

# UNDERSTANDING Business Statistics

NED FREED | STACEY JONES | TIM BERGQUIST

WILEY



# WileyPLUS

**WileyPLUS is a research-based online environment for effective teaching and learning.**

*WileyPLUS* builds students' confidence because it takes the guesswork out of studying by providing students with a clear roadmap:

- what to do
- how to do it
- if they did it right

It offers interactive resources along with a complete digital textbook that help students learn more. With *WileyPLUS*, students take more initiative so you'll have greater impact on their achievement in the classroom and beyond.



For more information, visit [www.wileyplus.com](http://www.wileyplus.com)

Now available for



Blackboard

# WileyPLUS

**ALL THE HELP, RESOURCES, AND PERSONAL SUPPORT YOU AND YOUR STUDENTS NEED!**

[www.wileyplus.com/resources](http://www.wileyplus.com/resources)

**1<sup>st</sup>** DAY OF CLASS  
**... AND BEYOND!**

2-Minute Tutorials and all of the resources you and your students need to get started

**WileyPLUS**

Student Partner Program

Student support from an experienced student user

**Wiley Faculty Network**



Collaborate with your colleagues, find a mentor, attend virtual and live events, and view resources

[www.WhereFacultyConnect.com](http://www.WhereFacultyConnect.com)

**WileyPLUS**

Quick Start

Pre-loaded, ready-to-use assignments and presentations created by subject matter experts



Technical Support 24/7  
FAQs, online chat,  
and phone support

[www.wileyplus.com/support](http://www.wileyplus.com/support)



© Courtney Keating/iStockphoto

Your *WileyPLUS* Account Manager, providing personal training and support

UNDERSTANDING

# Business Statistics



# UNDERSTANDING Business Statistics

---

---

**NED FREED**

University of Portland

**STACEY JONES**

Seattle University

**TIM BERGQUIST**

Northwest Christian University

**WILEY**



VICE PRESIDENT & EXECUTIVE PUBLISHER	George Hoffman
EXECUTIVE EDITOR	Lise Johnson
SENIOR EDITOR	Franny Kelly
PROJECT EDITOR	Brian Kamins
EDITORIAL ASSISTANT	Jacqueline Hughes
DIRECTOR OF MARKETING	Amy Scholz
SENIOR MARKETING MANAGER	Margaret Barrett
MARKETING ASSISTANT	Juliette San Filipo
DEVELOPMENTAL EDITOR	Susan McLaughlin
DESIGN DIRECTOR	Harry Nolan
COVER AND INTERIOR DESIGNER	Wendy Lai
PRODUCTION MANAGER	Dorothy Sinclair
PRODUCTION EDITOR	Sandra Dumas
SENIOR PRODUCT DESIGNER	Allison Morris
PRODUCT DESIGNER	Greg Chaput
EDITORIAL OPERATIONS MANAGER	Yana Mermel
SENIOR PHOTO EDITOR	MaryAnn Price

This book was typeset in 10/12 Adobe Garamond Pro Regular at Aptara, Inc. and printed and bound by Courier Companies. The cover was printed by Courier Companies.

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: [www.wiley.com/go/citizenship](http://www.wiley.com/go/citizenship).

This book is printed on acid-free paper.

Copyright © 2014 by John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201) 748-6011, fax (201) 748-6008.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at [www.wiley.com/go/returnlabel](http://www.wiley.com/go/returnlabel). If you have chosen to adopt this textbook for use in your course, please accept this book as your complimentary desk copy. Outside of the United States, please contact your local representative.

ISBN: 978-1118-14525-8

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1



# Brief Contents



<b>1</b>	An Introduction to Statistics	2
<b>2</b>	Descriptive Statistics I: Elementary Data Presentation and Description	16
<b>3</b>	Descriptive Statistics II: Additional Descriptive Measures and Data Displays	62
<b>4</b>	Probability	108
<b>5</b>	Discrete Probability Distributions	158
<b>6</b>	Continuous Probability Distributions	194
<b>7</b>	Statistical Inference: Estimating a Population Mean	228
<b>8</b>	Interval Estimates For Proportions, Mean Differences, and Proportion Differences	272
<b>9</b>	Statistical Hypothesis Testing: Hypothesis Tests for a Population Mean	310
<b>10</b>	Hypothesis Tests for Proportions, Mean Differences, and Proportion Differences	348
<b>11</b>	Basic Regression Analysis	388
<b>12</b>	Multiple Regression	440
<b>13</b>	<i>F</i> Tests and Analysis of Variance	488
<b>14</b>	Chi-Square Tests	516



# Praise for *Understanding Business Statistics*

“The text is wonderfully articulate with great explanations and lots of homework problems.”

—**PAUL D. BERGER**, Bentley University

“The authors have written an excellent business statistics textbook. It is so readable that diligent students should be able to read and understand it on their own. It will be a great teaching aid for instructors and a great learning aid for students.”

—**HARVEY A. SINGER**,  
George Mason University

“This is a textbook written in simplest terms, making it easy for students to understand and appreciate statistics, and making learning statistics more fun. The authors present the material in an engaging conversational manner, using current issues and cases that students can easily relate to.”

—**RENE ORDONEZ**, Southern Oregon University

“This is a great book for learning basic statistics and the scientific method affording the learner a sophisticated accessibility to statistics fundamentals and experimental design.”

—**COURTNEY PHAM**,  
Missouri State University

“This is a student friendly book written with a touch of added humor. The examples are very relevant for today’s business students.”

—**KAY BALLARD**,  
University of Nebraska-Lincoln

“The authors have written a book that is easy to read and understand, guiding students through detailed problems, with plenty of repetition.”

—**WALTER MAYER**,  
University of Mississippi

“Great writing style. The book does an excellent job of conveying the information in a simple and direct manner. The examples are relevant and explained in great detail, step by step.”

—**CHRISTOPHER O’BYRNE**, Cuyamaca College

“The visual presentation of the examples and the step by step approach in solving them is excellent. I believe this is a key element missing from the textbook that I currently use (Anderson Essentials).”

—**CHRISTOPHER O’BYRNE**,  
Cuyamaca College

“The coverage in this book is very user-friendly. It is well organized, easy to read and has lots of interesting and useful examples/exercises.”

—**BHARATENDRA RAJ**,  
University of Massachusetts, Dartmouth

“I see many statistics books ... and for the most part ... can’t stand them, because they are written from a statistics department perspective ... which completely leaves the majority of business students guessing ... this text is business student focused with business related examples and exercises.”

—**MICHAEL KING**, Virginia Tech

“From the perspective of the students, I believe that this text is much clearer and simpler for the students to understand than my current text (Keller) ... The authors have given careful attention to their intended audience. The chapters are written in an easy to follow style and structure.”

—**NIMER ALRUSHIEDAT**, California State University, Fullerton

“For the students, the textbook is readable, with clear explanations of statistical concepts. For the professor, there are a lot of good exercises and the solution manual provides the students with detailed solutions that cut down on the number of questions that have to be addressed during office hours.”

—**ERIC HOWINGTON**, Valdosta State University



# About the Authors

## **Ned Freed**

With more than 30 years of experience as an award-winning teacher and researcher in the area of statistics and quantitative methods, Dr. Ned Freed has served on the faculties of the Naval Postgraduate School in Monterey, California, and the University of Portland in Portland, Oregon. Professor Freed graduated from Boston College with a BS in mathematics, earned an MBA from the Wharton School at the University of Pennsylvania, and holds a Ph. D. in management science from the University of Colorado. He has authored a number of articles dealing with mathematical programming applications to statistical problems and has worked extensively with Prof. Fred Glover at the University of Colorado to produce linear programming alternatives to conventional methods of multivariate statistics.

## **Stacey Jones**

Stacey Jones is a member of the Economics Department of the Albers School of Business and Economics at Seattle University, where she teaches several sections of introductory statistics and quantitative methods courses each year. Her students speak highly of her clarity, humor, and concreteness. Many of her courses incorporate service-learning projects that help students understand the relevance of statistical methods in addressing real-world problems. She holds a BA in philosophy from Carleton College and a Ph. D. in economics from Stanford University, and has authored a number of publications on discrimination and inequality in higher education and the labor market.

## **Tim Bergquist**

Dr. Tim Bergquist has been a Professor of Quantitative Analysis at Northwest Christian University in Eugene, Oregon, since 1996. He has more than 25 years of teaching experience at schools that include South Dakota State University, the University of Oregon, and Oregon State University. He earned a BS in Mathematics from the University of Portland, an MS in Statistics from the University of Louisiana at Lafayette, an MBA from Santa Clara University, an MS in Operations Research from Oregon State University, and a Ph. D. in Decision Sciences from the University of Oregon. Dr. Bergquist has published or written numerous articles and manuals, and has presented more than 50 papers at professional conferences. He is the co-winner of the 2010 Decision Sciences Institute Instructional Innovation Award Competition and the recipient of the 2013 President's Award for Teaching Excellence and Campus Leadership.



## 1

### An Introduction to Statistics 2

---

- 1.1 Statistics Defined 4
- 1.2 Branches of Statistics 4
  - Descriptive Statistics 5
  - Statistical Inference 5
  - Probability Theory: The Link 6
- 1.3 Two Views of the World 7
  - A World of Certainty 7
  - A World of Uncertainty 7
- 1.4 The Nature of Data 7
  - Qualitative and Quantitative Data 7
  - Time Series and Cross-Sectional Data 8
  - Levels of Measurement 9
- 1.5 Data Sources 11
  - The Internet 11
  - Government Agencies and Private-Sector Companies 11
  - Original Studies 12
- 1.6 Ethics in Statistics 12
- 1.7 The Text 13
  - Goals 13
  - Key Elements 13
- 1.8 A Final Comment 13
- Glossary 14
- Chapter Exercises 14

## 2

### Descriptive Statistics I: Elementary Data Presentation and Description 16

---

- 2.1 Measures of Central Location or Central Tendency 18
  - Mean 18
  - Median 19
  - Mode 20

### 2.2 Measures of Dispersion 22

- Range 22
- Mean Absolute Deviation 23
- Variance 24
- Standard Deviation 25

### 2.3 Frequency Distributions 28

- Frequency Distribution Shapes 29
- Computing Descriptive Measures for Frequency Distributions 32
- The Effect of Distribution Shape on Descriptive Measures 35

### 2.4 Relative Frequency Distributions 36

- Relative Frequency Bar Charts 37
- Computing Descriptive Measures for Relative Frequency Distributions 38

### 2.5 Cumulative Distributions 41

- Cumulative Frequency Distributions 41
- Cumulative Relative Frequency Distribution 42

### 2.6 Grouped Data 44

- Histograms 45
- Approximating Descriptive Measures for Grouped Data 46
- A Final Note on Grouped Data 50
- Key Formulas 50
- Glossary 51
- Chapter Exercises 51
- Excel Exercises (Excel 2013) 56

## 3

### Descriptive Statistics II: Additional Descriptive Measures and Data Displays 62

---

### 3.1 Percentiles and Quartiles 64

- Percentiles 64
- Quartiles 67
- Measuring Dispersion with the Interquartile Range 70

### 3.2 Exploratory Data Analysis 71

- Stem-and-Leaf Diagrams 71
- An Ordered Stem-and-Leaf Diagram 72

Box Plots 75

**3.3 Identifying Outliers** 78

1.5 × Interquartile Range 79

Chebyshev's Rule 79

The Empirical Rule 79

**3.4 Measures of Association** 82

Covariance 83

Correlation Coefficient 87

Covariance and Correlation Coefficients for Samples 88

A Final Note 89

**3.5 Additional Descriptive Measures** 91

Coefficient of Variation 91

The Geometric Mean 94

Weighted Average (Weighted Mean) 96

Key Formulas 98

Glossary 99

Chapter Exercises 99

Excel Exercises (Excel 2013) 105

# 4

**Probability** 108**4.1 Basic Concepts** 110

Defining Probability 110

Assigning Basic Probabilities 110

Classical Approach 110

Relative Frequency Approach 111

Subjective Approach 111

**4.2 The Rules of Probability** 112

Simple Probabilities 113

Conditional Probabilities 114

Statistical Independence 116

Joint Probabilities—the Multiplication Rule 118

Mutually Exclusive Events 120

Either/Or Probabilities—the Addition Rule 121

The “Conditional Equals Joint Over Simple” Rule 123

Complementary Events 125

**4.3 Venn Diagrams** 126

Showing the Addition Rule 126

Showing Conditional Probability 127

Showing Complementary Events 127

Showing Mutually Exclusive Events 127

**4.4 A General Problem Solving Strategy** 129

Probability Trees 129

Using a Probability Tree to Revise Probabilities 133

Joint Probabilities and Cross-Tabulation Tables 138

Choosing the Right Visual Aid 141

**4.5 Counting Outcomes** 142

Multiplication Method 142

Combinations 143

Permutations 144

Key Formulas 147

Glossary 147

Chapter Exercises 148

Excel Exercises (Excel 2013) 154

# 5

**Discrete Probability Distributions** 158**5.1 Probability Experiments and Random Variables** 160**5.2 Building a Discrete Probability Distribution** 162**5.3 Displaying and Summarizing the Distribution** 166

Graphing the Distribution 166

The Distribution Mean (Expected Value) 166

The Variance and Standard Deviation of the Distribution 167

**5.4 The Binomial Probability Distribution** 169

The Binomial Conditions 170

The Binomial Probability Function 171

Logic of the Binomial Function 171

Descriptive Measures for the Binomial Distribution 173

The Binomial Table 175

Shape(s) of the Binomial Distribution 177

**5.5 The Poisson Distribution** 178

Conditions of a Poisson Experiment 178

The Poisson Probability Function 179

The Poisson Table 181

Graphing the Poisson Distribution 182

Descriptive Measures 183

Using the Poisson Distribution to Approximate Binomial Probabilities 185

Key Formulas	186
Glossary	187
Chapter Exercises	187
Excel Exercises (Excel 2013)	191

## 6

### Continuous Probability Distributions 194

---

<b>6.1</b> Continuous vs. Discrete Distributions	196
<b>6.2</b> The Uniform Probability Distribution	197
Assigning Probabilities in a Uniform Distribution	197
Total Area under the Curve	198
General Distribution Characteristics	198
<b>6.3</b> The Normal Distribution	201
The Normal Probability Density Function	202
Key Normal Distribution Properties	202
The Standard Normal Table	204
Calculating z-scores for Any Normal Distribution	206
Normal Table Applications	207
Using the Normal Table in Reverse	211
<b>6.4</b> The Exponential Distribution	214
The Exponential Probability Density Function	214
Descriptive Measures for the Exponential Distribution	217
The Memoryless Nature of the Exponential Distribution	219
Key Formulas	220
Glossary	221
Chapter Exercises	221
Excel Exercises (Excel 2013)	225

## 7

### Statistical Inference: Estimating A Population Mean 228

---

<b>7.1</b> The Nature of Sampling	230
Defining Statistical Inference	230
Why Sample?	230
<b>7.2</b> Developing a Sampling Plan	231
Choosing a Sample Size	231
Random Sampling	231

Alternative Selection Procedures	232
Using a Random Number Generator or a Random Number Table	233
Sampling with or without Replacement	233
A Note on the Use of Random Numbers	234
<b>7.3</b> Confidence Intervals and the Role of the Sampling Distribution	236
The Basic Nature of a Confidence Interval	236
Sampling Distributions	237
The Sampling Distribution of the Sample Mean	237
Properties of the Sampling Distribution of the Sample Mean	240
Using Sampling Distribution Properties to Build a Confidence Interval	245
Visualizing the Role of the Sampling Distribution in Interval Estimation	249
Standard Error versus Margin of Error	250
<b>7.4</b> Building Intervals when the Population Standard Deviation Is Unknown	252
Estimating the Population Standard Deviation with the Sample Standard Deviation, $s$	252
Using the $t$ Distribution when $s$ Estimates $\sigma$	253
Constructing Intervals with the $t$ Distribution	254
Application to the Social Media Example	255
The Normal Distribution as an Approximation to the $t$ Distribution	257
<b>7.5</b> Determining Sample Size	259
Factors Influencing Sample Size	259
The Basic Procedure	259
Key Formulas	261
Glossary	262
Chapter Exercises	263
Excel Exercises (Excel 2013)	267

## 8

### Interval Estimates for Proportions, Mean Differences, and Proportion Differences 272

---

<b>8.1</b> Estimating a Population Proportion	274
An Example	274
The Sampling Distribution of the Sample Proportion	275
Predictable Sampling Distribution Properties	277



Using Sampling Distribution Properties to Build a Confidence Interval	280
Determining Sample Size	283
Determining Sample Size with No Information about $\pi$	285
<b>8.2 Estimating the Difference between Two Population Means (Independent Samples)</b>	<b>286</b>
An Example	287
The Sampling Distribution of the Sample Mean Difference	287
Predictable Sampling Distribution Properties	288
Building the Interval	288
Small Sample Adjustments	291
<b>8.3 Estimating the Difference between Two Population Proportions</b>	<b>293</b>
An Example	293
The Sampling Distribution of the Sample Proportion Difference	294
Building the Interval	295
<b>8.4 Matched Samples</b>	<b>297</b>
An Example	297
Key Formulas	300
Glossary	301
Chapter Exercises	301
Excel Exercises (Excel 2013)	308

Choosing a Significance Level	318
Establishing a Decision Rule	318
Applying the Decision Rule	319
Accepting vs. Failing to Reject the Null Hypothesis	320
Summarizing Our Approach	321
Another Way to State the Decision Rule	323
<i>p-values</i>	326
Generalizing the Test Procedure	329
<b>9.4 The Possibility of Error</b>	<b>330</b>
The Risk of Type I Error	330
The Risk of Type II Error	330
Choosing a Significance Level	331
<b>9.5 Two-Tailed Tests</b>	<b>332</b>
Designing a Two-Tailed Test	332
Two-Tailed Tests and Interval Estimation	333
Deciding Whether a Two-Tailed Test Is Appropriate	334
<b>9.6 Using the <i>t</i> Distribution</b>	<b>337</b>
An Illustration	337
<i>p-value</i> Approach	338
Key Formulas	340
Glossary	341
Chapter Exercises	341
Excel Exercises (Excel 2013)	345

## 9

### Statistical Hypothesis Testing: Hypothesis Tests For A Population Mean 310

<b>9.1 The Nature of Hypothesis Testing</b>	<b>312</b>
Comparing Hypothesis Testing to Interval Estimation	312
Illustrating the Logic of Hypothesis Testing	312
<b>9.2 Establishing the Hypotheses</b>	<b>313</b>
Choosing the Null Hypothesis	313
Standard Forms for the Null and Alternative Hypotheses	315
<b>9.3 Developing a One-Tailed Test</b>	<b>316</b>
A Preliminary Step: Evaluating Potential Sample Results	316
The Key: The Sampling Distribution of the Sample Mean	317

## 10

### Hypothesis Tests for Proportions, Mean Differences, and Proportion Differences 348

<b>10.1 Tests for a Population Proportion</b>	<b>350</b>
Forming the Hypotheses	350
The Sampling Distribution of the Sample Proportion	350
The Null Sampling Distribution	351
Choosing a Significance Level	351
Establishing the Critical Value	351
Putting the Sample Result through the Test	352
Reporting the Critical $\bar{p}$	353
<i>p-value</i> Approach	353
The Possibility of Error	357
The Probability of Error	357

## 10.2 Tests for the Difference Between Two Population Means (Independent Samples) 358

- Forming the Hypotheses 358
- The Sampling Distribution of the Sample Mean Difference 358
- The Null Sampling Distribution 359
- Separating Likely from Unlikely Sample Results 359
- Putting the Sample Result through the Test 359
- p*-value Approach 360
- When Population  $\sigma$ s Are Unknown 362
- A Final Note 366

## 10.3 Tests for the Difference Between Two Population Proportions 367

- Forming the Hypotheses 367
- The Sampling Distribution 367
- The Null Sampling Distribution 368
- Establishing the Critical Value 368
- Computing the Value of the Test Statistic 369
- Computing the Standard Error of the Null Sampling Distribution 369
- Completing the Test 370
- p*-value Approach 371
- Minimum Sample Sizes 371

## 10.4 Matched Samples 373

- An Example 373
- Key Formulas 376
- Chapter Exercises 377
- Excel Exercises (Excel 2013) 383

# 11

## Basic Regression Analysis 388

### 11.1 An Introduction to Regression 390

- The Nature of Regression Analysis 390
- Regression Analysis Variations 390
- Simple vs. Multiple Regression 390
- Linear vs. Nonlinear Regression 391
- The Base Case: Simple Linear Regression 391

### 11.2 Simple Linear Regression: The Basic Procedure 391

- The Data 392

- Showing the Scatter Diagram 392

- Fitting a Line to the Data 393

- The Least Squares Criterion 394

- Identifying the Least Squares Line 395

- Producing the Slope and Intercept of the Best-Fitting Line 395

- Locating the Values for *a* and *b* in a Computer Printout 396

### 11.3 Performance Measures in Regression: How Well Did We Do? 400

- Standard Error of Estimate 400

- Coefficient of Determination ( $r^2$ ) 404

- Total Variation 404

- Explained Variation 404

- Correlation Coefficient (*r*) 406

- Reading an Expanded Computer Printout 408

### 11.4 The Inference Side of Regression Analysis 409

- Treating the Set of Observations as a Sample 409

- Bridging the Gap between Sample and Population 410

### 11.5 Estimating the Intercept and Slope Terms for the Population Regression Line 411

- The Sampling Distribution of the Sample Intercept 411

- The Sampling Distribution of the Sample Slope 412

- Building Confidence Intervals 412

- Identifying Interval Estimates of  $\alpha$  and  $\beta$  in a Computer Printout 415

### 11.6 The Key Hypothesis Test in Simple Regression 416

- The Competing Positions 416

- The Slope is the Key 416

- Formalizing the Test 417

- Identifying Hypothesis Testing Information in the Computer Printout 420

### 11.7 Estimating Values of *y* 421

- Estimating an Expected Value of *y* 421

- Estimating an Individual Value of *y* 423

### 11.8 A Complete Printout for Simple Linear Regression 425

### 11.9 Checking Errors (Residuals) 425

- Identifying Problems 426

- Autocorrelation 426

- A Final Comment 427

Key Formulas	427
Glossary	428
Chapter Exercises	428
Excel Exercises (Excel 2013)	432

# 12

## Multiple Regression 440

---

<b>12.1</b> The <i>F</i> Distribution	442
Basics of the <i>F</i> Distribution	442
Reading an <i>F</i> Table	443
<b>12.2</b> Using the <i>F</i> Distribution in Simple Regression	444
The Mobile Apps Example Revisited	444
The <i>t</i> Test	445
The <i>F</i> Test	446
Reasonableness of the <i>F</i> test	448
Connection Between the <i>t</i> Test and the <i>F</i> Test	449
<b>12.3</b> An Introduction to Multiple Regression	451
Getting Started	452
Interpreting the Coefficients	453
Performance Measures in Multiple Regression	454
<b>12.4</b> The Inference Side of Multiple Regression	457
Testing the Statistical Significance of the Relationship	459
<i>F</i> Test Results	459
Using <i>t</i> tests to Test Individual Coefficients	461
Interpreting <i>t</i> Test Results	462
Confidence Intervals for Individual Coefficients	464
<b>12.5</b> Building a Regression Model	466
What do we Learn from a Multiple Regression Model?	466
Why Not Conduct a Series of Simple Regressions?	466
The “Best” Set of Independent Variables	466
Adding Variables	466
Multicollinearity	467
Adjusted $r^2$	467
Qualitative Variables	469
Interaction Effects	472
A Final Note: Be Prepared for Anything	472

Key Formulas	472
Glossary	473
Chapter Exercises	473
Excel Exercises (Excel 2013)	477

# 13

## *F* Tests and Analysis of Variance 488

---

<b>13.1</b> The <i>F</i> Distribution	490
Basics of the <i>F</i> Distribution	490
Reading the <i>F</i> Table	491
<b>13.2</b> Testing the Equality of Two Population Variances	492
Two-Tailed Tests	492
One-Tailed Tests	493
<b>13.3</b> Testing the Equality of Means for Multiple Populations: One-Way Analysis of Variance	494
Preliminary Comments	495
The Formal Test	498
Within-Groups Sum of Squares	498
Between-Groups Sum of Squares	499
Mean Squares	500
Computing the Variance Ratio for Sample Results	500
Applying the Critical Value Rule	501
<i>p</i> -value Version of the Test	501
ANOVA Table	501
Summarizing the Test	502
Determining Which Means Are Different	503
<b>13.4</b> Experimental Design	506
Completely Randomized Design	507
Block Designs	507
Factorial Designs	507
Other Designs	508
Key Formulas	508
Glossary	508
Chapter Exercises	509
Excel Exercises (Excel 2013)	512

**Part Two of Chapter 13 (Experimental Design) is available on [www.wiley.com/college/freed](http://www.wiley.com/college/freed) and in WileyPLUS.**

# 14

## Chi-Square Tests 516

---

### 14.1 The Chi-Square ( $\chi^2$ ) Distribution 518

Basics of the Chi-Square Distribution 518

Distribution Shapes 518

Reading the Chi-Square Table 518

### 14.2 Chi-Square Tests for Differences in Population Proportions 520

Setting Up the Test 520

Calculating z-scores for Sample Results 520

Computing  $\chi_{stat}^2$  521

Using the  $\chi_{stat}^2$  Value in the Test 522

Reaching a Conclusion 523

Summarizing the Test 523

A Table Format to Test Proportion Differences 526

### 14.3 Chi-Square Goodness-of-Fit Tests 532

An Example 532

Summarizing the Test 534

Extending the Approach 534

### 14.4 Chi-Square Tests of Independence 537

The Hypotheses 537

Calculating Expected Frequencies 538

Computing the Chi-Square Statistic 539

Reaching a Conclusion 539

Summarizing the Test 540

Minimum Cell Sizes 540

Key Formulas 543

Glossary 544

Chapter Exercises 544

Excel Exercises (Excel 2013) 549

## End Notes 552

---

## Appendix A: Tables 553

---

## Appendix B: Selected Exercise Solutions 574

---

## Index 589

---



# Preface

Years of teaching introductory courses in business statistics have taught us a good deal about what students value in a textbook. Number one is clarity of presentation. Students want a book that explains things clearly and concisely, with a sharp focus and a minimum amount of jargon. They also want a text that is examples-driven and divided into digestible sections, with plenty of opportunities to check their progress and test their understanding. Beyond merely mastering theory and concept, today's students are motivated by the prospect of real-world applications and current relevance. They want to know not only *how* things work, but *where* and *when* they work. And while students value a book that's relaxed in tone and manageable in content, they don't want a text that's condescending or simplistic. It's with these principles firmly in mind that we chose to write this book.

As you'll shortly see, each chapter of the book begins with a vignette highlighting an everyday application of statistics or focusing on an important contribution to the field. These brief introductions, intended to stimulate student interest and offer context for the upcoming topic, exemplify our commitment to engaging students from the outset and motivating their curiosity. We think you'll find the presentation that follows relevant and concise, and well-suited to the needs and challenges of today's business student.

## Overview

---

*Understanding Business Statistics* is intended for a general audience of business students enrolled in a two- or four-year program of study. It represents the authors' many years of experience teaching introductory classes in business statistics at the undergraduate level. The book's effective organization of topics and efficient discussion of key ideas make it appealing to teachers, while its informal, intuitive style—supported by clear and constant reinforcement—makes it accessible even to students who may not be completely confident in their math skills.

## Organization of Topics

Choosing to write a book with a sharper focus than most, the authors have organized key topics in a compelling manner, simplifying the instructor's job of identifying which topics fit best into available class time. The presentation of topics is concise without being choppy, and no critical topics have been left out. Although the general sequencing of topics is fairly standard, there are a number of important differences that allow instructors to cover the material efficiently:

- The distinctive clustering of both graphical and numerical methods in Chapter 2 (Descriptive Statistics I) introduces the core of descriptive statistics in a smooth-flowing and cohesive discussion of key descriptive elements. Chapter 3 presents somewhat less commonly covered descriptive methods and is set up to allow instructors to conveniently pick and choose those topics that they wish to cover and omit those that they do not.
- Chapter 4 introduces the crucial elements of probability theory in an intuitive way, establishing all of the basics with one simple, unambiguous example. A general problem-solving strategy is firmly established, supported by clear and practical visual tools to aid in the implementation of that strategy.
- By integrating their introduction to sampling distributions into the Chapter 7 discussion of interval estimation, the authors offer a unique approach that motivates student interest and allows students to see immediately how sampling distributions drive the inference process. This approach is not only intuitively appealing but also efficient in reducing the time needed for coverage of these two crucial topics.

- Chapter 7's coverage of interval estimation (Estimating a Population Mean) is followed immediately by Chapter 8's discussion of interval estimation for proportions, mean differences, and proportion differences, giving students a chance to solidify their understanding of interval estimation before moving on to the somewhat more challenging ideas of hypothesis testing covered in Chapters 9 and 10.
- Chapter 11's treatment of simple linear regression uses only the basic  $t$  test, allowing instructors to defer discussion of the  $F$  distribution until Chapter 12 (Multiple Regression).
- Chapters 12, 13, and 14 allow instructors to add topics beyond the basic core—including multiple regression, one-way ANOVA, and contingency tables. Usefully, the three chapters are designed so that they can be covered in any order.

## Intuitive Approach, Informal Style, and Constant Reinforcement

This is a student-oriented text, intended to be that rarest of birds—a math book that's actually readable and understandable for students with widely varying mathematical backgrounds. Reviews have described the chapters as “engaging,” and the writing as “wonderfully direct.” The text is more accessible to students in several ways:

- The authors provide an intuitive discussion of basic statistical principles rather than a mathematically rigorous development. Simple examples are used to introduce and develop concepts and procedures. The presentation of formulas is nearly always preceded—or followed immediately—by an exploration of student instincts regarding the reasoning behind the formula. For example, prior to their first look at the formula for building confidence intervals, students are led through an intuitive discussion of those factors that could reasonably be expected to influence interval precision. Chapter 9, which introduces hypothesis testing, is an exceptionally effective presentation of the topic using this same highly intuitive approach.
- The authors' clarity of presentation puts even the most complicated issues within reach of every student.
- For ease of reading, chapter sections are designed to ensure easy-to-follow continuity from one section to the next.
- To help establish and maintain a supportive and informal tone, the authors use near-conversational language and periodically inject an interesting and/or humorous illustration or side note.
- Each chapter begins with a vignette highlighting an everyday application of statistics or focusing on an important contribution to the field. These brief and pertinent introductions serve to engage student interest, offer context for the upcoming topic, and provide a basis for active classroom discussion.
- Students are given frequent opportunities to check their understanding of topics as they proceed through the chapters, with exercises included at the end of nearly every section and many subsections. Each exercise set is preceded by a step-by-step Demonstration Exercise that clearly establishes the solution pattern for the exercises that follow.
- Numerous exercises—ranging from elementary to quite challenging—appear at the end of each chapter. In many cases, they have been designed to augment and extend chapter discussions rather than solely provide opportunities for drill and repetition. In a significant number of the exercises, “real” data are used, taken from sources like the *Wall Street Journal*, the Department of Transportation, and the Bureau of Economic Analysis. (All the exercises that give a source citation involve “real” data.) Solutions to the even-numbered exercises are provided. “Next Level” problems are included to allow students to explore more advanced topics and applications.
- To familiarize students with the capabilities of commonly available statistical packages, a set of Excel exercises, with step-by-step instructions, is provided at the end of most chapters.

## Beyond the Numbers

---

As students advance through the course, it's important that they not get lost in the details and lose sight of important broader issues. Here are some general principles that we think will help students look beyond the numbers:

- **Data quality matters.** Just as the most gifted chef can't make a fine meal from poor ingredients, the most brilliant statistician can't draw sound conclusions from flawed data. In statistics, the old adage "garbage in, garbage out" is in full effect. When you embark on a project that involves data, a crucial first step is to learn where the data came from, and whether the data accurately and objectively measure what they are supposed to measure.
- **A picture is worth a thousand words.** A graph that provides a simple, visual summary of your results is often more memorable and useful than a verbal summary of the same information. Think carefully about how you might use graphs and figures to present information. If you would like to pursue this further, the field of data visualization is growing rapidly and is worth exploring!
- **It ain't necessarily so.** When we use a sample to learn about an entire population, the claims we make are *probably*, but not *necessarily*, true. Claims based on sample data should be framed in language that shows this. For example, based on a survey, you might say: "We have strong evidence that the proportion of teens who own smart phones has risen," or "The survey results suggest a rise in the share of teens that own smart phone." The only way to be 100 percent certain of what is going on in an entire population is to take a survey of the entire population—in other words, a census.
- **The numbers don't always tell the whole story.** You might think of statistics as "storytelling with numbers." But keep in mind that the numbers don't always tell the whole story. For example, if production at a company is rising, there may be nonquantitative factors at work, such as a boost in morale or improved teamwork. The best research in business and economics takes into account both quantitative and qualitative information.
- **Metrics matter.** This may seem to contradict the previous claim, but when people do have a quantitative measurement, they often focus solely on the measurement, to the exclusion of other important dimensions of an issue. Economic policy makers, for example, focus on gross domestic product (GDP) as a measure of the economic well-being of a country. Sometimes this focus on GDP is at the expense of other dimensions of economic well-being such as innovation or equity. When you propose a metric, be aware that it may become a powerful guide to decision making.
- **Choose your methods to fit the question at hand.** An old expression says, "When all you have is a hammer, everything looks like a nail." You will soon have a rich set of statistical techniques at your disposal, from confidence intervals to multiple regression analysis. That said, with any problem you are presented, your first task is to define the problem clearly and figure out precisely what it is you would like to learn from the data. You may be able to answer the question with the techniques you learn here. But if not, you will have a strong enough foundation in statistics to explore new techniques when presented with a new type of question.
- **In business, having the numbers and understanding them is key to leading successfully.** Decision making is increasingly data-driven. The knowledge you acquire in this course should prepare you to be a critical and effective user of information. Your ability to use data will allow you to make better business decisions, an essential quality of a twenty-first-century business leader.
- **Small is beautiful.** In a big data age, it is easy to forget that much can be learned from taking a small sample and exploring it deeply. Large data sets reveal a great deal about correlation, but in order to gain insight about causality, sometimes zooming in close and looking deeply into a situation is the better strategy.
- **Keep it simple.** The goal of statistics is not to confuse, impress, or intimidate others into submission with Greek letters and long mathematical formulas. If a simple technique can transform data into actionable information, use it. If the situation truly demands a more complex technique, learn and apply it.
- **It's what you say, and how you say it.** A lot of hard work and valid statistical analysis can go to waste if it is not communicated effectively. Effective communication means different



things to different people. If you have a case to make, and a strong statistical finding to make your case, consider your audience, and prepare to present that finding in words, numbers, *and* pictures. Different people and audiences will be more receptive to a verbal, quantitative, or visual presentation.

- **It's not all in the math.** Business scholar Edward Deming made the point long ago that “good statistical work is the product of several kinds of knowledge working together.” Sometimes the answer to a problem must come from knowledge of the subject matter, not from the numbers. A good statistician knows when solving a problem requires a more sophisticated mathematical technique, and when solving a problem requires a deeper understanding of the context and subject matter.

## Ancillary Teaching and Learning Materials

---

A number of supplementary materials developed by the author team, are available to provide students and instructors with valuable support:

- PowerPoint slides for each chapter provide tables, charts, illustrative examples and more.
- Practice Problems allow students to focus their study when preparing for tests and quizzes.
- Student Solution Sets offer detailed solutions to the even-numbered exercises in each chapter.
- Instructor Solution Sets provide detailed solutions to every exercise in the text.
- Test Bank questions of varying levels of difficulty are organized to enable instructors to efficiently create quizzes and exams. (Over 1200 test questions in all.)
- Excel Manual provides students with detailed explanations and examples that show how to solve key statistical concepts and problems using Excel
- Excel Exercise Solutions provide the student with all the guidance necessary to develop an easy facility with Excel's statistical features.



# Acknowledgments

*This book has benefited greatly from the input of focus-group participants, manuscript reviewers, proofreaders, and many others who took the time to provide us with their insights and advice.*

Subrata Acharya, Townson University  
Nimer Alrushiedat, California State University, Fullerton  
Kay Ballard, University of Nebraska  
Frank Bensics, Central Connecticut State University  
John Bird, West Virginia State University  
Jackie Blackstock, Hillsdale College  
Stacy Bourgeois Roberts, University of North Carolina, Wilmington  
Mark Dantonio, Northern Virginia Community College  
Leonard Deaton, California Polytechnic State University  
Gretchen Donahue, University of Minnesota  
Joan M. Donohue, University of South Carolina  
Todd Easton, University of Portland  
James Gaugler, Texas A & M Kingsville  
Mark Gebert, University of Kentucky  
Seth Giertz, University of Nebraska  
Malcolm Getz, Vanderbilt University  
Bassam Hasan, University of Toledo  
Johnny Ho, Columbus State University  
Eric Howington, Valdosta State University  
Kuang-Chung Hsu, University of Central Oklahoma  
Donny Hurwitz, Austin Community College  
Brian Jue, California State University, Stanislaus  
Jerzy Kamburowski, University of Toledo  
Lara Khansa, Virginia Tech  
Michael A. King, Virginia Tech  
Clint Knox, Wayne County Community College  
Subhash Kochar, Portland State University  
Jose Lobo, Arizona State University  
Walter Mayer, University of Mississippi  
Margaret Niehaus, Gordon College  
Eric Nielsen, St. Louis Community College, Meramec  
Christopher O'Byrne, Cuyamaca College  
Mavis Pararai, Indiana University of Pennsylvania  
Scott Paulsen, Illinois Central College  
Courtney Pham, Missouri State University  
Bharatendra Rai, University of Massachusetts, Dartmouth  
Mohsen Sahebame, California State University Long Beach  
Yvonne Sandoval, Pima Community College, West Campus  
Harvey Singer, George Mason University

**Class Testers**

Jackie Blackstock, Hillsdale College

Bharatendra Rai, University of Massachusetts, Dartmouth

Margaret Niehaus, Gordon College

We appreciate the exemplary support and professional commitment given us by the development, marketing, production, and editorial teams at Wiley. We specifically want to thank the following people at Wiley for their commitment to this project: Franny Kelly, Kelly Simmons, George Hoffman, Lise Johnson, Margaret Barrett, Brian Kamins, Jackie Hughes, and Susan McLaughlin.

Of special note is the work of four people: Franny Kelly, who, from the beginning, has provided invaluable guidance and encouragement; Brian Kamins, whose patience and support has been terrific throughout; Susan McLaughlin, whose editorial recommendations and content suggestions were always spot-on; and Sandra Dumas, who performed her very challenging role with efficiency and good humor.



UNDERSTANDING

# Business Statistics

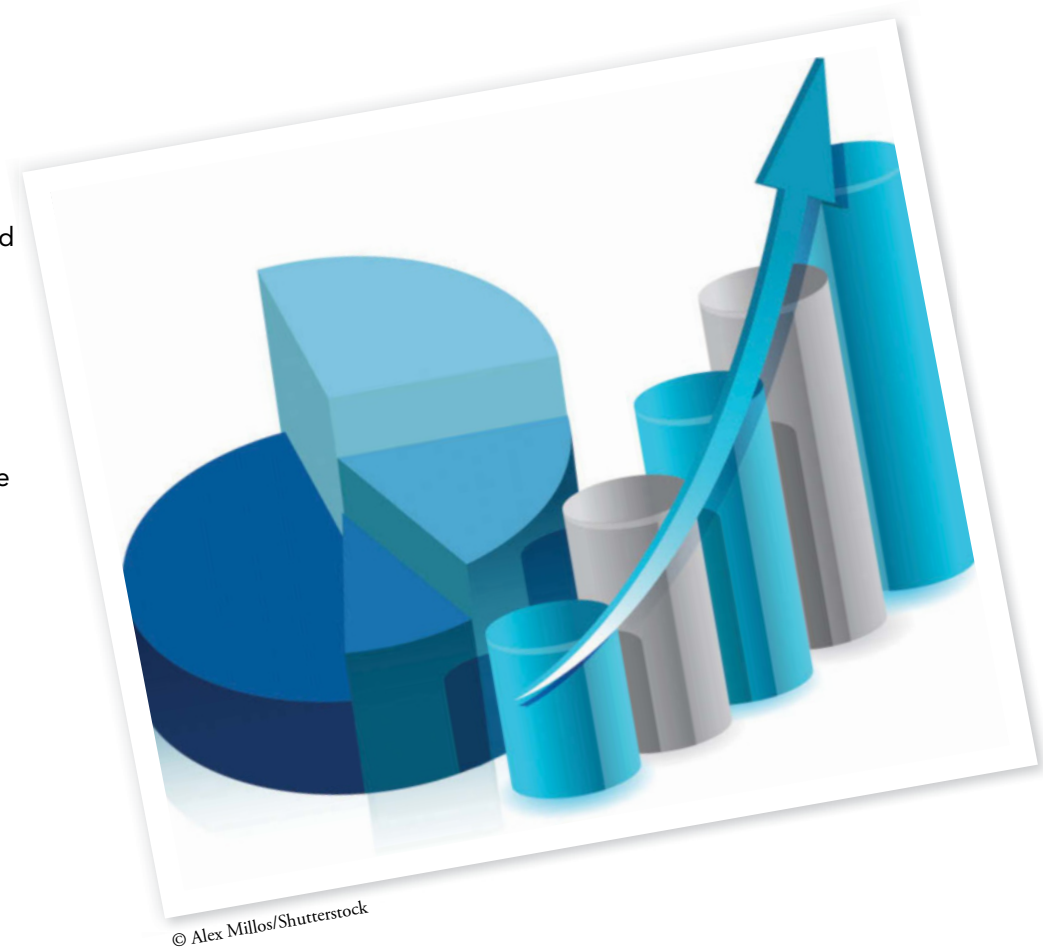


# An Introduction to Statistics

## LEARNING OBJECTIVES

After completing the chapter, you should be able to

1. Define *statistics*.
2. Identify the two major branches of statistics, describe their function, and discuss the role of probability as the link between the two.
3. Distinguish between a deterministic and probabilistic world view.
4. Differentiate qualitative from quantitative data, times series data from cross-sectional data, and define the four levels of measurement.
5. Discuss the various sources of data.
6. Summarize the main points in ASA's *Ethical Guidelines for Statistical Practice*.



# EVERYDAY STATISTICS

## Data, Data Everywhere

The past decade has been marked by an incredible growth in the collection and analysis of data. Today more and more companies are assembling large databases in an effort to learn more about their customers, their competitors, and the environment they live in. In the frenzy to mine this data bonanza, statisticians suddenly find themselves front and center. "It's like an arms race to hire statisticians nowadays," Andreas Weigend, the former chief scientist at Amazon, told the *New York Times*. "Mathematicians are suddenly sexy." (Well, that last part might be a bit of a stretch.)

How much information is out there? To answer that question, we first need a suitable unit of measurement.



© Aaron Bacall/www.CartoonStock.com

Digitally stored information is measured in bits and bytes. One *bit* (abbreviated "b") represents one choice between 0 and 1, or "yes" and "no." One *byte* (abbreviated "B") is made up of 8 bits, about the amount of information required to represent a single letter in the English language. One thousand bytes makes a *kilobyte* (KB), one million bytes makes a *megabyte* (MB), and one *million million* bytes ( $10^{12}$ ) is called a *terabyte*. The amount of information stored in the world has now grown to the point where we need new terms of measurement: enter *exabytes* ( $10^{18}$  bytes) and *zettabytes* ( $10^{21}$  bytes).

Researchers estimate that as of 2011, there were 1.8 *zettabytes* of data stored in the world. They arrived at that number by adding up the huge quantities of data stored on 60 different media, from PC hard drives and DVDs to books, newspapers, and vinyl records. (The number of books needed to hold this much data would cover a country the size of China 13 times over.) Not surprisingly, most of these data are now stored in digital form. In fact, that's been true since 2002, when, for the first time, digital data storage surpassed nondigital forms. According to Martin Hilbert, a scholar of the information age, "You could say the digital age started in 2002. That was the turning point." Since then, the amount of stored information has continued to grow exponentially. Facebook users alone upload more than 1,000 photos per second, adding up to 3 billion stored photos to the world's stock of digital data each month. Between 2002 and the present, it's estimated that the share of the world's information stored in digital form grew to nearly 94 percent.

Where is all this data? Most of it is currently on the hard drives of personal computers, but data storage is increasingly moving to data *clouds*—large-scale, off-site storage systems owned and maintained by companies like Google, Apple, EMC, and Microsoft. Cloud storage offers a low-cost option that allows users—both individual and corporate—to access their data from any location with an Internet connection. Needless to say, this has only added to the appetite for collecting more data.

So what are all these zettabytes of data good for? Frankly, not much, until they're transformed into useful information and communicated effectively. That's where statisticians play their most important role—distilling the numbers into meaningful information in order to answer questions and provide insight.

**WHAT'S AHEAD:** In this chapter and those that follow, we'll provide the statistical tools needed to turn data into information.

*Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write. —H.G. Wells*



“The Nasdaq gave up all of yesterday’s gains and then some, as it dropped nearly 16 points to close at 2744.” “The Angels slugger hit two towering home runs, lifting his RBI total to 132 and his batting average to .316.” “Teamsters officials argued that median pay, in constant dollars, was down 16% from five years ago and vowed to make no more concessions.” “According to NEA researchers, the relationship between income and SAT score appears undeniable; their study showed an  $r^2$  of .91 and a  $p$ -value of .0002.”

Sound a little like a foreign language? These kinds of statements, laden with technical terms and loaded with numbers, now seem an inescapable part of our everyday experience. Confronted daily by a barrage of data, most of us filter out what seems too complex and deal only with those bits and pieces that we can easily digest.

As we’ll shortly see, the study of statistics will greatly expand our ability to understand and manage the flood of facts and figures rising around us. It will allow us to crack the code of numeric and nonnumeric data and give us the capacity to intelligently assess the information they contain. Crucially for us, an understanding of statistics will open the door to valuable sources of information that we might otherwise have ignored or overlooked.

The future that H.G. Wells predicted—a future in which statistical skills will be an indispensable part of life—has surely arrived. To survive and prosper in this age of easy data access, we’ll need to prepare both for its unique challenges and for its unprecedented opportunities. In truth, to do otherwise is to surrender the future to someone else.

## 1.1 Statistics Defined

---

In common usage, the term *statistics* can take on a variety of meanings. It’s frequently used to describe data of almost any sort—height, weight, stock prices, batting average, GPA, temperature, and the like. Some people might connect the term to the results of surveys, polls, and questionnaires. In this text, we’ll use **statistics** primarily to designate a specific academic discipline focused on methods of data collection, analysis, and presentation. In virtually every case, statistics involves the transformation of data into information.



### Defining Statistics

**Statistics** is the art and science of collecting, analyzing, interpreting, and presenting data in an effort to transform the data into useful information.

In a business setting, statistical analysis generally implies translating raw numbers into meaningful information in order to establish the basis for sound *decision making*, whether we’re sorting out the day’s stock quotations to make a more informed investment decision or unscrambling the latest market research data so we can better respond to customer needs and wants.

## 1.2 Branches of Statistics

---

The area of statistics can be divided into two principal branches: *descriptive statistics* and *statistical inference* (or *inferential statistics*).

**TABLE 1.1**  
**Cross-tabulation Table Showing Survey Results for 511**  
**Executives from Companies of Various Sizes**

Opinion of Economy Today vs. 6 Months Ago	Company Size			Totals
	Small	Medium	Large	
Better	179	91	89	359
Same	61	37	25	123
Worse	23	3	3	29
Totals	263	131	117	511

Source: Bay Area Business Confidence Survey, Evans-McDonough

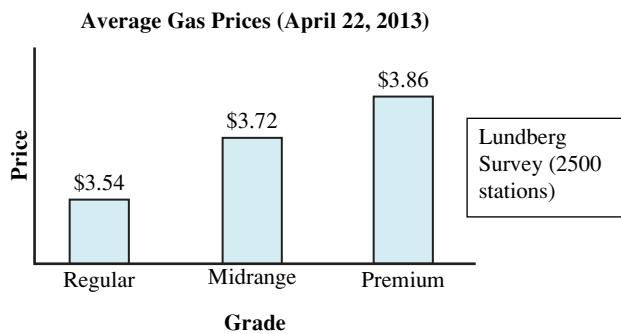
## Descriptive Statistics

**Descriptive statistics** focuses directly on summarizing and presenting data. It's here that we'll find ways to refine, distill, and describe data so that the numbers involved become more than simply numbers. We'll look at summary measures as basic as the *average* that will allow us to communicate the essence of a data set using only one or two values. We'll work with various table formats to help organize data. (See, for example, Table 1.1.)

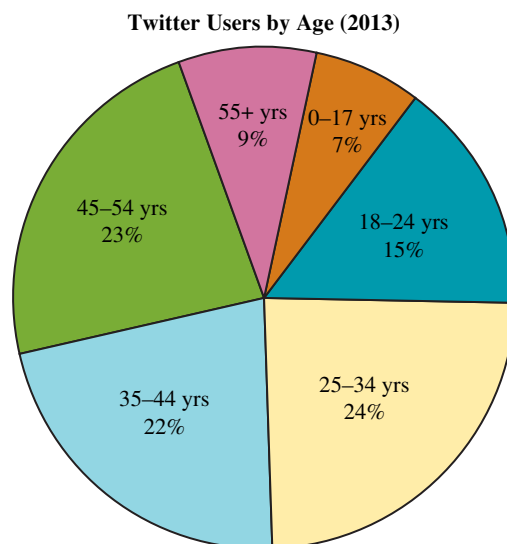
We'll also use graphical forms to display data, following the familiar saying that a picture is worth a thousand words—or numbers. Pictures as simple as *bar charts* (Figure 1.1) and *pie charts* (Figure 1.2) have the ability to communicate a great deal of information quickly and effectively.

## Statistical Inference

When dealing with especially large or complex data sets, it is sometimes necessary, for reasons of time, cost, or convenience, to draw conclusions about the entire data set by examining only a

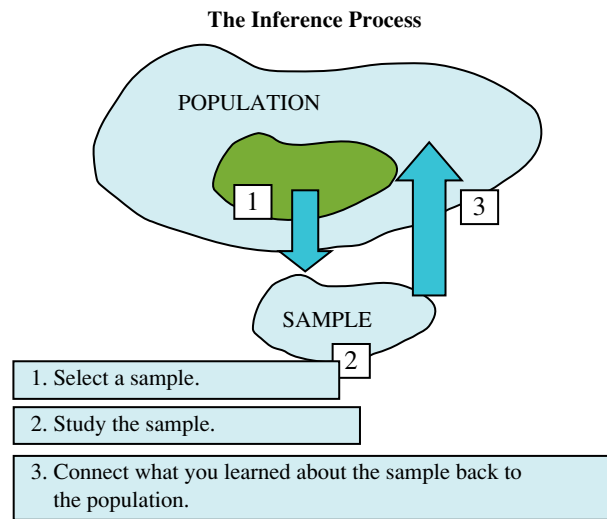


**FIGURE 1.1** Bar Chart Showing Results of a National Gas Price Survey



**FIGURE 1.2** Pie Chart Showing Twitter Users by Age Group

Source: Jobstock.com

**FIGURE 1.3** The Steps of Statistical Inference

subset—a **sample**—of the values involved. **Statistical inference**, or **inferential statistics**, deals with the selection and use of sample data to produce information about the larger **population** from which the sample was selected. Following the rules of statistical inference will enable us to *infer* what’s likely to be true of the population based solely on the data in the sample (see Figure 1.3).

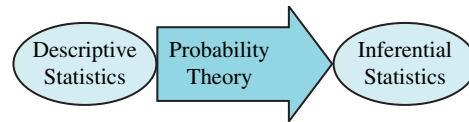
In this setting, the term **statistic** is often used to refer specifically to summary characteristics of a *sample*. For example, the average purchase amount for a sample of 10 customers selected to represent a population of 1000 customers would be considered a sample “statistic.” In contrast, summary characteristics of a *population* are frequently referred to as **parameters**. (In our example, the average purchase amount for the population of 1000 customers would be a population *parameter*.) In statistical inference, we routinely use sampling theory to connect sample statistics to population parameters. Table 1.2 shows more examples of parameters and statistics.

## Probability Theory: The Link

As we’ve already implied, **probability theory**—the study of likelihood or chance—can be seen as the link between descriptive and inferential statistics. Knowledge of basic probability theory will allow us to connect what we see in a sample to what we would *likely* see in the population being represented by the sample (see Figure 1.4).

**TABLE 1.2**  
**Population Parameters and Sample Statistics**

Population	Population Parameter	Sample	Sample Statistic
All 1200 units in a shipment of electronic components	Percentage of the components in the shipment that are defective	50 units selected from the shipment	Percentage of sample items that are defective
All 1500 employees of Company XYZ	Average years at the company for the 1500 employees	25 employees selected at random	Average years at the company for the 25 employees in the sample
All 130 million US registered voters	Percentage of all registered voters who support income tax reform	1200 voters contacted randomly by telephone	Percentage of sample members who support income tax reform
All 15,000 students at University A; All 20,000 students at University B	Difference in average SAT scores for the two student populations	200 students selected from each of the two university populations	Difference in average SAT scores for students in the two samples



**FIGURE 1.4** The Linking Role of Probability Theory

## 1.3 Two Views of the World

There's no such thing as chance. —*Friedrich Schiller*  
 Chance governs all. —*John Milton*

To put statistical analysis into a somewhat broader context, it's useful to identify two distinct analytic views of the world—two sets of beliefs about “how things are”: a *deterministic* view and a *probabilistic* view.

### A World of Certainty

With a *deterministic* view of the world, the world we see is essentially a world of certainty. It's a world that consists exclusively of one-to-one, cause-and-effect relationships—given some cause, there's a unique and identifiable effect; given some effect, there's a definite, identifiable cause. The world here, while not necessarily simple, is at least well defined. In this world, the direction and trajectory of a golf ball are perfectly predictable, and the stopping point of a spinning roulette wheel is purely a matter of force and friction.

### A World of Uncertainty

In contrast, under a *probabilistic* view of the world, things don't appear quite so straightforward. Here, given some cause, there may well be a number of possible effects; given some effect, there may be a number of possible causes. In this view of reality, there's randomness in the world. Here, the world is dominated by real, and complicating, *uncertainty*. In this world, rain clouds don't always bring rain, and “sure things” sometimes fizzle.

It turns out that for each of these two contrasting worldviews, an appropriate set of analytic tools is available. Tools such as linear programming and differential calculus are often applied to problems encountered in a certain (that is, deterministic) world. By contrast, the tools of probability and statistics, especially inferential statistics, are designed to deal with problems arising in a distinctly uncertain world.

**NOTE:** For centuries there's been active scientific, philosophical, and theological debate over whether the world is completely deterministic (that is, nonrandom) or whether there are truly random or “chance” forces at work in nature. If the world is deterministic—that is, if the world consists solely of perfect cause-and-effect relationships—probability is not really an objective measure of chance or likelihood since, in a deterministic world, there's no such thing as chance. Rather, probability is just an “expression of our ignorance” (in the words of French mathematician Pierre-Simon Laplace) or our response to a lack of information. In this sort of world, the results of tossing a coin are uncertain only for someone unable to apply the laws of physics. With the emergence of quantum mechanics in the twentieth century, the strictly deterministic view of the world was challenged by evidence that randomness is, in fact, an inherent part of nature. Although this is an interesting issue and clearly deserving of discussion, we'll leave it to the philosophers and physicists to sort it out. For whether we view uncertainty as an intrinsic part of the natural world or merely the result of having less than perfect information, understanding the principles of probability and statistics can add significantly to our ability to make informed decisions.

## 1.4 The Nature of Data

### Qualitative and Quantitative Data

To this point we've used the term **data** pretty loosely. In statistics, the term refers specifically to facts or figures that are subject to summarization, analysis, and presentation. A **data set** is a collection of data having some common connection. Data can be either *numeric* or *nonnumeric*. **Numeric data** are data expressed as numbers. **Nonnumeric data** are represented in

other ways, often with words or letters. Telephone numbers and golf scores are examples of numeric data. Nationalities and nicknames are nonnumeric.

Data can also be classified as *qualitative* or *quantitative*. **Qualitative data** are just names and labels for things. For example, when interviewers in a Harris poll record the gender and the ZIP code of poll participants, they're collecting *qualitative* data. Similarly, when participants in the poll are asked whether they agree or disagree with a particular statement, they're providing *qualitative* data. Table 1.3 shows one person's answers to a survey question designed to elicit *qualitative* responses.

Qualitative data are usually nonnumeric—not represented by a number—but can sometimes be numeric. For example, ZIP codes are *numeric* qualitative data since they use numbers just to label, not to count or measure. We could transform nonnumeric qualitative data like gender to a numeric form by letting “1” represent female and “2” represent male, but we'd be using the numbers only as category labels, so the data would remain qualitative.

Whenever we're recording things like age or income or commuting time, we're collecting *quantitative* data. **Quantitative data** represent measurements or counts and are always numeric.

Statistical data are typically the result of successive observations of some characteristic of a group. The characteristic being observed is referred to as a **variable**. Variables associated with qualitative data are **qualitative** (or categorical) **variables**. Variables associated with quantitative data are **quantitative variables**. Gender, student ID number, and marital status, then, are *qualitative* variables; income, age, and IQ are *quantitative* variables.

Table 1.4 shows more examples of qualitative and quantitative variables.

You might take a minute to decide which of the qualitative variables in Table 1.4 are generally *numeric* and which are generally *nonnumeric*.

## Time Series and Cross-Sectional Data

Data collected over time or representing values of a variable in two or more time periods are called **time series data**. If, for example, we show the US cost of living index for each of the years 2007 to 2014, we're dealing with *time series* data. By contrast, a chart indicating the 2010 cost of living index for each of 10 countries involves **cross-sectional data**—data collected at, or representing, the same point in time. Importantly, time series data can create special statistical challenges (for example, how should we compare incomes in 1960 with incomes in 2014?) that aren't generally present in cross-sectional studies.

**TABLE 1.3**  
**One Person's Answers to a Survey Question Seeking a Qualitative Response**

Question: How much confidence do you have in ____?				
	A Great Deal	Only Some	Hardly Any	Not Sure
The military	✓			
Colleges and universities	✓			
Medicine		✓		
The White House		✓		
The U.S. Supreme Court				✓
Organized religion		✓		
The press			✓	
Organized labor		✓		
Congress			✓	
Major companies		✓		
Law firms			✓	

Source: Harris Interactive Survey, harrisinteractive.com.

**TABLE 1.4**  
**Qualitative and Quantitative Variables**

Variable	Qualitative	Quantitative
Years of job seniority		X
Favorite books	X	
Temperature		X
Height		X
Number of siblings		X
College attended	X	
Color of hair	X	
Exchange rate		X
Country of origin	X	
Social Security Number	X	
Occupation	X	
Size of the national debt		X
Political party affiliation	X	
Weight		X
Area code	X	
Apartment number	X	
Golf handicap		X

## Levels of Measurement

The main focus in our text is on numeric data, and whenever numeric data are collected, whether from historical records, observation, or controlled experimentation, the process of measurement—assigning proper values to observed phenomena—is involved. And whenever the process of measurement is involved, the issue of *levels of measurement* becomes important. Since most statistical techniques are suited to data measured only at certain levels, any user of statistical analysis should be aware of the level of measurement involved in his or her data. We can, in fact, describe four distinct measurement levels.

### Nominal Data

**Nominal data** represent the lowest level of measurement. With nominal data, each value serves strictly as a label or a name. For example, a country of origin data set could include the possible values France (designated as country 1), the United States (country 2), Japan (country 3), and so on. With this sort of data, there's no natural ordering of values; the “value” of France (1) is no larger or smaller than the “value” of Japan (3). Each value is just an identifier.

### Ordinal Data

**Ordinal data** represent a step up on the measurement scale. Here, in contrast to the strictly nominal case, values can be meaningfully *rank-ordered*. For example, in the Harris survey cited in Table 1.3), possible responses “A Great Deal,” “Only Some,” and “Hardly Any” could be easily compared and ranked in order of most-to-least favorable. If we assigned a number to each of the possible responses—for example, 1 to “A Great Deal,” 2 to “Only Some,” and 3 to “Hardly Any”—the order, in terms of favorability, is clear: 1 is a higher rank than 2, and 2 is a higher rank than 3. It's important to note, however, that with ordinal data, even though rank ordering is possible, measuring or interpreting the precise difference between successive ranks is normally difficult or impossible. For example, a rank of 2 is higher than a rank of 3, but *how much* higher? “A Great Deal” is a more favorable response than “Only Some,” but *how much* more favorable is it? With ordinal data, there's no reason to believe that the distance between a number 1 ranking and a number 2 ranking