Elementary Statistics

A STEP BY STEP APPROACH

Ninth Edition

ALLAN G. BLUMAN

Elementary Statistics

A STEP BY STEP APPROACH

Ninth Edition

McGraw Hill Education

ALLAN G. BLUMAN
PROFESSOR EMERITUS
COMMUNITY COLLEGE OF ALLEGHENY COUNTY

ELEMENTARY STATISTICS: A STEP BY STEP APPROACH, NINTH EDITION

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 DOW/DOW 1 0 9 8 7 6 5 4

ISBN 978–0–07–353498–5
MHID 0–07–353498–6

ISBN 978–0–07–766572–2 (Annotated Instructor's Edition)
MHID 0–07–766572–4

Senior Vice President, Products & Markets: *Kurt L. Strand*
Vice President, General Manager, Products & Markets: *Marty Lange*
Vice President, Content Production & Technology Services: *Kimberly Meriwether David*
Managing Director: *Ryan Blankenship*
Brand Manager: *Holly Rhodes*
Director of Development: *Rose Koos*
Development Editor: *Ashley Zellmer McFadden*
Director of Digital Content: *Rob Brieler, Ph.D.*
Digital Development Editor: *Adam Fischer*
Digital Product Analyst: *Michael Lemke*
Director of Marketing: *Alex Gay*
Director, Content Production: *Terri Schiesl*
Content Project Manager: *Vicki Krug*
Senior Buyer: *Sandy Ludovissy*
Design: *Matthew Backhaus*
Cover Photo: *Image 1: Lonely Planet Images/Bob Charlton/Getty Images; Image 2: Stuart Dee/Getty Images*
Senior Content Licensing Specialist: *Lori Hancock*
Compositor: *MPS Limited*
Typeface: *10.5/12 Times Roman*
Printer: *R. R. Donnelley*

Library of Congress Cataloging-in-Publication Data

Bluman, Allan G., author.
Elementary statistics : a step by step approach / Allan G. Bluman, professor emeritus, Community College of Allegheny County. – Ninth edition.
pages cm
Includes index.
ISBN 978–0–07–353498–5 — ISBN 0–07–353498–6 (hard copy : alk. paper) — ISBN 978–0–07–766572–2 — ISBN 0–07–766572–4 (annotated instructor's edition : hard copy) 1. Statistics–Textbooks. I. Title.
QA276.12.B59 2014
519.5–dc23
2013031502

www.mhhe.com

ABOUT THE AUTHOR

Allan G. Bluman

Allan G. Bluman is a professor emeritus at the Community College of Allegheny County, South Campus, near Pittsburgh, Pennsylvania. He has taught mathematics and statistics for over 35 years. He received an Apple for the Teacher award in recognition of his bringing excellence to the learning environment at South Campus. He has also taught statistics for Penn State University at the Greater Allegheny (McKeesport) Campus and at the Monroeville Center. He received his master's and doctor's degrees from the University of Pittsburgh.

He is also author of *Elementary Statistics: A Brief Version* and coauthor of *Math in Our World*. In addition, he is the author of four mathematics books in the McGraw-Hill DeMystified Series. They are *Pre-Algebra, Math Word Problems, Business Math,* and *Probability*.

He is married and has two sons, a granddaughter, and a grandson.

Dedication: *To Betty Bluman, Earl McPeek, and Dr. G. Bradley Seager, Jr.*

This page intentionally left blank

CONTENTS

All examples and exercises in this textbook (unless cited) are hypothetical and are presented to enable students to achieve a basic understanding of the statistical concepts explained. These examples and exercises should not be used in lieu of medical, psychological, or other professional advice. Neither the author nor the publisher shall be held responsible for any misuse of the information presented in this textbook.

CHAPTER 4

Probability and Counting Rules *185*

CHAPTER 5

Discrete Probability Distributions *257*

CHAPTER 6

The Normal Distribution *311*

CHAPTER 7

Confidence Intervals and Sample Size *369*

CHAPTER 12

Analysis of Variance 647

CHAPTER 13

Nonparametric Statistics 689

CHAPTER 14

Sampling and Simulation 741

APPENDICES

ADDITIONAL TOPICS ONLINE

(www.mhhe.com/bluman)

PREFACE

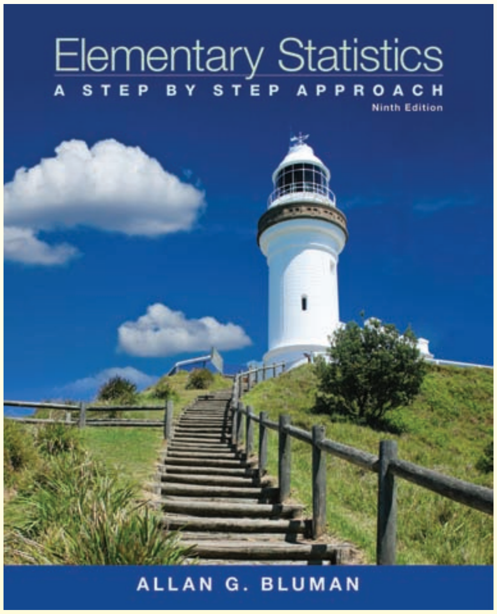
Approach

Elementary Statistics: A Step by Step Approach was written as an aid in the beginning statistics course to students whose mathematical background is limited to basic algebra. The book follows a nontheoretical approach without formal proofs, explaining concepts intuitively and supporting them with abundant examples. The applications span a broad range of topics certain to appeal to the interests of students of diverse backgrounds, and they include problems in business, sports, health, architecture, education, entertainment, political science, psychology, history, criminal justice, the environment, transportation, physical sciences, demographics, eating habits, and travel and leisure.

About This Book

While a number of important changes have been made in the ninth edition, the learning system remains untouched and provides students with a useful framework in which to learn and apply concepts. Some of the retained features include the following:

- Over **1800 exercises** are located at the end of major sections within each chapter.
- **Hypothesis-Testing Summaries** are found at the end of Chapter 9 (z, t, χ^2, and F tests for testing means, proportions, and variances), Chapter 12 (correlation, chi-square, and ANOVA), and Chapter 13 (nonparametric tests) to show students the different types of hypotheses and the types of tests to use.
- A **Data Bank** listing various attributes (educational level, cholesterol level, gender, etc.) for 100 people and several additional data sets using real data are included and referenced in various exercises and projects throughout the book.
- An updated **reference card** containing the formulas and the z, t, χ^2, and PPMC tables is included with this textbook.
- End-of-chapter **Summaries, Important Terms,** and **Important Formulas** give students a concise summary of the chapter topics and provide a good source for quiz or test preparation.
- **Review Exercises** are found at the end of each chapter.



- Special sections called **Data Analysis** require students to work with a data set to perform various statistical tests or procedures and then summarize the results. The data are included in the Data Bank in Appendix B and can be downloaded from the book's website at **www.mhhe.com/bluman**.
- **Chapter Quizzes,** found at the end of each chapter, include multiple-choice, true/false, and completion questions along with exercises to test students' knowledge and comprehension of chapter content.
- The **Appendixes** provide students with extensive reference tables, a glossary, and answers to all quiz questions and odd-numbered exercises. New to this edition, the additional Online Appendixes include algebra review, an outline for report writing, Bayes' theorem, and an alternative method for using the standard normal distribution. These can be found at **www.mhhe.com/bluman**.
- The **Applying the Concepts** feature is included in all sections and gives students an opportunity to think about the new concepts and apply them to examples and scenarios similar to those found in newspapers, magazines, and radio and television news programs.

Changes in the Ninth Edition

Global Changes

- Updated and redesigned the Technology Boxes to reflect the latest technology updates
- Over 60 new examples and more than 400 new exercises have been added or revised throughout the book
- Design of graphs, photos and art has been revised for clarity
- New interior design was incorporated for ease of reading and flow
- Matched the odd and even exercises

Chapter 1 Nearly 100 new section exercises have been added
Statistics Today updated
Added new material on sampling and the three types of observational studies

Chapter 2 New subsection on dotplots added
More material on class boundaries added
Statistics Today updated

Chapter 3 New Procedure Table for median added
New Procedure Table for variance and standard deviation added
Summary for Chebyshev's Theorem added
New procedure Table for boxplots included

Chapter 4 New subsection for permutations with identical objects added

Chapter 5 New subsection on the geometric probability distribution added

Chapter 6 Improved flow with a shortened introduction
Added New Procedure Table for finding data values for specific probabilities

Chapter 7 Statistics Today updated
Included summary of the Characteristics of the Chi-Square Distribution

Chapter 8 Reorganized Section 8-1 for a better flow
Simplified diagram on stating the conclusion from a hypothesis test

Chapter 9 Expanded explanation of independent samples
Added 20 new exercises

Chapter 10 Statistics Today updated
Revised and Shortened Introduction
Added additional information on analyzing scatter plots
Included six new Procedure Tables

Chapter 11 Expanded explanation on finding expected frequencies

Chapter 12 Added additional material on the Computation of the F test
Added two new figures

Chapter 13 Boxed definitions and listed assumptions for the nonparametric tests
Added a new Procedure Table for finding and testing the significance of the Spearman's Rank Correlation Coefficient
Added a new Procedure Table for the runs test

Chapter 14 Added a formal definition for convenience sample
Improved definitions for systematic sampling and stratified sampling

Acknowledgments

It is important to acknowledge the many people whose contributions have gone into the Ninth Edition of *Elementary Statistics.* Very special thanks are due to Jackie Miller of the University of Michigan for her provision of the Index of Applications, her exhaustive accuracy check of the page proofs, and her general availability and advice concerning all matters statistical. The Technology Step by Step sections were provided by Gerry Moultine of Northwood University (MINITAB), John Thomas of College of Lake County (Excel), and William Vezko of Saint Johns River State College-Orange Park (TI-84 Plus).

I would also like to thank Diane P. Cope for providing the new exercises; Kelly Jackson for writing the new Data Projects; Lisa Collette, developmental copyeditor, for her thoughtful suggestions; Phyllis Barnidge for her error checking; and Sally Robinson for error checking, adding technology-accurate answers to Appendix E, and writing the Solutions Manuals.

Finally, at McGraw-Hill Education, thanks to Ryan Blankenship, Managing Director; Holly Rhodes, Brand Manager; Ashley Zellmer McFadden, Developmental Editor; Alex Gay, Marketing Director; Rob Brieler, Director of Digital Content; and Vicki Krug, Content Project Manager.

—Allan G. Bluman

Special thanks for their advice and recommendations for the Ninth Edition go to:

Wesley Anderson, *Northwest Vista College*
Elizabeth Betzel, *Columbus State Community College*
David Bosworth, *Hutchinson Community College*
Christine M. Brady, *Suffolk County Community College*
Gregory Cianflone, *Suffolk County Community College*
Lisa M. Crawford, *Piedmont Technical College*
Hemangini Deshmukh, *Mercyhurst College*
Angela Everett, *Chattanooga State Community College*
Robert Feinerman, *Lehman College (CUNY)*
Kathi Fields, *Blue Ridge Community College*
Maggie Flint, *Northeast State Community College*
Jesus Gutierrez, *Citrus College*
Jaffar Ali Shahul Hameed, *Florida Gulf Coast University*
Kelly Jackson, *Camden County College*
Mohammad Kazemi, *University of North Carolina Charlotte*
Brian Kelly, *Bryant University*
Patrick Kelly, *Mercyhurst College*
Jong Kim, *Portland State University*
James E. Martin, *Christopher Newport University*
Angie Schirck Matthews, *Broward College*
Carla Monticelli, *Camden County College*
Keith Oberlander, *Pasadena City College*
Dr. Elaine H. Olaoye, *Brookdale Community College*
Sherri Rankin, *Hutchinson Community College*
Fred Rispoli, *Dowling College*
Pam Stogsdill, *Bossier Parish Community College*
Fernando Rincon Tellez, *Piedmont Technical College*
Malissa Trent, *Northeast State Community College*
Barbara Villatoro, *Diablo Valley College*
Cassandra Vincent, *State University of New York College at Plattsburgh*
Henry Wakhungu, *Indiana University*
Jane West, *Trident Technical College*
Bashar Zogheib, *American University of Kuwait*

A STEP BY STEP APPROACH

Each chapter begins with an **outline,** a **list of learning objectives,** and a feature titled **Statistics Today;** in which a real-life problem shows students the relevance of the material. This problem is solved near the end of the chapter using statistical techniques presented in the chapter.

7

Confidence Intervals and Sample Size

STATISTICS TODAY

Stress and the College Student

A recent poll conducted by the mtvU/Associated Press found that 85% of college students reported that they experience stress daily. The study said, "It is clear that being stressed is a fact of life on college campuses today."

The study also reports that 74% of students' stress comes from school work, 71% from grades, and 62% from financial woes. The report stated that 2240 undergraduate students were selected and that the poll has a margin of error of ±3.0%.

In this chapter you will learn how to make a true estimate of a parameter, what is meant by the margin of error, and whether or not the sample size was large enough to represent all college students.

See Statistics Today—Revisited at the end of this chapter for more details.

OUTLINE

OBJECTIVES

After completing this chapter, you should be able to

1. Find the confidence interval for the mean when σ is known.
2. Determine the minimum sample size for finding a confidence interval for the mean.
3. Find the confidence interval for the mean when σ is unknown.
4. Find the confidence interval for a proportion.
5. Determine the minimum sample size for finding a confidence interval for a proportion.
6. Find a confidence interval for a variance and a standard deviation.

EXAMPLE 8–6 Cost of College Tuition

A researcher wishes to test the claim that the average cost of tuition and fees at a four-year public college is greater than $5700. She selects a random sample of 36 four-year public colleges and finds the mean to be $5950. The population standard deviation is $659. Is there evidence to support the claim at $\alpha = 0.05$? Use the *P*-value method.

Source: Based on information from the College Board.

SOLUTION

Step 1 State the hypotheses and identify the claim.

H_0: $\mu = \$5700$ and H_1: $\mu > \$5700$ (claim).

Step 2 Compute the test value.

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{5950 - 5700}{659/\sqrt{36}} = 2.28$$

Step 3 Find the *P*-value. Using Table E in Appendix A, find the corresponding area under the normal distribution for $z = 2.28$. It is 0.9887. Subtract this value for the area from 1.0000 to find the area in the right tail.

$$1.0000 - 0.9887 = 0.0113$$

Hence, the *P*-value is 0.0113.

Step 4 Make the decision. Since the *P*-value is less than 0.05, the decision is to reject the null hypothesis. See Figure 8–17.

Hundreds of examples with detailed solutions serve as models to help students solve problems on their own. Examples are solved by using a step by step explanation, and illustrations provide a clear display of results.

Procedure Table

Solving Hypothesis-Testing Problems (Traditional Method)

Step 1 State the hypotheses and identify the claim.

Step 2 Find the critical value(s) from the appropriate table in Appendix A.

Step 3 Compute the test value.

Step 4 Make the decision to reject or not reject the null hypothesis.

Step 5 Summarize the results.

Numerous **Procedure Tables** summarize processes for students' quick reference.

Critical Thinking sections at the end of each chapter challenge students to apply what they have learned to new situations while deepening conceptual understanding.

Critical Thinking Challenges

The power of a test $(1 - \beta)$ can be calculated when a specific value of the mean is hypothesized in the alternative hypothesis; for example, let H_0: $\mu = 50$ and let H_1: $\mu = 52$. To find the power of a test, it is necessary to find the value of β. This can be done by the following steps:

Step 1 For a specific value of α find the corresponding value of $\overline{X}$, using $z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$, where μ is the hypothesized value given in H_0. Use a right-tailed test.

Step 2 Using the value of $\overline{X}$ found in step 1 and the value of μ in the alternative hypothesis, find the area corresponding to z in the formula $z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$.

Step 3 Subtract this area from 0.5000. This is the value of β.

Step 4 Subtract the value of β from 1. This will give you the power of a test. See Figure 8–41.

Technology Step by Step boxes instruct students how to use Excel, TI-84 Plus graphing calculators, and MINITAB to solve the types of problems covered in the section. Numerous computer or calculator screens are displayed as well as numbered steps.

Technology Step by Step

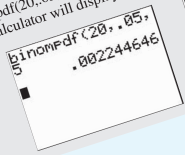
TI-84 Plus Step by Step

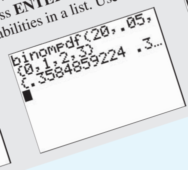
Binomial Random Variables

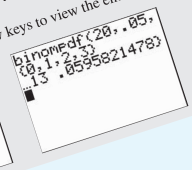
To find the probability for a binomial variable:
Press **2nd [DISTR]** then **A (ALPHA MATH)** for binompdf.
The form is binompdf(n,p,X).

Example: $n = 20$, $X = 5$, $p = .05$ (Example 5–20a from the text)
binompdf(20,.05,5), then press **ENTER** for the probability.

Example: $n = 20$, $X = 0, 1, 2, 3$, $p = .05$ (Example 5–20b from the text).
binompdf(20,.05,{0,1,2,3}), then press **ENTER**.
The calculator will display the probabilities in a list. Use the arrow keys to view the entire display.







Applying the Concepts are end-of-section exercises that reinforce the concepts explained in the section. They give students an opportunity to think about the concepts and apply them to hypothetical examples similar to real-life ones.

Applying the Concepts 4–5

Counting Rules and Probability

One of the biggest problems for students when doing probability problems is to decide which formula or formulas to use. Another problem is to decide whether two events are independent or dependent. Use the following problem to help develop a better understanding of these concepts.

Assume you are given a five-question multiple-choice quiz. Each question has 5 possible answers: A, B, C, D, and E.

1. How many events are there?
2. Are the events independent or dependent?
3. If you guess at each question, what is the probability that you get all of them correct?
4. What is the probability that a person guesses answer A for each question?

Data Projects, which appear at the end of each chapter, further challenge students' understanding and application of the material presented in the chapter. Many of these require the student to gather, analyze, and report on real data.

Data Projects

1. **Business and Finance** Select a pizza restaurant and a sandwich shop. For the pizza restaurant look at the menu to determine how many sizes, crust types, and toppings are available. How many different pizza types are possible? For the sandwich shop determine how many breads, meats, veggies, cheeses, sauces, and condiments are available. How many different sandwich choices are possible?
2. **Sports and Leisure** When poker games are shown on television, there are often percentages displayed that show how likely it is that a certain hand will win. Investigate how these percentages are determined. Show an example with two competing hands in a Texas Hold 'Em game. Include the percentages that each hand will win after the deal, the flop, the turn, and the river.
3. **Technology** A music player or music organization program can keep track of how many different artists are in a library. First note how many different artists are in your music library. Then find the probability that if 25 songs are selected at random, none will have the same artist.
4. **Health and Wellness** Assume that the gender distribution of babies is such that one-half the time females are born and one-half the time males are born. In a family of 3 children, what is the probability that all are girls? In a family of 4? Is it unusual that in a family with 4 children all would be girls? In a family of 5?
5. **Politics and Economics** Consider the U.S. Senate. Find out about the composition of any three of the Senate's standing committees. How many different committees of Senators are possible, knowing the party composition of the Senate and the number of committee members from each party for each committee?
6. **Your Class** Research the famous Monty Hall probability problem. Conduct a simulation of the Monty Hall problem online using a simulation program or in class using live "contestants." After 50 simulations compare your results to those stated in the research you did. Did your simulation support the conclusions?

SUPPLEMENTS

McGraw-Hill conducted in-depth research to create a new learning experience that meets the needs of students and instructors today. The result is a reinvented learning experience rich in information, visually engaging, and easily accessible to both instructors and students.

- McGraw-Hill's Connect is a Web-based assignment and assessment platform that helps students connect to their coursework and prepares them to succeed in and beyond the course.
- Connect enables math and statistics instructors to create and share courses and assignments with colleagues and adjuncts with only a few clicks of the mouse. All exercises, learning objectives, and activities are vetted and developed by math instructors to ensure consistency between the textbook and the online tools.
- Connect also links students to an interactive eBook with access to a variety of media assets and a place to study, highlight, and keep track of class notes.

To learn more, contact your sales rep or visit **www.connectmath.com**.

ALEKS is a Web-based program that uses artificial intelligence to assess a student's knowledge and provide personalized instruction on the exact topics the student is most ready to learn. By providing individualized assessment and learning, ALEKS helps students to master course content quickly and easily. ALEKS allows students to easily move between explanations and practice, and it provides intuitive feedback to help students correct and analyze errors. ALEKS also includes a powerful instructor module that simplifies course management so instructors spend less time with administrative tasks and more time directing student learning.

ALEKS 360 is a new cost-effective total course solution: fully integrated, interactive eBook, including lecture and exercise videos tied to the textbook, combined with ALEKS personalized assessment and learning.

To learn more about ALEKS and ALEKS 360, contact your sales rep or visit **www.aleks.com**.

Give students the tools they need. Right when they need them. McGraw-Hill LearnSmart adaptively assesses students' skill levels to determine which topics students have mastered and which require further practice. Then it delivers customized learning content based on their strengths and weaknesses. The result: students learn faster and more efficiently because they get the help they need, right when they need it—instead of getting stuck on lessons or being continually frustrated with stalled progress.

SmartBook is the first and only adaptive reading experience available for the higher education market. Powered by the intelligent and adaptive LearnSmart engine, SmartBook facilitates the reading process by identifying what content a student knows and doesn't know. As a student reads, the material continuously adapts to ensure the student is focused on the content he or she needs the most to close specific knowledge gaps.

With **McGraw-Hill Create™,** you can easily rearrange chapters, combine material from other content sources, and quickly upload content you have written such as your course syllabus or teaching notes. Find the content you need in Create by searching through thousands of leading McGraw-Hill textbooks. Arrange your book to fit your teaching style. Create even allows you to personalize your book's appearance by selecting the cover and adding your name, school, and course information. Assemble a Create book, and you'll receive a complimentary print review copy in 3–5 business days or a complimentary electronic review copy (eComp) via email in minutes. Go to **www.mcgrawhillcreate.com** today and experience how McGraw-Hill Create™ empowers you to teach your students your way.

Instructor's Testing and Resource Online

This computerized test bank, available online to adopting instructors, utilizes TestGen® cross-platform test generation software to quickly and easily create customized exams. Using hundreds of test items taken directly from the text, TestGen allows rapid test creation and flexibility for instructors to create their own questions from scratch with the ability to randomize number values. Powerful search and sort functions help quickly locate questions and arrange them in any order, and built-in mathematical templates let instructors insert stylized text, symbols, graphics, and equations directly into questions without need for a separate equation editor.

MegaStat®

MegaStat® is a statistical add-in for Microsoft Excel, handcrafted by J. B. Orris of Butler University. When MegaStat is installed, it appears as a menu item on the Excel menu bar and allows you to perform statistical analysis on data in an Excel workbook. The MegaStat plug-in can be purchased at **www.mhhe.com/megastat**.

MINITAB Student Release 14

The student version of MINITAB statistical software is available with copies of the text. Ask your McGraw-Hill representative for details.

SPSS Student Version for Windows

A student version of SPSS statistical software is available with copies of this text. Consult your McGraw-Hill representative for details.

MINITAB 14 Manual

This manual provides the student with how-to information on data and file management, conducting various statistical analyses, and creating presentation-style graphics while following each text chapter.

TI-84 Plus Graphing Calculator Manual

This friendly, practical manual teaches students to learn about statistics and solve problems by using these calculators while following each text chapter.

Excel Manual

This workbook, specially designed to accompany the text, provides additional practice in applying the chapter concepts while using Excel.

Instructor's Solutions Manual (instructors only)

By Sally Robinson of South Plains College, this manual includes worked-out solutions to all the exercises in the text and answers to all quiz questions. This manual can be found online at **www.mhhe.com/bluman**.

Student's Solutions Manual

By Sally Robinson of South Plains College, this manual contains detailed solutions to all odd-numbered text problems and answers to all quiz questions.

INDEX OF APPLICATIONS

CHAPTER 4

Probability and Counting Rules

CHAPTER 5
Discrete Probability Distributions

CHAPTER 6
The Normal Distribution

CHAPTER 7
Confidence Intervals and Sample Size

CHAPTER 8
Hypothesis Testing

CHAPTER 9

Testing the Difference Between Two Means, Two Proportions, and Two Variances

CHAPTER 10

Correlation and Regression

CHAPTER 11
Other Chi-Square Tests

CHAPTER 12
Analysis of Variance

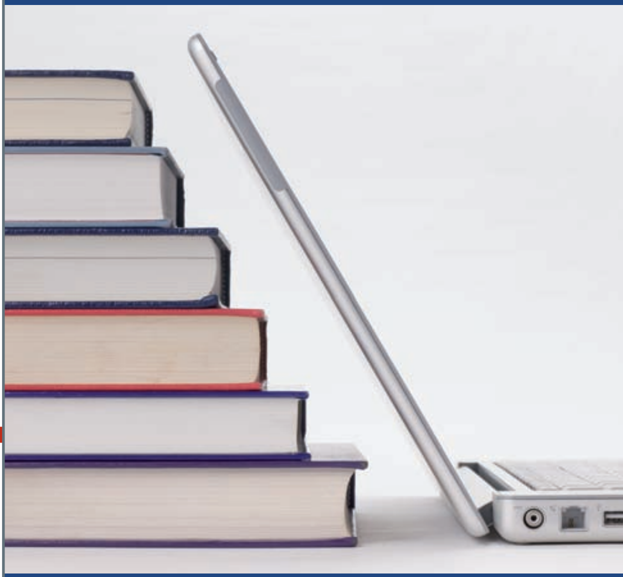
CHAPTER 13
Nonparametric Statistics

CHAPTER 14
Sampling and Simulation

This page intentionally left blank

1

The Nature of Probability and Statistics



STATISTICS TODAY

Is Higher Education "Going Digital"?

Today many students take college courses online and use eBooks. Also, many students use a laptop, smartphone, or computer tablet in the classroom. With the increased use of technology, some questions about the effectiveness of this technology have been raised. For example,

- How many colleges and universities offer online courses?
- Do students feel that the online courses are equal in value to the traditional classroom presentations?
- Approximately how many students take online courses now?
- Will the number of students who take online courses increase in the future?
- Has plagiarism increased since the advent of computers and the Internet?
- Do laptops, smartphones, and tablets belong in the classroom?
- Have colleges established any guidelines for the use of laptops, smartphones, and tablets?

To answer these questions, Pew Research Center conducted a study of college graduates and college presidents in 2011. The procedures they used and the results of the study are explained in this chapter. See Statistics Today—Revisited at the end of the chapter.

OUTLINE

OBJECTIVES

After completing this chapter, you should be able to

1. Demonstrate knowledge of statistical terms.
2. Differentiate between the two branches of statistics.
3. Identify types of data.
4. Identify the measurement level for each variable.
5. Identify the four basic sampling techniques.
6. Explain the difference between an observational and an experimental study.
7. Explain how statistics can be used and misused.
8. Explain the importance of computers and calculators in statistics.

Introduction

You may be familiar with probability and statistics through radio, television, newspapers, and magazines. For example, you may have read statements like the following found in newspapers.

- The FBI reported that violent crimes were down by 6.4% in 2011.
- *USA TODAY* reported that the average college graduate student loan debt was about $19,000.
- The College Stress and Mental Illness Poll reported that 85% of college and university students reported feeling stress daily; 77% reported stress from school work, and 74% experienced stress from grades.
- The *Occupational Outlook Handbook* reported that the median hourly wage for a registered nurse was $31.10 per hour.
- *Reader's Digest* reported that the average cost of using a plasma television is $0.1152 per hour.
- In 2013, the number of sales of smartphones is estimated to be 832.5 million units globally.

Unusual Stats

Of people in the United States, 14% said that they feel happiest in June, and 14% said that they feel happiest in December.

Interesting Fact

Every day in the United States about 120 golfers claim that they made a hole-in-one.

Statistics is used in almost all fields of human endeavor. In sports, for example, a statistician may keep records of the number of yards a running back gains during a football game, or the number of hits a baseball player gets in a season. In other areas, such as public health, an administrator might be concerned with the number of residents who contract a new strain of flu virus during a certain year. In education, a researcher might want to know if new methods of teaching are better than old ones. These are only a few examples of how statistics can be used in various occupations.

Furthermore, statistics is used to analyze the results of surveys and as a tool in scientific research to make decisions based on controlled experiments. Other uses of statistics include operations research, quality control, estimation, and prediction.

Statistics is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.

Historical Note

A Scottish landowner and president of the Board of Agriculture, Sir John Sinclair introduced the word *statistics* into the English language in the 1798 publication of his book on a statistical account of Scotland. The word *statistics* is derived from the Latin word *status,* which is loosely defined as a statesman.

Students study statistics for several reasons:

1. Like professional people, you must be able to read and understand the various statistical studies performed in your fields. To have this understanding, you must be knowledgeable about the vocabulary, symbols, concepts, and statistical procedures used in these studies.
2. You may be called on to conduct research in your field, since statistical procedures are basic to research. To accomplish this, you must be able to design experiments; collect, organize, analyze, and summarize data; and possibly make reliable predictions or forecasts for future use. You must also be able to communicate the results of the study in your own words.
3. You can also use the knowledge gained from studying statistics to become better consumers and citizens. For example, you can make intelligent decisions about what products to purchase based on consumer studies, about government spending based on utilization studies, and so on.

These reasons can be considered some of the goals for studying statistics.

It is the purpose of this chapter to introduce the goals for studying statistics by answering questions such as the following:

What are the branches of statistics?

What are data?

How are samples selected?

1–1 Descriptive and Inferential Statistics

OBJECTIVE 1

Demonstrate knowledge of statistical terms.

To gain knowledge about seemingly haphazard situations, statisticians collect information for *variables,* which describe the situation.

A **variable** is a characteristic or attribute that can assume different values.

Data are the values (measurements or observations) that the variables can assume. Variables whose values are determined by chance are called **random variables.**

Suppose that an insurance company studies its records over the past several years and determines that, on average, 3 out of every 100 automobiles the company insured were involved in accidents during a 1-year period. Although there is no way to predict the specific automobiles that will be involved in an accident (random occurrence), the company can adjust its rates accordingly, since the company knows the general pattern over the long run. (That is, on average, 3% of the insured automobiles will be involved in an accident each year.)

A collection of data values forms a **data set.** Each value in the data set is called a **data value** or a **datum.**

In statistics it is important to distinguish between a sample and a population.

Historical Note

The 1880 Census had so many questions on it that it took 10 years to publish the results.

Historical Note

The origin of descriptive statistics can be traced to data collection methods used in censuses taken by the Babylonians and Egyptians between 4500 and 3000 B.C. In addition, the Roman Emperor Augustus (27 B.C.—A.D. 17) conducted surveys on births and deaths of the citizens of the empire, as well as the number of livestock each owned and the crops each citizen harvested yearly.

A **population** consists of all subjects (human or otherwise) that are being studied.

When data are collected from every subject in the population, it is called a *census*.

For example, every 10 years the United States conducts a census. The primary purpose of this census is to determine the apportionment of the seats in the House of Representatives.

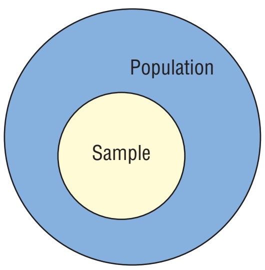
The first census was conducted in 1790 and was mandated by Article 1, Section 2 of the Constitution. As the United States grew, the scope of the census also grew. Today the Census limits questions to populations, housing, manufacturing, agriculture, and mortality. The Census is conducted by the Bureau of the Census, which is part of the Department of Commerce.

Most of the time, due to the expense, time, size of population, medical concerns, etc., it is not possible to use the entire population for a statistical study; therefore, researchers use samples.

A **sample** is a group of subjects selected from a population.

If the subjects of a sample are properly selected, most of the time they should possess the same or similar characteristics as the subjects in the population. See Figure 1–1.

FIGURE 1–1
Population and Sample



OBJECTIVE 2

Differentiate between the two branches of statistics.

However, the information obtained from a statistical sample is said to be *biased* if the results from the sample of a population are radically different from the results of a census of the population. Also, a sample is said to be biased if it does not represent the population from which it has been selected. The techniques used to properly select a sample are explained in Section 1–3.

The body of knowledge called statistics is sometimes divided into two main areas, depending on how data are used. The two areas are

1. Descriptive statistics
2. Inferential statistics

Descriptive statistics consists of the collection, organization, summarization, and presentation of data.

In *descriptive statistics* the statistician tries to describe a situation. Consider the national census conducted by the U.S. government every 10 years. Results of this census give you the average age, income, and other characteristics of the U.S. population. To obtain this information, the Census Bureau must have some means to collect relevant data. Once data are collected, the bureau must organize and summarize them. Finally, the bureau needs a means of presenting the data in some meaningful form, such as charts, graphs, or tables.

The second area of statistics is called *inferential statistics.*

Inferential statistics consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships among variables, and making predictions.

Historical Note

Inferential statistics originated in the 1600s, when John Graunt published his book on population growth, *Natural and Political Observations Made upon the Bills of Mortality.* About the same time, another mathematician/astronomer, Edmond Halley, published the first complete mortality tables. (Insurance companies use mortality tables to determine life insurance rates.)

Unusual Stat

Twenty-nine percent of Americans want their boss's job.

Here, the statistician tries to make inferences from *samples* to *populations.* Inferential statistics uses **probability,** i.e., the chance of an event occurring. You may be familiar with the concepts of probability through various forms of gambling. If you play cards, dice, bingo, or lotteries, you win or lose according to the laws of probability. Probability theory is also used in the insurance industry and other areas.

The area of inferential statistics called **hypothesis testing** is a decision-making process for evaluating claims about a population, based on information obtained from samples. For example, a researcher may wish to know if a new drug will reduce the number of heart attacks in men over age 70 years of age. For this study, two groups of men over age 70 would be selected. One group would be given the drug, and the other would be given a placebo (a substance with no medical benefits or harm). Later, the number of heart attacks occurring in each group of men would be counted, a statistical test would be run, and a decision would be made about the effectiveness of the drug.

Statisticians also use statistics to determine *relationships* among variables. For example, relationships were the focus of the most noted study in the 20th century, "Smoking and Health," published by the Surgeon General of the United States in 1964. He stated that after reviewing and evaluating the data, his group found a definite relationship between smoking and lung cancer. He did not say that cigarette smoking actually causes lung cancer, but that there is a relationship between smoking and lung cancer. This conclusion was based on a study done in 1958 by Hammond and Horn. In this study, 187,783 men were observed over a period of 45 months. The death rate from lung cancer in this group of volunteers was 10 times as great for smokers as for nonsmokers.

Finally, by studying past and present data and conditions, statisticians try to make predictions based on this information. For example, a car dealer may look at past sales records for a specific month to decide what types of automobiles and how many of each type to order for that month next year.

EXAMPLE 1–1 Descriptive or Inferential Statistics

Determine whether descriptive or inferential statistics were used.

a. The average jackpot for the top five lottery winners was $367.6 million.

b. A study done by the American Academy of Neurology suggests that older people who had a high caloric diet more than doubled their risk of memory loss.

c. Based on a survey of 9317 consumers done by the National Retail Federation, the average amount that consumers spent on Valentine's Day in 2011 was $116.

d. Scientists at the University of Oxford in England found that a good laugh significantly raises a person's pain level tolerance.

SOLUTION

a. Descriptive statistics were used because this is an average, and it is based on data obtained from the top five lottery winners at this time.

b. Inferential statistics were used since this is a generalization made from a sample to a population.

c. Descriptive statistics were used since this is an average based on a sample of 9317 respondents.

d. Inferential statistics were used since an inference is made from a sample to a population.

Applying the Concepts 1–1

Attendance and Grades

Read the following on attendance and grades, and answer the questions.

A study conducted at Manatee Community College revealed that students who attended class 95 to 100% of the time usually received an A in the class. Students who attended class 80 to 90% of the time usually received a B or C in the class. Students who attended class less than 80% of the time usually received a D or an F or eventually withdrew from the class.

Based on this information, attendance and grades are related. The more you attend class, the more likely it is you will receive a higher grade. If you improve your attendance, your grades will probably improve. Many factors affect your grade in a course. One factor that you have considerable control over is attendance. You can increase your opportunities for learning by attending class more often.

1. What are the variables under study?
2. What are the data in the study?
3. Are descriptive, inferential, or both types of statistics used?
4. What is the population under study?
5. Was a sample collected? If so, from where?
6. From the information given, comment on the relationship between the variables.

See page 39 for the answers.

Unusual Stat

Only one-third of crimes committed are reported to the police.

Exercises 1–1

1. Define statistics.

2. What is a variable?

3. What is meant by a census?

4. How does a population differ from a sample?

5. Explain the difference between descriptive and inferential statistics.

6. Name two areas where probability is used.

7. Why is information obtained from samples used more often than information obtained from populations?

8. What is meant by a biased sample?

For Exercises 9–17, determine whether descriptive or inferential statistics were used.

9. Because of the current economy, 49% of 18 to 34 year-olds have taken a job to pay the bills. (Source: Pew Research Center)

10. In 2025, the world population is predicted to be 8 billion people. (Source: United Nations)

11. In 2011, there were 34 deaths from the avian flu. (Source: World Health Organization)

12. Based on a sample of 2739 respondents, it is estimated that pet owners spent a total of 14 billion dollars on veterinarian care for their pets. (Source: American Pet Products Association, Pet Owners Survey)

13. In 2011, 79% of U.S. adults used the Internet. (Source: Pew Research Center)

14. In 2010, a total of 68,905 people died from diabetes. (Source: Centers for Disease Control and Prevention)

15. In an online survey of 500 Virginia Tech students between spring 2010 and spring 2011, 31% said that they had missed class because of alcohol consumption. (Source: Center for Applied Behavior Systems at Virginia Tech)

16. In 2008–2009, a total of 260,327 U.S. students were studying abroad. (Source: Institute of International Education)

17. Forty-four percent of the people in the United States have type O blood. (Source: American Red Cross)

Extending the Concepts

18. Find three statistical studies and explain whether they used descriptive or inferential statistics.

19. Find a gambling game and explain how probability was used to determine the outcome.

1–2 Variables and Types of Data

OBJECTIVE 3

Identify types of data.

As stated in Section 1–1, statisticians gain information about a particular situation by collecting data for random variables. This section will explore in greater detail the nature of variables and types of data.

Variables can be classified as qualitative or quantitative.

Qualitative variables are variables that have distinct categories according to some characteristic or attribute.

For example, if subjects are classified according to gender (male or female), then the variable *gender* is qualitative. Other examples of qualitative variables are religious preference and geographic locations.

Quantitative variables are variables that can be counted or measured.

For example, the variable *age* is numerical, and people can be ranked in order according to the value of their ages. Other examples of quantitative variables are heights, weights, and body temperatures.

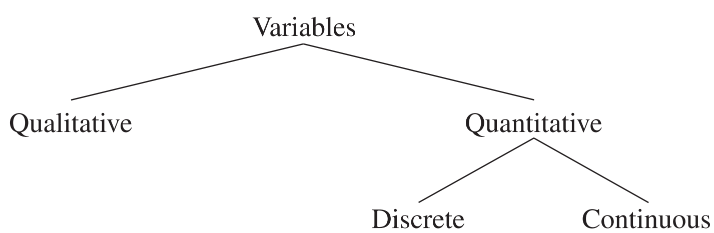
Quantitative variables can be further classified into two groups: discrete and continuous. *Discrete variables* can be assigned values such as 0, 1, 2, 3 and are said to be *countable.* Examples of discrete variables are the number of children in a family, the number of students in a classroom, and the number of calls received by a switchboard operator each day for a month.

Discrete variables assume values that can be counted.

Continuous variables, by comparison, can assume an infinite number of values in an interval between any two specific values. Temperature, for example, is a continuous variable, since the variable can assume an infinite number of values between any two given temperatures.

Continuous variables can assume an infinite number of values between any two specific values. They are obtained by measuring. They often include fractions and decimals.

The classification of variables can be summarized as follows:



EXAMPLE 1–2 Discrete or Continuous Variables

Classify each variable as a discrete variable or a continuous variable.

a. The highest wind speed of a hurricane
b. The weight of baggage on an airplane
c. The number of pages in a statistics book
d. The amount of money a person spends per year for online purchases

SOLUTION

a. Continuous, since wind speed must be measured
b. Continuous, since weight is measured
c. Discrete, since the number of pages is countable
d. Discrete, since the smallest value that money can assume is in cents

Unusual Stat

Fifty-two percent of Americans live within 50 miles of a coastal shoreline.

Since continuous data must be measured, answers must be rounded because of the limits of the measuring device. Usually, answers are rounded to the nearest given unit. For example, heights might be rounded to the nearest inch, weights to the nearest ounce, etc. Hence, a recorded height of 73 inches could mean any measure from 72.5 inches up to but not including 73.5 inches. Thus, the boundary of this measure is given as 72.5–73.5 inches. The **boundary** of a number, then, is defined as a class in which a data value would be placed before the data value was rounded. *Boundaries are written for convenience as 72.5–73.5 but are understood to mean all values up to but not including 73.5.* Actual data values of 73.5 would be rounded to 74 and would be included in a class with boundaries of 73.5 up to but not including 74.5, written as 73.5–74.5. As another example, if a recorded weight is 86 pounds, the exact boundaries are 85.5 up to but not including 86.5, written as 85.5–86.5 pounds. Table 1–1 helps to clarify this concept. The boundaries of a continuous variable are given in one additional decimal place and always end with the digit 5.

TABLE 1–1 Recorded Values and Boundaries

Variable	Recorded value	Boundaries
Length	15 centimeters (cm)	14.5–15.5 cm
Temperature	86 degrees Fahrenheit (°F)	85.5–86.5°F
Time	0.43 second (sec)	0.425–0.435 sec
Mass	1.6 grams (g)	1.55–1.65 g

EXAMPLE 1–3 Class Boundaries

Find the boundaries of each variable.

a. 8.4 quarts
b. 138 mmHg
c. 137.63 mg/dL

SOLUTION

a. 8.35–8.45 quarts
b. 137.5–138.5 mmHg
c. 137.625–137.635 mg/dL

In addition to being classified as qualitative or quantitative, variables can be classified by how they are categorized, counted, or measured. For example, can the data be organized into specific categories, such as area of residence (rural, suburban, or urban)? Can the data values be ranked, such as first place, second place, etc.? Or are the values obtained from measurement, such as heights, IQs, or temperature? This type of classification—i.e., how variables are categorized, counted, or measured—uses **measurement scales,** and four common types of scales are used: nominal, ordinal, interval, and ratio.

OBJECTIVE 4

Identify the measurement level for each variable.

The first level of measurement is called the *nominal level* of measurement. A sample of college instructors classified according to subject taught (e.g., English, history, psychology, or mathematics) is an example of nominal-level measurement. Classifying survey subjects as male or female is another example of nominal-level measurement. No ranking or order can be placed on the data. Classifying residents according to zip codes is also an example of the nominal level of measurement. Even though numbers are assigned as zip codes, there is no meaningful order or ranking. Other examples of nominal-level data are political party (Democratic, Republican, Independent, etc.), religion (Christianity, Judaism, Islam, etc.), and marital status (single, married, divorced, widowed, separated).

Unusual Stat

Sixty-three percent of us say we would rather hear the bad news first.

Historical Note

When data were first analyzed statistically by Karl Pearson and Francis Galton, almost all were continuous data. In 1899, Pearson began to analyze discrete data. Pearson found that some data, such as eye color, could not be measured, so he termed such data *nominal data.* Ordinal data were introduced by a German numerologist Frederich Mohs in 1822 when he introduced a hardness scale for minerals. For example, the hardest stone is the diamond, which he assigned a hardness value of 1500. Quartz was assigned a hardness value of 100. This does not mean that a diamond is 15 times harder than quartz. It only means that a diamond is harder than quartz. In 1947, a psychologist named Stanley Smith Stevens made a further division of continuous data into two categories, namely, interval and ratio.

The **nominal level of measurement** classifies data into mutually exclusive (nonoverlapping) categories in which no order or ranking can be imposed on the data.

The next level of measurement is called the *ordinal level.* Data measured at this level can be placed into categories, and these categories can be ordered, or ranked. For example, from student evaluations, guest speakers might be ranked as superior, average, or poor. Floats in a homecoming parade might be ranked as first place, second place, etc. *Note that precise measurement of differences in the ordinal level of measurement* does not *exist.* For instance, when people are classified according to their build (small, medium, or large), a large variation exists among the individuals in each class.

Other examples of ordinal data are letter grades (A, B, C, D, F).

The **ordinal level of measurement** classifies data into categories that can be ranked; however, precise differences between the ranks do not exist.

The third level of measurement is called the *interval level.* This level differs from the ordinal level in that precise differences do exist between units. For example, many standardized psychological tests yield values measured on an interval scale. IQ is an example of such a variable. There is a meaningful difference of 1 point between an IQ of 109 and an IQ of 110. Temperature is another example of interval measurement, since there is a meaningful difference of 1°F between each unit, such as 72 and 73°F. *One property is lacking in the interval scale: There is no true zero.* For example, IQ tests do not measure people who have no intelligence. For temperature, 0°F does not mean no heat at all.

The **interval level of measurement** ranks data, and precise differences between units of measure do exist; however, there is no meaningful zero.

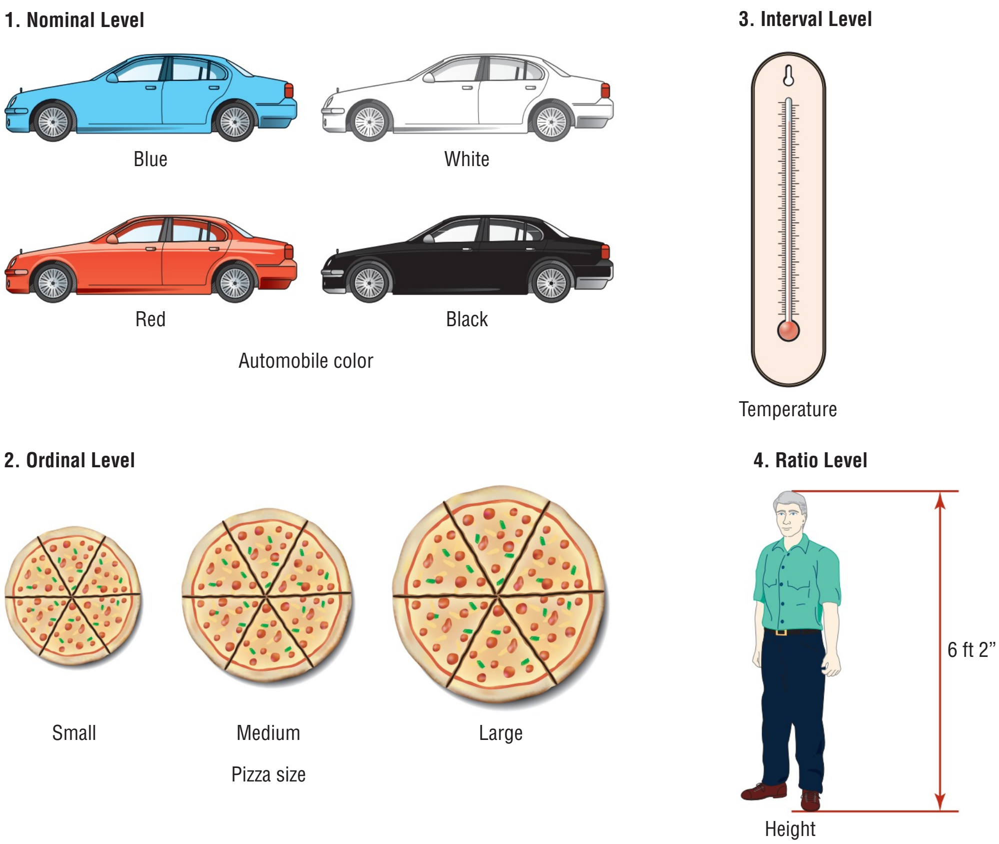
The final level of measurement is called the *ratio level.* Examples of ratio scales are those used to measure height, weight, area, and number of phone calls received. Ratio scales have differences between units (1 inch, 1 pound, etc.) and a true zero. In addition, the ratio scale contains a true ratio between values. For example, if one person can lift 200 pounds and another can lift 100 pounds, then the ratio between them is 2 to 1. Put another way, the first person can lift twice as much as the second person.

The **ratio level of measurement** possesses all the characteristics of interval measurement, and there exists a true zero. In addition, true ratios exist when the same variable is measured on two different members of the population.

TABLE 1–2 Examples of Measurement Scales

Nominal-level data	Ordinal-level data	Interval-level data	Ratio-level data
Zip code Gender (male, female) Eye color (blue, brown, green, hazel) Political affiliation Religious affiliation Major field (mathematics, computers, etc.) Nationality	Grade (A, B, C, D, F) Judging (first place, second place, etc.) Rating scale (poor, good, excellent) Ranking of tennis players	SAT score IQ Temperature	Height Weight Time Salary Age

FIGURE 1–2
Measurement Scales



There is not complete agreement among statisticians about the classification of data into one of the four categories. For example, some researchers classify IQ data as ratio data rather than interval. Also, data can be altered so that they fit into a different category. For instance, if the incomes of all professors of a college are classified into the three categories of low, average, and high, then a ratio variable becomes an ordinal variable. Table 1–2 gives some examples of each type of data. See Figure 1–2.

EXAMPLE 1–4 Measurement Levels

What level of measurement would be used to measure each variable?

a. The ages of patients in a local hospital
b. The ratings of movies released this month
c. Colors of athletic shirts sold by Oak Park Health Club
d. Temperatures of hot tubs in local health clubs