allows for a course to present an extensive introduction to probability or light coverage of probability. For instructors wishing to present light coverage of probability, they may cover Section 5.1 and skip the remaining sections. A word of caution is in order, however. Instructors who will be covering the Chi-Square Test for Independence will want to cover Sections 5.1 through 5.3. In addition, any instructor who will be covering Binomial Probabilities will want to cover Independence in Section 5.3 and Combinations in Section 5.5.

Chapter by Chapter Content

Chapter 1 Data Collection

This chapter deals with the methods of obtaining data. There is a detailed presentation of the various sampling techniques along with circumstances under which each is used. In addition, there is an entire section dedicated to nonsampling errors and how to control them. The chapter ends with a detailed discussion of experimental design.

Chapter 2 Organizing and Summarizing Data

This chapter addresses methods for summarizing qualitative data (Section 2.1) and quantitative data (Sections 2.2 and 2.3). For instructors who do not wish to cover topics such as frequency polygons and ogives, Section 2.3 can be skipped without loss of continuity. The chapter ends with a discussion of graphical misrepresentations of data. This section can be covered as a reading assignment.

Chapter 3 Numerically Summarizing Data

Sections 3.1 and 3.2 present numerical measures of central tendency and dispersion. Section 3.3 is optional and can be skipped without loss of continuity. However, if it is skipped and Section 6.1 is covered, proceed slowly through the mean and standard deviation of a discrete random variable. Section 3.4 discusses measures of position including the z-score, percentiles, and outliers. Section 3.5 presents exploratory data analysis.

Chapter 4 Describing the Relation between Two Variables

Section 4.1 introduces scatter diagrams and correlation. Section 4.2 presents the least-squares regression line. Section 4.3 presents the coefficient of determination, residual analysis, and influential observations. This section is particularly important because it allows students to graphically assess whether the data they are analyzing are, in fact, linearly related. Section 4.4 presents methods for using least-squares to fit data to exponential and power models. It is optional and can be skipped without loss of continuity. In addition, the material in this chapter can be postponed until after Chapter 11.

Chapter 5 Probability

Section 5.1 introduces the basic concepts of probability and unusual events. Section 5.2 presents the Addition Rule. Section 5.3 presents the Multiplication Rule. Section 5.4 presents Conditional Probability and can be skipped without loss of continuity. Section 5.5 presents Counting Techniques and can be skipped without loss of continuity with the exception of Combinations.

Chapter 6 Discrete Probability Distributions

The entire chapter is optional. Section 6.1 introduces the concept of a random variable and discrete probability distributions along with expected value. Section 6.2 presents the binomial probability formula and an introduction to inference using binomial probabilities. If you intend to cover Section 6.2, then it is a good idea to also cover Section 5.3. Section 6.3 presents the Poisson probability distribution.

Chapter 7 The Normal Probability Distribution

Sections 7.1 through 7.3 introduce the normal probability distribution. Section 7.4 presents normal probability plots as a means for assessing normality and is required in order to cover topics presented in Chapters 8–13. Section 7.4 does not require the use of technology because the output generated by technology is presented. This section is necessary in order to help students see that verifying normality is necessary before proceeding with inference for small samples. Section 7.5 introduces sampling distributions. Section 7.6 discusses the normal approximation to the binomial and is optional.

Chapter 8 Confidence Intervals about a Single Parameter

Section 8.1 introduces the construction of a confidence interval when the population standard deviation is known, while Section 8.2 constructs confidence intervals when the population standard deviation is unknown. This approach is different from that in some other texts, but it is logical. It's as simple as "σ known, use z; σ unknown, use t." In both cases, small samples require that the population from which the sample was drawn must be normal–we check this requirement with a normal probability plot. Section 8.3 covers confidence intervals about a population proportion, while Section 8.4 covers confidence intervals about a population standard deviation. Section 8.4 is optional and can be skipped without loss of continuity.

Chapter 9 Hypothesis Testing

Section 9.1 provides an introduction to the language of hypothesis testing. Sections 9.2 and 9.3 test hypotheses regarding a population mean, again segmented by "σ known, use z; σ unknown, use t." Section 9.4 presents hypothesis testing about a population proportion. An interesting feature in this section is that it includes how to use the binomial probability distribution to compute exact P-values. This is especially important if the requirement for using the normal approximation to the binomial is not satisfied. Section 9.5 discusses hypothesis testing about a population standard deviation, while Section 9.6 discusses the power of the test and probability of Type II errors. Both Sections 9.5 and 9.6 are optional.

Chapter 10 Inferences on Two Samples

Section 10.1 presents the analysis required for matched-pairs design. Section 10.2 presents the analysis for comparing two means from independent samples. Notice that the discussion regarding pooled estimates of σ is absent. This is because the pooled estimate approach requires that the two populations have a common variance and this is an extremely difficult requirement to test. Because "pooling" versus "not pooling" often provide the same results, it is not necessary at this level to cloud the students' thought process any further. Section 10.3 discusses comparing two population proportions. Section 10.4 presents the comparison of two variances and is optional.

Chapter 11 Chi-Square Procedures

Section 11.1 presents the chi-square goodness of fit. It presents a discussion of the chi-square distribution for those who skipped Sections 8.4 and 9.5. If you did skip these sections, be sure to introduce students to the chi-square table. Section 11.2 presents contingency tables, marginal distributions, and conditional distributions. If you skipped Section 5.4, you may want to proceed at a slower pace. Section 11.3 discusses chi-square tests for independence and homogeneity. Again, if Sections 5.3 and/or 5.4 were skipped, proceed slowly.

Chapter 12 Inference on the Least-Squares Regression Model; ANOVA

Sections 12.1 and 12.2 are independent of Section 12.3. Sections 12.1 and 12.2 require that Sections 4.1 and 4.2 were covered. Section 12.3 is simply a continuation of the discussion in Section 10.2.

Chapter 13 Nonparametric Statistics

Section 13.2 is on the runs test for randomness and can be covered any time after Section 9.1. Section 13.3 can be presented right after Section 9.3. Section 13.4 can be covered after Section 10.1. Section 13.5 can be covered after Section 10.2. Section 13.6 can be covered after Section 12.2. Section 13.7 can be covered after Section 12.3.

Acknowledgments

Textbooks evolve into their final form through the efforts and contributions of many people. First and foremost, I would like to thank my family, whose dedication to this project was just as much as mine: my wife, Yolanda, whose words of encouragement and support were unabashed, and my children, Michael, Kevin, and Marissa, who would come and greet me every morning with smiles that only children can deliver. I owe each of them my sincerest gratitude.

I would also like to thank the entire Mathematics Department at Joliet Junior College, who provided support, ideas, and encouragement to help me complete this project.

Countless ideas were bounced off of each of them and their responses are present throughout the text. Special thanks also go to the Joliet Junior College community, who also supported this project.

I sincerely appreciate the comments and suggestions of Erica Egizio, who had the courage to use the text in its nascent stages. Her comments and suggestions have proven to be invaluable. Erica is also to be commended for her "Note to the Instructor" annotations. Special thanks to Michael McCraith, who helped locate interesting data and also made the PowerPoint presentations friendly for both PCs and Macs. Thank you to Faye Dang and Priscilla Gathoni for checking all the answers. Thank you to Elena Catoiu, Kathleen Miranda, and Gigi Williams for proofreading the text to help ensure its accuracy. Thanks to Lindsay Packer, who took on the daunting task of writing the solutions manuals; and to Beverly J. Dretzke, Kate McLaughlin, and Dorothy Wakefield, who wrote the technology guides; and David Lane, who wrote the applets workbook. I would also like to thank Max Linn, who wrote the Case Studies at the end of each chapter. A big thank you to the development editors: Frank Purcell of Twin Prime Editorial and Anne Scanlan-Rohrer. I would also like to thank my father, who has taught me the ropes. It is always nice to learn from the best.

There are many colleagues I would like to thank for their input. I owe each of them my sincerest thanks. I apologize for any omissions.

Fusan Akman, *Coastal Carolina Community College*
Grant Alexander, *Joliet Junior College*
Randall Allbritton, *Daytona Beach Community College*
Margaret M. Balachowski, *Michigan Technological University*
Charles Biles, *Humboldt College*
Joanne Brunner, *Joliet Junior College*
Jim Butterbach, *Joliet Junior College*
Roxanne Byrne, *University of Colorado–Denver*
Elena Catoiu, *Joliet Junior College*
Pinyuen Chen, *Syracuse University*
Faye Dang, *Joliet Junior College*
Erica Egizio, *Joliet Junior College*
Richard Einsporn, *The University of Akron*
Franco Fedele, *University of West Florida*
Joe Gallegos, *Salt Lake City Community College*
Frieda Ganter, *California State University–Fresno*
Kim Jones, *Virginia Commonwealth University*
Maryann E. Justinger, *Erie Community College–South Campus*
Iraj Kalantari, *Western Illinois University*
Donna Katula, *Joliet Junior College*
Mohammad Kazemi, *University of North Carolina–Charlotte*
Melody Lee, *Joliet Junior College*
Mike Mays, *West Virginia University*
Jean McArthur, *Joliet Junior College*
Marilyn McCollum, *North Carolina State University*
Michael McCraith, *Joliet Junior College*
Jill McGowan, *Howard University*
Judy Meckley, *Joliet Junior College*
Kathleen Miranda, *SUNY at Old Westbury*
Lindsay Packer, *College of Charleston*
Linda Padilla, *Joliet Junior College*
Ingrid Peterson, *University of Kansas*
Nancy Pevey, *Pellissippi State Technical Community College*
Philip Pina, *Florida Atlantic University*
David Ray, *University of Tennessee–Martin*
Mike Rosenthal, *Florida International University*
Robert Sackett, *Erie Community College–North Campus*
Karen Spike, *University of North Carolina–Wilmington*
Timothy D. Stebbins, *Kalamazoo Valley Community College*
Patrick Stevens, *Joliet Junior College*
Sharon Stokero, *Michigan Technological University*
Cathleen M. Zucco Teveloff, *Trinity College*
Kirk Trigsted, *University of Idaho*
Robert Tuskey, *Joliet Junior College*
Farroll Tim Wright, *University of Missouri–Columbia*

Recognition and thanks are due particularly to the following individuals for their valuable assistance in the preparation of this text: Sally Yagan for her support, interest, and friendship; Patrice Jones for his sage advice and marketing skills; Krista M. Bettino for her innovative marketing efforts; Barbara Mack for her organizational skills as a production supervisor; and the entire Prentice Hall sales staff for their confidence.

Michael Sullivan, III

This proverb applies to anything where one wants to succeed. Would a pianist expect to be the best if she didn't practice? The only way you are going to learn statistics is by doing statistics. When you get problems wrong, ask for help.

4. If you have questions, visit your professor during office hours.

Just like your previous mathematics courses, statistics is a building process. This means the material is used throughout the course. If a topic is not understood, then this lack of understanding could come back to haunt you later in the class. This is the source of a lot of frustration in learning statistics or mathematics. You need to build a strong foundation before you continue to build the house.

How to Use This Book Effectively and Efficiently

First, and foremost, this text was written to be read. You will find that the text has additional explanations and examples that will help you. As mentioned previously, be sure to read the text before attending class.

Many sections begin with "Preparing for This Section," a list of concepts that will be used in the section. Take the short amount of time required to refresh your memory. This will make the section easier to understand and will actually save you time and effort. Each section that has the "Preparing for This Section" feature will have a "Preparing for This Section" quiz on the companion Web site (www.prenhall.com/sullivanstats). The quiz asks questions related to the concepts that are listed. Objectives are provided at the beginning of each section. Read them. They will help you recognize the important ideas and skills developed in the section.

After a concept has been introduced and an example given, you will see **NW** *Now Work Problem xx.* Go to the exercises at the end of the section, work the problem cited, and check your answer in the back of the book. If you get it right, you can be confident in continuing on in the section. If you don't get it right, go back over the explanations and examples to see what you might have missed. Then rework the problem. Ask for help if you miss it again.

From left to right, Michael IV, Michael III, Marissa, Yolanda, and Kevin

I have included an "In Your Own Words" feature that explains definitions, theorems, and concepts using everyday language. This is meant to help you understand the concepts presented in the text. There are also "Caution" statements. These are meant to make you aware of common errors that occur in statistics, so that you don't make these mistakes.

The chapter review contains a list of formulas and vocabulary introduced in the chapter. Be sure you understand how to use the formulas and that you know the definitions of the vocabulary. There is also a list of objectives along with review problems that correspond to the objective. If you can't do the problems listed for a particular objective, go back to the page indicated and review the material.

Please do not hesitate to contact me, through Prentice Hall, with any suggestions or comments that would improve the text. I look forward to hearing from you.

Michael Sullivan, III

A USER'S GUIDE TO THE TEXT

Utilizing the features found throughout this text will help you learn and study the content for this course. Many of these features were developed by the author's students for students.

Putting It All Together

Each chapter opens with a discussion of topics that were covered and how they relate to topics that are about to be discussed. This allows students to see the "big picture" of how the topics relate to each other.

Putting It All Together

In Chapter 1, we learned that statistics is a process. The process begins with asking a research question. In order to determine the answer to the question, information (data) must be collected. The information is obtained from a census, existing data sources, surveys or designed experiments. When data are collected from a survey or designed experiment, they must be organized into a manageable form. Data that are not organized are referred to as **raw data**.

Methods for organizing raw data include the creation of tables or graphs, which allow for a quick overview of the information collected. The organization of data is the third step in the statistical process. The procedures used in organizing data into tables and graphs depend upon whether the data are qualitative, discrete, or continuous.

PAGE 50

Companion Web Site

Many additional free resources for this text can be found at: *www.prenhall.com/sullivanstats*

CHAPTER 2

Organizing and Summarizing Data

Outline

For additional study help, go to www.prenhall.com/sullivanstats

Materials include

- Self-Graded Quizzes
- "Preparing for This Section" Quizzes
- STATLETs
- PowerPoint Downloads
- Step-by-Step Technology Guide
- Graphing Calculator Help

PAGE 51

Preparing for This Section

Most sections open with a referenced list (by section and page number) of key items to review in preparation for the section ahead. This provides a just-in-time review for the students.

Objectives

A numbered list of key objectives appears in the beginning of each section. As the topic corresponding to the objective is addressed, the number appears in the column.

2.1 Organizing Qualitative Data

Preparing for This Section Before getting started, review the following:

✓ Qualitative data (Section 1.1, p. 6)

Objectives

1. Construct frequency and relative frequency distributions from qualitative data
2. Construct bar graphs
3. Construct pie charts

In this section we will concentrate on tabular and graphical summaries of qualitative data. Sections 2.2 and 2.3 discuss methods for summarizing quantitative data.

1 **Organizing Qualitative Data**

Recall that qualitative data provide nonnumerical measures that categorize or classify an individual. When qualitative data are collected, we are often interested in determining the number of individuals that occur within each category.

Definition A **frequency distribution** lists the number of occurrences for each category of data.

PAGE 52

In Your Own Words

When a definition or concept is presented, the "In Your Own Words" feature presents the definition or concept using everyday language while maintaining accuracy.

In Your Own Words

A frequency distribution shows how often each category occurs. A relative frequency distribution shows the percent of the observations that belong in each category.

Definition The **relative frequency** is the proportion or percent of observations within a category and is found using the formula

$$\text{Relative frequency} = \frac{\text{frequency}}{\text{sum of all frequencies}} \quad (1)$$

A **relative frequency distribution** lists the relative frequency of each category of data.

*History considers Cleveland to be two different presidents, even though he is the same individual, because he was not elected to two consecutive terms.

PAGE 53

2 **Bar Graphs**

One of the most common devices for graphically representing qualitative data is a bar graph.

Definition A **bar graph** is constructed by labeling each category of data on a horizontal axis and the frequency or relative frequency of the category on the vertical axis. A rectangle of equal width is drawn for each category. The height of the rectangle is equal to the category's frequency or relative frequency.

▶ EXAMPLE 3 **Constructing a Frequency and Relative Frequency Bar Graph**

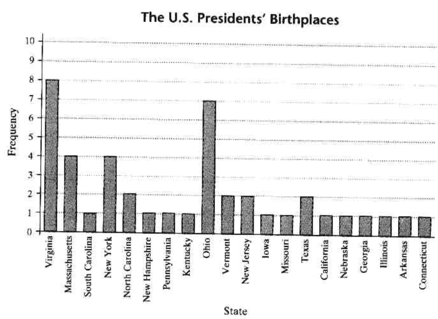
Problem: Use the data summarized in Table 3 to construct (a) a frequency bar graph and (b) a relative frequency bar graph.

Approach: A horizontal axis is used to indicate the category of the data (states, in this case), and a vertical axis is used to represent the frequency or relative frequency (in this case, number of presidential births). Draw bars, or rectangles of equal width for each category whose height is the frequency or relative frequency. The bars do not touch each other.

Solution:

(a) Figure 1 shows the frequency bar graph.

Figure 1



Caution

Whenever constructing bar graphs, be sure to include labels for the axes as well as a title for the graph!

PAGE 55

Caution

This alerts students to some of the pitfalls in statistical analysis.

Historical Notes

For many of the concepts introduced, a short historical note regarding the individual responsible is presented. These historical notes are meant to "humanize" famous statisticians.

Section 2.2 Organizing Quantative Data I **71**

Historical Note

Florence Nightingale was born in Italy on May 12, 1820. She was named after the city of her birth. Florence was educated by her father, who attended Cambridge University. Between 1849 and 1851, Florence studied nursing throughout Europe. In 1854, she was asked to oversee the introduction of female nurses into the military hospitals in Turkey. While there, she greatly improved the mortality rate of wounded soldiers. She collected data and invented graphs (the polar area diagram), tables, and charts to show that improving sanitary conditions would lead to decreased mortality rates. In 1869, Florence founded the Nightingale School Home for Nurses. Florence died on August 13, 1910.

TABLE 13

Class (Three-Year Rate of Return)	Tally	Frequency	Relative Frequency
10.0–14.9	𝍸 II	7	7/40 = 0.175
15.0–19.9	𝍸 𝍸 I	11	11/40 = 0.275
20.0–24.9	𝍸 III	8	8/40 = 0.2
25.0–29.9	𝍸 I	6	0.15
30.0–34.9	III	3	0.075
35.0–39.9	III	3	0.075
40.0–44.9		0	0
45.0–49.9	II	2	0.05

Only 5% of the mutual funds had a three-year rate of return between 45.0% and 49.9%. We might conclude that mutual funds with a rate of return at this level outperform their peers and consider them worthy of our investment. This type of information would be difficult to obtain from the raw data. ◀◀

Notice that the choice of the lower class limit of the first class and the class width was rather arbitrary. While formulas and procedures do exist for creating frequency distributions from raw data, they do not necessarily provide "better" summaries. It is incorrect to say that one particular frequency distribution is the correct one. Constructing frequency distributions is somewhat of an art form in which the distribution that seems to provide the best overall summary of the data should be used.

Consider the frequency distribution in Table 14, which also summarizes the three-year rate of return data discussed in Example 3. Here, the lower class limit of the first class is 10.0 and the class width is 8.0. Do you think Table 13 or Table 14 provides a better summary of the distribution of three-year rates of return? In forming your opinion, consider the following: Too few classes will cause a "bunching" effect. Too many classes will spread the data out, thereby not revealing any pattern.

The goal in constructing a frequency distribution is to reveal interesting features of the data. With that said, when constructing frequency distributions, we typically want the number of classes to be between 5 and 20. When the data set is small, we want fewer classes; when the data set is large, we want more classes. Why do you think this is reasonable?

In Your Own Words

Creating the classes for summarizing continuous data is an art form. There is no such thing as the correct frequency distribution. However, there can be incorrect frequency distributions. The larger the class width, the fewer classes a frequency distribution will have.

TABLE 14

Class	Tally	Frequency
10.0–17.9	𝍸 𝍸 I	11
18.0–25.9	𝍸 𝍸 𝍸 I	16
26.0–33.9	𝍸 III	8
34.0–41.9	III	3
42.0–49.9	II	2

NW *Now Work Problems 19(a) and (b).*

PAGE 71

Examples

Examples are set up in "Problem," "Approach," and "Solution" structure. The "Problem" explains the situation and the question that is being asked. The "Approach" walks through the steps of how to analyze the information. Finally, the "Solution" applies the "Approach" and solves the "Problem."

▶ EXAMPLE 5 **Constructing a Pie Chart**

Problem: The data presented in Table 6 represent the educational attainment of residents of the United States 25 years or older, based upon data obtained from the 2000 United States Census. Construct a pie chart of the data.

TABLE 6

Educational Attainment	2000
Less than 9th grade	12,327,601
9th–12th grade, no diploma	20,343,848
High school diploma	52,395,507
Some college, no degree	36,453,108
Associate's degree	11,487,194
Bachelor's degree	28,603,014
Graduate/professional degree	15,930,061
Totals	**177,540,333**

Approach: The pie chart will have seven parts, or sectors, corresponding to the seven categories of data. The area of each sector is proportional to the frequency of each category. For example,

$$\frac{12{,}327{,}601}{177{,}540{,}333} = 0.0694 = 6.94\%$$

of all U.S. residents 25 years or older have less than a 9th-grade education. The category "less than 9th grade" will make up 6.94% of the pie chart. Since a circle has 360°, the degree measure of the sector for the category "less than 9th-grade education" will be $(0.0694)360° \approx 25.0°$. Use a protractor to measure each angle.

Solution: We follow the approach presented for the remaining categories of data to obtain Table 7.

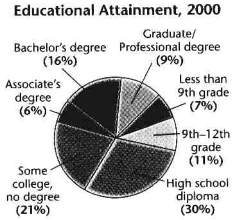
TABLE 7

Education	Frequency	Relative Frequency	Degree Measure of Each Sector
Less than 9th grade	12,327,601	0.0694	25.0
9th to 12th grade, no diploma	20,343,848	0.1146	41.3
High school diploma	52,395,507	0.2951	106.2
Some college, no degree	36,453,108	0.2053	73.9
Associate's degree	11,487,194	0.0647	23.3
Bachelor's degree	28,603,014	0.1611	58.0
Graduate/professional degree	15,930,061	0.0897	32.3

PAGE 59

Figure 6

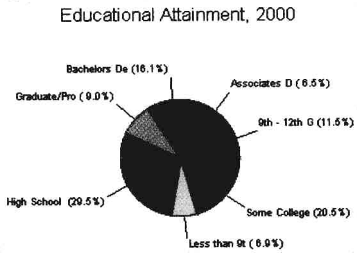
Educational Attainment, 2000



To construct a pie chart by hand, we use a protractor to approximate the angles for each sector. See Figure 6. ◀◀

Using Technology: Certain statistical spreadsheets have the ability to draw pie charts. Figure 7 shows the pie chart of the data in Table 6, using Minitab.

Figure 7



Pie charts can be created only if all the categories of the variable under consideration are represented. In other words, we could create a bar chart that lists the proportion of presidents born in the states of Virginia, Texas, and New York, but it would not make sense to construct a pie chart for this situation. Do you see why? Only 34.88% of the pie would be accounted for.

NW *Now Work Problem 19(e).*

PAGE 60

Using Technology

Immediately following an example whenever appropriate, the output from a TI-83 Plus, MINITAB, or Excel is presented.

Now Work Problem xxx

A NW icon NW along with a corresponding problem to be solved appears after a concept has been introduced. This directs the student to a problem in the exercises that tests the concept, ensuring that the concept has been mastered before moving on. The Now Work problems are identified in the exercises using orange numbers and a NW icon.