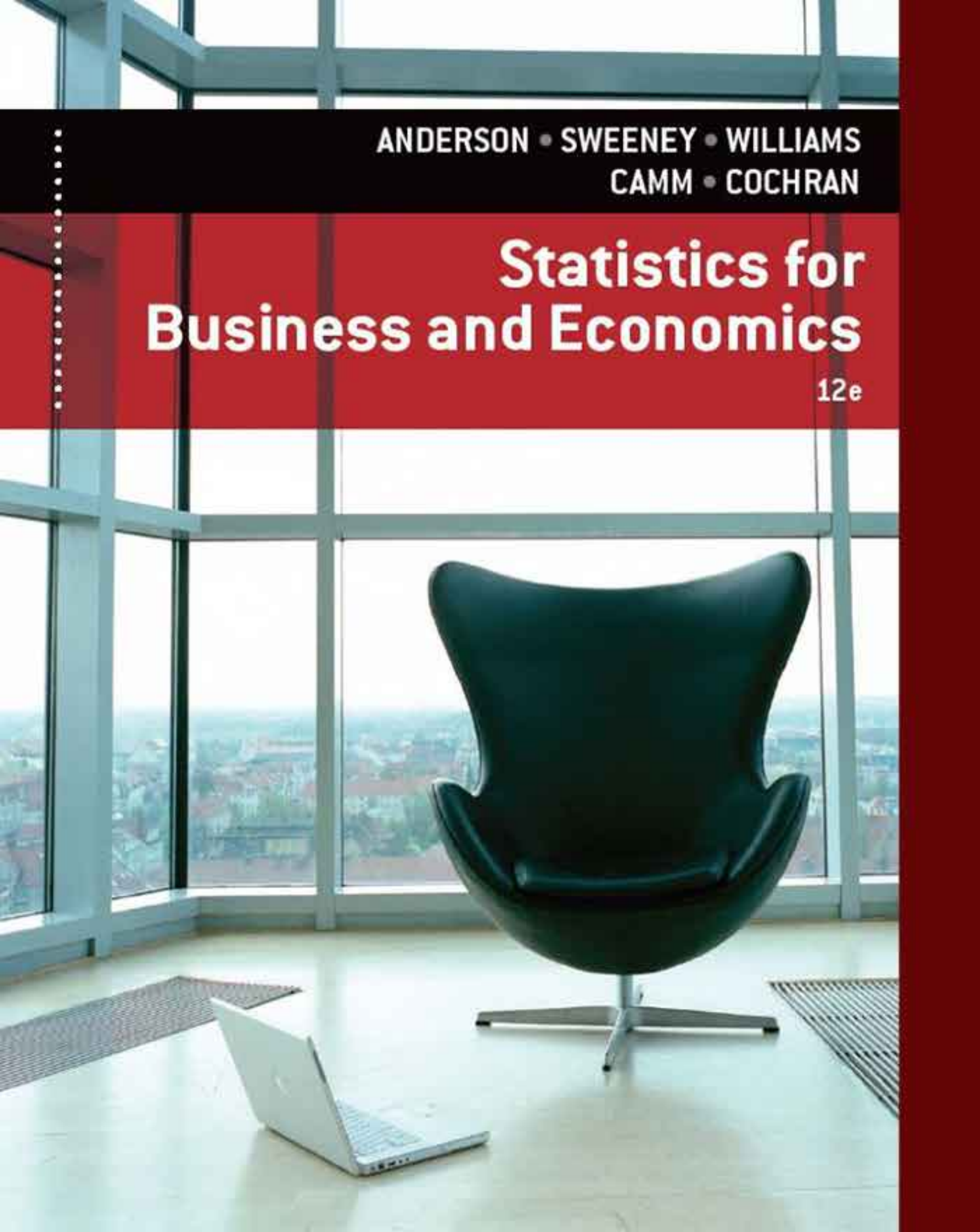


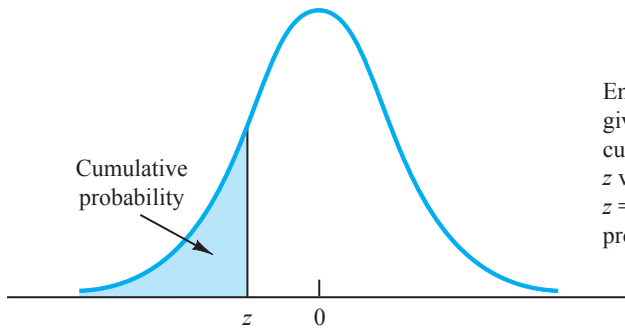
ANDERSON • SWEENEY • WILLIAMS  
CAMM • COCHRAN

# Statistics for Business and Economics

12e



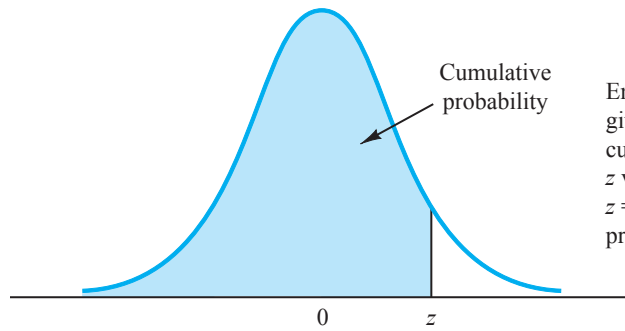
## CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION



Entries in this table give the area under the curve to the left of the  $z$  value. For example, for  $z = -.85$ , the cumulative probability is .1977.

$z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
- .9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
- .8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
- .7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
- .6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
- .5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
- .4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
- .3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
- .2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
- .1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
- .0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

## CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION



Entries in the table give the area under the curve to the left of the  $z$  value. For example, for  $z = 1.25$ , the cumulative probability is .8944.

$z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

This is an electronic version of the print textbook. Due to electronic rights restrictions, some third party content may be suppressed. The publisher reserves the right to remove content from this title at any time if subsequent rights restrictions require it. For valuable information on pricing, previous editions, changes to current editions, and alternate formats, please visit [www.nelson.com](http://www.nelson.com) to search by ISBN#, author, title, or keyword for materials in your areas of interest.

# STATISTICS FOR BUSINESS AND ECONOMICS 12e



# STATISTICS FOR BUSINESS AND ECONOMICS 12e

**David R. Anderson**  
University of Cincinnati

**Dennis J. Sweeney**  
University of Cincinnati

**Thomas A. Williams**  
Rochester Institute of Technology

**Jeffrey D. Camm**  
University of Cincinnati

**James J. Cochran**  
Louisiana Tech University



---

Australia · Brazil · Canada · Mexico · Singapore · Spain · United Kingdom · United States

**Statistics for Business and Economics,  
Twelfth Edition**

David R. Anderson, Dennis J. Sweeney,  
Thomas A. Williams, Jeffrey D. Camm,  
James J. Cochran

Senior Vice President, LRS/Acquisitions  
& Solutions Planning:  
Jack W. Calhoun

Editorial Director, Business & Economics:  
Erin Joyner

Editor-In-Chief: Joe Sabatino

Sr. Acquisitions Editor:  
Charles McCormick, Jr.

Sr. Brand Manager: Kristen Hurd

Developmental Editor:  
Maggie Kubale

Sr. Content Project Manager:  
Tamborah Moore

Media Editor: Chris Valentine

Manufacturing Planner:  
Ron Montgomery

Production Service: MPS Limited

Sr. Art Director: Stacy Jenkins Shirley

Internal Designer:  
Michael Stratton/cmiller design

Cover Designer: Craig Ramsdell

Cover Image:  
Eric O'Connell/Getty Images

Rights Acquisitions Specialist:  
Anne Sheroff

Text permissions researcher:  
Sarah Carey/PMG

Image permissions researcher:  
Sheeja Mohan/PMG

© 2014, 2012 South-Western, Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance, contact us at  
**Cengage Learning Customer & Sales Support, 1-800-354-9706**

For permission to use material from this text or product,  
submit all requests online at [www.cengage.com/permissions](http://www.cengage.com/permissions)  
Further permissions questions can be emailed to  
[permissionrequest@cengage.com](mailto:permissionrequest@cengage.com)

**ExamView®** is a registered trademark of eInstruction Corp.  
Windows is a registered trademark of the Microsoft Corporation  
used herein under license. Macintosh and Power Macintosh are  
registered trademarks of Apple Computer, Inc. used herein under  
license.

© 2008 Cengage Learning. All Rights Reserved.

*Microsoft Excel®* is a registered trademark of Microsoft Corporation.  
© 2014 Microsoft.

Library of Congress Control Number: 2012941311

ISBN-13: 978-1-133-27453-7

ISBN-10: 1-133-27453-6

**South-Western**

5191 Natorp Boulevard  
Mason, OH 45040  
USA

Cengage Learning is a leading provider of customized learning  
solutions with office locations around the globe, including  
Singapore, the United Kingdom, Australia, Mexico, Brazil, and  
Japan. Locate your local office at: [www.cengage.com/global](http://www.cengage.com/global)

Cengage Learning products are represented in Canada by  
Nelson Education, Ltd.

For your course and learning solutions, visit  
[www.cengage.com](http://www.cengage.com)

Purchase any of our products at your local college store or at our  
preferred online store [www.cengagebrain.com](http://www.cengagebrain.com)

Printed in Canada  
1 2 3 4 5 6 7 16 15 14 13 12



*Dedicated to  
Marcia, Cherri, Robbie, Karen, and Teresa*



# Brief Contents

<b>Preface</b>	xxv
<b>About the Authors</b>	xxxi
<b>Chapter 1</b>	<b>Data and Statistics 1</b>
<b>Chapter 2</b>	<b>Descriptive Statistics: Tabular and Graphical Displays 33</b>
<b>Chapter 3</b>	<b>Descriptive Statistics: Numerical Measures 99</b>
<b>Chapter 4</b>	<b>Introduction to Probability 169</b>
<b>Chapter 5</b>	<b>Discrete Probability Distributions 215</b>
<b>Chapter 6</b>	<b>Continuous Probability Distributions 265</b>
<b>Chapter 7</b>	<b>Sampling and Sampling Distributions 298</b>
<b>Chapter 8</b>	<b>Interval Estimation 342</b>
<b>Chapter 9</b>	<b>Hypothesis Tests 382</b>
<b>Chapter 10</b>	<b>Inference About Means and Proportions with Two Populations 441</b>
<b>Chapter 11</b>	<b>Inferences About Population Variances 482</b>
<b>Chapter 12</b>	<b>Comparing Multiple Proportions, Test of Independence and Goodness of Fit 507</b>
<b>Chapter 13</b>	<b>Experimental Design and Analysis of Variance 545</b>
<b>Chapter 14</b>	<b>Simple Linear Regression 598</b>
<b>Chapter 15</b>	<b>Multiple Regression 682</b>
<b>Chapter 16</b>	<b>Regression Analysis: Model Building 751</b>
<b>Chapter 17</b>	<b>Time Series Analysis and Forecasting 800</b>
<b>Chapter 18</b>	<b>Nonparametric Methods 870</b>
<b>Chapter 19</b>	<b>Statistical Methods for Quality Control 916</b>
<b>Chapter 20</b>	<b>Index Numbers 951</b>
<b>Chapter 21</b>	<b>Decision Analysis On Website</b>
<b>Chapter 22</b>	<b>Sample Survey On Website</b>
<b>Appendix A</b>	<b>References and Bibliography 972</b>
<b>Appendix B</b>	<b>Tables 974</b>
<b>Appendix C</b>	<b>Summation Notation 1001</b>
<b>Appendix D</b>	<b>Self-Test Solutions and Answers to Even-Numbered Exercises 1003</b>
<b>Appendix E</b>	<b>Microsoft Excel 2010 and Tools for Statistical Analysis 1064</b>
<b>Appendix F</b>	<b>Computing <math>p</math>-Values Using Minitab and Excel 1076</b>
<b>Index</b>	<b>1080</b>



# Contents

<b>Preface</b>	<b>xxv</b>
<b>About the Authors</b>	<b>xxxi</b>

## **Chapter 1** Data and Statistics 1

### **Statistics in Practice: Bloomberg Businessweek** 2

#### **1.1 Applications in Business and Economics** 3

Accounting	3
Finance	4
Marketing	4
Production	4
Economics	4
Information Systems	5

#### **1.2 Data** 5

Elements, Variables, and Observations	5
Scales of Measurement	7
Categorical and Quantitative Data	8
Cross-Sectional and Time Series Data	8

#### **1.3 Data Sources** 11

Existing Sources	11
Statistical Studies	12
Data Acquisition Errors	14

#### **1.4 Descriptive Statistics** 14

#### **1.5 Statistical Inference** 16

#### **1.6 Computers and Statistical Analysis** 18

#### **1.7 Data Mining** 18

#### **1.8 Ethical Guidelines for Statistical Practice** 19

#### **Summary** 21

#### **Glossary** 21

#### **Supplementary Exercises** 22

#### **Appendix: An Introduction to StatTools** 29

## **Chapter 2** Descriptive Statistics: Tabular and Graphical Displays 33

### **Statistics in Practice: Colgate-Palmolive Company** 34

#### **2.1 Summarizing Data for a Categorical Variable** 35

Frequency Distribution	35
------------------------	----

	Relative Frequency and Percent Frequency Distributions	36
	Bar Charts and Pie Charts	36
<b>2.2</b>	<b>Summarizing Data for a Quantitative Variable</b>	<b>42</b>
	Frequency Distribution	42
	Relative Frequency and Percent Frequency Distributions	43
	Dot Plot	44
	Histogram	44
	Cumulative Distributions	46
	Stem-and-Leaf Display	47
<b>2.3</b>	<b>Summarizing Data for Two Variables Using Tables</b>	<b>55</b>
	Crosstabulation	55
	Simpson's Paradox	58
<b>2.4</b>	<b>Summarizing Data for Two Variables Using Graphical Displays</b>	<b>64</b>
	Scatter Diagram and Trendline	64
	Side-by-Side and Stacked Bar Charts	65
<b>2.5</b>	<b>Data Visualization: Best Practices in Creating Effective Graphical Displays</b>	<b>70</b>
	Creating Effective Graphical Displays	71
	Choosing the Type of Graphical Display	72
	Data Dashboards	72
	Data Visualization in Practice: Cincinnati Zoo and Botanical Garden	74
	<b>Summary</b>	<b>77</b>
	<b>Glossary</b>	<b>78</b>
	<b>Key Formulas</b>	<b>79</b>
	<b>Supplementary Exercises</b>	<b>79</b>
	<b>Case Problem 1: Pelican Stores</b>	<b>84</b>
	<b>Case Problem 2: Motion Picture Industry</b>	<b>85</b>
	<b>Appendix 2.1 Using Minitab for Tabular and Graphical Presentations</b>	<b>86</b>
	<b>Appendix 2.2 Using Excel for Tabular and Graphical Presentations</b>	<b>88</b>
	<b>Appendix 2.3 Using StatTools for Tabular and Graphical Presentations</b>	<b>98</b>
	<b>Chapter 3 Descriptive Statistics: Numerical Measures</b>	<b>99</b>
	<b>Statistics in Practice: Small Fry Design</b>	<b>100</b>
<b>3.1</b>	<b>Measures of Location</b>	<b>101</b>
	Mean	101
	Weighted Mean	103
	Median	104
	Geometric Mean	106
	Mode	107
	Percentiles	108
	Quartiles	109

<b>3.2 Measures of Variability</b>	<b>116</b>
Range	116
Interquartile Range	117
Variance	117
Standard Deviation	118
Coefficient of Variation	119
<b>3.3 Measures of Distribution Shape, Relative Location, and Detecting Outliers</b>	<b>123</b>
Distribution Shape	123
z-Scores	123
Chebyshev's Theorem	125
Empirical Rule	126
Detecting Outliers	127
<b>3.4 Five-Number Summaries and Box Plots</b>	<b>130</b>
Five-Number Summary	131
Box Plot	131
<b>3.5 Measures of Association Between Two Variables</b>	<b>136</b>
Covariance	136
Interpretation of the Covariance	138
Correlation Coefficient	140
Interpretation of the Correlation Coefficient	141
<b>3.6 Data Dashboards: Adding Numerical Measures to Improve Effectiveness</b>	<b>145</b>
<b>Summary</b>	<b>149</b>
<b>Glossary</b>	<b>149</b>
<b>Key Formulas</b>	<b>150</b>
<b>Supplementary Exercises</b>	<b>152</b>
<b>Case Problem 1: Pelican Stores</b>	<b>157</b>
<b>Case Problem 2: Motion Picture Industry</b>	<b>158</b>
<b>Case Problem 3: Business Schools of Asia-Pacific</b>	<b>159</b>
<b>Case Problem 4: Heavenly Chocolates Website Transactions</b>	<b>161</b>
<b>Case Problem 5: African Elephant Populations</b>	<b>162</b>
<b>Appendix 3.1 Descriptive Statistics Using Minitab</b>	<b>163</b>
<b>Appendix 3.2 Descriptive Statistics Using Excel</b>	<b>165</b>
<b>Appendix 3.3 Descriptive Statistics Using StatTools</b>	<b>167</b>
<b>Chapter 4 Introduction to Probability</b>	<b>169</b>
<b>Statistics in Practice: Probability to the Rescue</b>	<b>170</b>
<b>4.1 Experiments, Counting Rules, and Assigning Probabilities</b>	<b>171</b>
Counting Rules, Combinations, and Permutations	172
Assigning Probabilities	176
Probabilities for the KP&L Project	178
<b>4.2 Events and Their Probabilities</b>	<b>181</b>

**4.3 Some Basic Relationships of Probability 185**

Complement of an Event 185

Addition Law 186

**4.4 Conditional Probability 192**

Independent Events 195

Multiplication Law 195

**4.5 Bayes' Theorem 200**

Tabular Approach 203

**Summary 206****Glossary 206****Key Formulas 207****Supplementary Exercises 208****Case Problem: Hamilton County Judges 212****Chapter 5 Discrete Probability Distributions 215****Statistics in Practice: CitiBank 216****5.1 Random Variables 217**

Discrete Random Variables 217

Continuous Random Variables 218

**5.2 Developing Discrete Probability Distributions 220****5.3 Expected Value and Variance 225**

Expected Value 225

Variance 225

**5.4 Bivariate Distributions, Covariance, and Financial Portfolios 230**

A Bivariate Empirical Discrete Probability Distribution 230

Financial Applications 233

Summary 236

**5.5 Binomial Probability Distribution 239**

A Binomial Experiment 240

Martin Clothing Store Problem 241

Using Tables of Binomial Probabilities 245

Expected Value and Variance for the Binomial Distribution 246

**5.6 Poisson Probability Distribution 250**

An Example Involving Time Intervals 250

An Example Involving Length or Distance Intervals 252

**5.7 Hypergeometric Probability Distribution 253****Summary 257****Glossary 258****Key Formulas 258****Supplementary Exercises 260****Appendix 5.1 Discrete Probability Distributions with Minitab 263****Appendix 5.2 Discrete Probability Distributions with Excel 263**



## **Chapter 6** Continuous Probability Distributions 265

### **Statistics in Practice: Procter & Gamble 266**

#### **6.1 Uniform Probability Distribution 267**

Area as a Measure of Probability 268

#### **6.2 Normal Probability Distribution 271**

Normal Curve 271

Standard Normal Probability Distribution 273

Computing Probabilities for Any Normal Probability Distribution 278

Great Tire Company Problem 279

#### **6.3 Normal Approximation of Binomial Probabilities 283**

#### **6.4 Exponential Probability Distribution 287**

Computing Probabilities for the Exponential Distribution 287

Relationship Between the Poisson and Exponential Distributions 288

### **Summary 290**

### **Glossary 291**

### **Key Formulas 291**

### **Supplementary Exercises 291**

### **Case Problem: Specialty Toys 294**

### **Appendix 6.1 Continuous Probability Distributions with Minitab 295**

### **Appendix 6.2 Continuous Probability Distributions with Excel 296**

## **Chapter 7** Sampling and Sampling Distributions 298

### **Statistics in Practice: Meadwestvaco Corporation 299**

#### **7.1 The Electronics Associates Sampling Problem 300**

#### **7.2 Selecting a Sample 301**

Sampling from a Finite Population 301

Sampling from an Infinite Population 303

#### **7.3 Point Estimation 306**

Practical Advice 308

#### **7.4 Introduction to Sampling Distributions 310**

#### **7.5 Sampling Distribution of $\bar{x}$ 312**

Expected Value of  $\bar{x}$  312

Standard Deviation of  $\bar{x}$  313

Form of the Sampling Distribution of  $\bar{x}$  314

Sampling Distribution of  $\bar{x}$  for the EAI Problem 316

Practical Value of the Sampling Distribution of  $\bar{x}$  317

Relationship Between the Sample Size and the Sampling Distribution of  $\bar{x}$  318

#### **7.6 Sampling Distribution of $\bar{p}$ 322**

Expected Value of  $\bar{p}$  323

Standard Deviation of  $\bar{p}$  323

Form of the Sampling Distribution of $\bar{p}$	324
Practical Value of the Sampling Distribution of $\bar{p}$	324
<b>7.7 Properties of Point Estimators</b>	<b>328</b>
Unbiased	328
Efficiency	329
Consistency	330
<b>7.8 Other Sampling Methods</b>	<b>331</b>
Stratified Random Sampling	331
Cluster Sampling	331
Systematic Sampling	332
Convenience Sampling	332
Judgment Sampling	333
<b>Summary</b>	<b>333</b>
<b>Glossary</b>	<b>334</b>
<b>Key Formulas</b>	<b>335</b>
<b>Supplementary Exercises</b>	<b>335</b>
<b>Appendix 7.1 The Expected Value and Standard Deviation of <math>\bar{x}</math></b>	<b>337</b>
<b>Appendix 7.2 Random Sampling with Minitab</b>	<b>339</b>
<b>Appendix 7.3 Random Sampling with Excel</b>	<b>340</b>
<b>Appendix 7.4 Random Sampling with StatTools</b>	<b>341</b>

## **Chapter 8 Interval Estimation 342**

<b>Statistics in Practice: Food Lion</b>	<b>343</b>
<b>8.1 Population Mean: <math>\sigma</math> Known</b>	<b>344</b>
Margin of Error and the Interval Estimate	344
Practical Advice	348
<b>8.2 Population Mean: <math>\sigma</math> Unknown</b>	<b>350</b>
Margin of Error and the Interval Estimate	351
Practical Advice	354
Using a Small Sample	354
Summary of Interval Estimation Procedures	356
<b>8.3 Determining the Sample Size</b>	<b>359</b>
<b>8.4 Population Proportion</b>	<b>362</b>
Determining the Sample Size	364
<b>Summary</b>	<b>367</b>
<b>Glossary</b>	<b>368</b>
<b>Key Formulas</b>	<b>369</b>
<b>Supplementary Exercises</b>	<b>369</b>
<b>Case Problem 1: Young Professional Magazine</b>	<b>372</b>
<b>Case Problem 2: Gulf Real Estate Properties</b>	<b>373</b>
<b>Case Problem 3: Metropolitan Research, Inc.</b>	<b>375</b>
<b>Appendix 8.1 Interval Estimation with Minitab</b>	<b>375</b>

**Appendix 8.2 Interval Estimation Using Excel 377**

**Appendix 8.3 Interval Estimation with StatTools 380**

## **Chapter 9 Hypothesis Tests 382**

**Statistics in Practice: John Morrell & Company 383**

### **9.1 Developing Null and Alternative Hypotheses 384**

The Alternative Hypothesis as a Research Hypothesis 384

The Null Hypothesis as an Assumption to Be Challenged 385

Summary of Forms for Null and Alternative Hypotheses 386

### **9.2 Type I and Type II Errors 387**

### **9.3 Population Mean: $\sigma$ Known 390**

One-Tailed Test 390

Two-Tailed Test 396

Summary and Practical Advice 398

Relationship Between Interval Estimation and Hypothesis Testing 400

### **9.4 Population Mean: $\sigma$ Unknown 405**

One-Tailed Test 405

Two-Tailed Test 406

Summary and Practical Advice 408

### **9.5 Population Proportion 411**

Summary 413

### **9.6 Hypothesis Testing and Decision Making 416**

### **9.7 Calculating the Probability of Type II Errors 417**

### **9.8 Determining the Sample Size for a Hypothesis Test About a Population Mean 422**

Summary 425

Glossary 426

Key Formulas 427

Supplementary Exercises 427

Case Problem 1: Quality Associates, Inc. 430

Case Problem 2: Ethical Behavior of Business Students at  
Bayview University 432

Appendix 9.1 Hypothesis Testing with Minitab 433

Appendix 9.2 Hypothesis Testing with Excel 435

Appendix 9.3 Hypothesis Testing with StatTools 439

## **Chapter 10 Inference About Means and Proportions with Two Populations 441**

**Statistics in Practice: U.S. Food and Drug Administration 442**

### **10.1 Inferences About the Difference Between Two Population Means: $\sigma_1$ and $\sigma_2$ Known 443**

Interval Estimation of  $\mu_1 - \mu_2$  443

	Hypothesis Tests About $\mu_1 - \mu_2$	445
	Practical Advice	447
<b>10.2</b>	<b>Inferences About the Difference Between Two Population Means: <math>\sigma_1</math> and <math>\sigma_2</math> Unknown</b>	<b>450</b>
	Interval Estimation of $\mu_1 - \mu_2$	450
	Hypothesis Tests About $\mu_1 - \mu_2$	452
	Practical Advice	454
<b>10.3</b>	<b>Inferences About the Difference Between Two Population Means: Matched Samples</b>	<b>458</b>
<b>10.4</b>	<b>Inferences About the Difference Between Two Population Proportions</b>	<b>464</b>
	Interval Estimation of $p_1 - p_2$	464
	Hypothesis Tests About $p_1 - p_2$	466
	<b>Summary</b>	<b>470</b>
	<b>Glossary</b>	<b>471</b>
	<b>Key Formulas</b>	<b>471</b>
	<b>Supplementary Exercises</b>	<b>472</b>
	<b>Case Problem: Par, Inc.</b>	<b>475</b>
	<b>Appendix 10.1 Inferences About Two Populations Using Minitab</b>	<b>476</b>
	<b>Appendix 10.2 Inferences About Two Populations Using Excel</b>	<b>478</b>
	<b>Appendix 10.3 Inferences About Two Populations Using StatTools</b>	<b>479</b>
	<b>Chapter 11 Inferences About Population Variances</b>	<b>482</b>
	<b>Statistics in Practice: U.S. Government Accountability Office</b>	<b>483</b>
<b>11.1</b>	<b>Inferences About a Population Variance</b>	<b>484</b>
	Interval Estimation	484
	Hypothesis Testing	488
<b>11.2</b>	<b>Inferences About Two Population Variances</b>	<b>494</b>
	<b>Summary</b>	<b>501</b>
	<b>Key Formulas</b>	<b>501</b>
	<b>Supplementary Exercises</b>	<b>501</b>
	<b>Case Problem: Air Force Training Program</b>	<b>503</b>
	<b>Appendix 11.1 Population Variances with Minitab</b>	<b>504</b>
	<b>Appendix 11.2 Population Variances with Excel</b>	<b>505</b>
	<b>Appendix 11.3 Single Population Standard Deviation with StatTools</b>	<b>505</b>
	<b>Chapter 12 Comparing Multiple Proportions, Test of Independence and Goodness of Fit</b>	<b>507</b>
	<b>Statistics in Practice: United Way</b>	<b>508</b>
<b>12.1</b>	<b>Testing the Equality of Population Proportions for Three or More Populations</b>	<b>509</b>
	A Multiple Comparison Procedure	514

<b>12.2</b>	<b>Test of Independence</b>	<b>519</b>
<b>12.3</b>	<b>Goodness of Fit Test</b>	<b>527</b>
	Multinomial Probability Distribution	527
	Normal Probability Distribution	530
	<b>Summary</b>	<b>536</b>
	<b>Glossary</b>	<b>536</b>
	<b>Key Formulas</b>	<b>537</b>
	<b>Supplementary Exercises</b>	<b>537</b>
	<b>Case Problem: A Bipartisan Agenda for Change</b>	<b>540</b>
	<b>Appendix 12.1 Chi-Square Tests Using Minitab</b>	<b>541</b>
	<b>Appendix 12.2 Chi-Square Tests Using Excel</b>	<b>542</b>
	<b>Appendix 12.3 Chi-Square Tests Using StatTools</b>	<b>544</b>

## **Chapter 13** Experimental Design and Analysis of Variance 545

	<b>Statistics in Practice: Burke Marketing Services, Inc.</b>	<b>546</b>
<b>13.1</b>	<b>An Introduction to Experimental Design and Analysis of Variance</b>	<b>547</b>
	Data Collection	548
	Assumptions for Analysis of Variance	549
	Analysis of Variance: A Conceptual Overview	549
<b>13.2</b>	<b>Analysis of Variance and the Completely Randomized Design</b>	<b>552</b>
	Between-Treatments Estimate of Population Variance	553
	Within-Treatments Estimate of Population Variance	554
	Comparing the Variance Estimates: The $F$ Test	555
	ANOVA Table	557
	Computer Results for Analysis of Variance	558
	Testing for the Equality of $k$ Population Means: An Observational Study	559
<b>13.3</b>	<b>Multiple Comparison Procedures</b>	<b>563</b>
	Fisher's LSD	563
	Type I Error Rates	566
<b>13.4</b>	<b>Randomized Block Design</b>	<b>569</b>
	Air Traffic Controller Stress Test	570
	ANOVA Procedure	571
	Computations and Conclusions	572
<b>13.5</b>	<b>Factorial Experiment</b>	<b>576</b>
	ANOVA Procedure	578
	Computations and Conclusions	578
	<b>Summary</b>	<b>583</b>
	<b>Glossary</b>	<b>584</b>
	<b>Key Formulas</b>	<b>584</b>
	<b>Supplementary Exercises</b>	<b>586</b>

<b>Case Problem 1: Wentworth Medical Center</b>	<b>591</b>
<b>Case Problem 2: Compensation for Sales Professionals</b>	<b>592</b>
<b>Appendix 13.1 Analysis of Variance with Minitab</b>	<b>592</b>
<b>Appendix 13.2 Analysis of Variance with Excel</b>	<b>594</b>
<b>Appendix 13.3 Analysis of a Completely Randomized Design Using StatTools</b>	<b>597</b>

## **Chapter 14 Simple Linear Regression 598**

### **Statistics in Practice: Alliance Data Systems 599**

<b>14.1 Simple Linear Regression Model</b>	<b>600</b>
Regression Model and Regression Equation	600
Estimated Regression Equation	601
<b>14.2 Least Squares Method</b>	<b>603</b>
<b>14.3 Coefficient of Determination</b>	<b>614</b>
Correlation Coefficient	618
<b>14.4 Model Assumptions</b>	<b>622</b>
<b>14.5 Testing for Significance</b>	<b>623</b>
Estimate of $\sigma^2$	623
<i>t</i> Test	624
Confidence Interval for $\beta_1$	626
<i>F</i> Test	627
Some Cautions About the Interpretation of Significance Tests	629
<b>14.6 Using the Estimated Regression Equation for Estimation and Prediction</b>	<b>632</b>
Interval Estimation	633
Confidence Interval for the Mean Value of <i>y</i>	634
Prediction Interval for an Individual Value of <i>y</i>	635
<b>14.7 Computer Solution</b>	<b>640</b>
<b>14.8 Residual Analysis: Validating Model Assumptions</b>	<b>644</b>
Residual Plot Against <i>x</i>	645
Residual Plot Against $\hat{y}$	646
Standardized Residuals	648
Normal Probability Plot	650
<b>14.9 Residual Analysis: Outliers and Influential Observations</b>	<b>653</b>
Detecting Outliers	653
Detecting Influential Observations	656

### **Summary 661**

### **Glossary 661**

### **Key Formulas 662**

### **Supplementary Exercises 664**

### **Case Problem 1: Measuring Stock Market Risk 671**

### **Case Problem 2: U.S. Department of Transportation 672**

### **Case Problem 3: Selecting a Point-and-Shoot Digital Camera 673**

<b>Case Problem 4: Finding the Best Car Value</b>	<b>674</b>
<b>Appendix 14.1 Calculus-Based Derivation of Least Squares Formulas</b>	<b>675</b>
<b>Appendix 14.2 A Test for Significance Using Correlation</b>	<b>677</b>
<b>Appendix 14.3 Regression Analysis with Minitab</b>	<b>678</b>
<b>Appendix 14.4 Regression Analysis with Excel</b>	<b>678</b>
<b>Appendix 14.5 Regression Analysis Using StatTools</b>	<b>681</b>

## **Chapter 15 Multiple Regression 682**

<b>Statistics in Practice: dunnhumby</b>	<b>683</b>
<b>15.1 Multiple Regression Model</b>	<b>684</b>
Regression Model and Regression Equation	684
Estimated Multiple Regression Equation	684
<b>15.2 Least Squares Method</b>	<b>685</b>
An Example: Butler Trucking Company	686
Note on Interpretation of Coefficients	688
<b>15.3 Multiple Coefficient of Determination</b>	<b>694</b>
<b>15.4 Model Assumptions</b>	<b>698</b>
<b>15.5 Testing for Significance</b>	<b>699</b>
<i>F</i> Test	699
<i>t</i> Test	702
Multicollinearity	703
<b>15.6 Using the Estimated Regression Equation for Estimation and Prediction</b>	<b>706</b>
<b>15.7 Categorical Independent Variables</b>	<b>709</b>
An Example: Johnson Filtration, Inc.	709
Interpreting the Parameters	711
More Complex Categorical Variables	713
<b>15.8 Residual Analysis</b>	<b>717</b>
Detecting Outliers	719
Studentized Deleted Residuals and Outliers	719
Influential Observations	720
Using Cook's Distance Measure to Identify Influential Observations	720
<b>15.9 Logistic Regression</b>	<b>724</b>
Logistic Regression Equation	725
Estimating the Logistic Regression Equation	726
Testing for Significance	728
Managerial Use	729
Interpreting the Logistic Regression Equation	729
Logit Transformation	732
<b>Summary</b>	<b>736</b>
<b>Glossary</b>	<b>736</b>
<b>Key Formulas</b>	<b>737</b>

<b>Supplementary Exercises</b>	<b>739</b>
<b>Case Problem 1: Consumer Research, Inc.</b>	<b>745</b>
<b>Case Problem 2: Predicting Winnings for NASCAR Drivers</b>	<b>746</b>
<b>Case Problem 3: Finding the Best Car Value</b>	<b>747</b>
<b>Appendix 15.1 Multiple Regression with Minitab</b>	<b>748</b>
<b>Appendix 15.2 Multiple Regression with Excel</b>	<b>748</b>
<b>Appendix 15.3 Logistic Regression with Minitab</b>	<b>750</b>
<b>Appendix 15.4 Multiple Regression Analysis Using StatTools</b>	<b>750</b>

## **Chapter 16 Regression Analysis: Model Building 751**

<b>Statistics in Practice: Monsanto Company</b>	<b>752</b>
<b>16.1 General Linear Model</b>	<b>753</b>
Modeling Curvilinear Relationships	753
Interaction	756
Transformations Involving the Dependent Variable	760
Nonlinear Models That Are Intrinsically Linear	763
<b>16.2 Determining When to Add or Delete Variables</b>	<b>767</b>
General Case	769
Use of $p$ -Values	770
<b>16.3 Analysis of a Larger Problem</b>	<b>773</b>
<b>16.4 Variable Selection Procedures</b>	<b>777</b>
Stepwise Regression	777
Forward Selection	778
Backward Elimination	779
Best-Subsets Regression	779
Making the Final Choice	780
<b>16.5 Multiple Regression Approach to Experimental Design</b>	<b>783</b>
<b>16.6 Autocorrelation and the Durbin-Watson Test</b>	<b>788</b>
<b>Summary</b>	<b>792</b>
<b>Glossary</b>	<b>792</b>
<b>Key Formulas</b>	<b>792</b>
<b>Supplementary Exercises</b>	<b>793</b>
<b>Case Problem 1: Analysis of PGA Tour Statistics</b>	<b>796</b>
<b>Case Problem 2: Rating Wines from the Piedmont Region of Italy</b>	<b>797</b>
<b>Appendix 16.1 Variable Selection Procedures with Minitab</b>	<b>798</b>
<b>Appendix 16.2 Variable Selection Procedures Using StatTools</b>	<b>799</b>

## **Chapter 17 Time Series Analysis and Forecasting 800**

<b>Statistics in Practice: Nevada Occupational Health Clinic</b>	<b>801</b>
<b>17.1 Time Series Patterns</b>	<b>802</b>
Horizontal Pattern	802



	Trend Pattern	804
	Seasonal Pattern	804
	Trend and Seasonal Pattern	805
	Cyclical Pattern	805
	Selecting a Forecasting Method	807
<b>17.2</b>	<b>Forecast Accuracy</b>	<b>808</b>
<b>17.3</b>	<b>Moving Averages and Exponential Smoothing</b>	<b>813</b>
	Moving Averages	813
	Weighted Moving Averages	816
	Exponential Smoothing	816
<b>17.4</b>	<b>Trend Projection</b>	<b>823</b>
	Linear Trend Regression	823
	Holt's Linear Exponential Smoothing	828
	Nonlinear Trend Regression	830
<b>17.5</b>	<b>Seasonality and Trend</b>	<b>836</b>
	Seasonality Without Trend	836
	Seasonality and Trend	838
	Models Based on Monthly Data	841
<b>17.6</b>	<b>Time Series Decomposition</b>	<b>845</b>
	Calculating the Seasonal Indexes	846
	Deseasonalizing the Time Series	849
	Using the Deseasonalized Time Series to Identify Trend	851
	Seasonal Adjustments	852
	Models Based on Monthly Data	852
	Cyclical Component	852
	<b>Summary</b>	<b>855</b>
	<b>Glossary</b>	<b>856</b>
	<b>Key Formulas</b>	<b>857</b>
	<b>Supplementary Exercises</b>	<b>857</b>
	<b>Case Problem 1: Forecasting Food and Beverage Sales</b>	<b>861</b>
	<b>Case Problem 2: Forecasting Lost Sales</b>	<b>862</b>
	<b>Appendix 17.1 Forecasting with Minitab</b>	<b>864</b>
	<b>Appendix 17.2 Forecasting with Excel</b>	<b>866</b>
	<b>Appendix 17.3 Forecasting Using StatTools</b>	<b>867</b>
	<b>Chapter 18 Nonparametric Methods</b>	<b>870</b>
	<b>Statistics in Practice: West Shell Realtors</b>	<b>871</b>
<b>18.1</b>	<b>Sign Test</b>	<b>872</b>
	Hypothesis Test About a Population Median	872
	Hypothesis Test with Matched Samples	877
<b>18.2</b>	<b>Wilcoxon Signed-Rank Test</b>	<b>880</b>
<b>18.3</b>	<b>Mann-Whitney-Wilcoxon Test</b>	<b>885</b>
<b>18.4</b>	<b>Kruskal-Wallis Test</b>	<b>895</b>

<b>18.5 Rank Correlation</b>	<b>900</b>
<b>Summary</b>	<b>905</b>
<b>Glossary</b>	<b>905</b>
<b>Key Formulas</b>	<b>906</b>
<b>Supplementary Exercises</b>	<b>907</b>
<b>Appendix 18.1 Nonparametric Methods with Minitab</b>	<b>910</b>
<b>Appendix 18.2 Nonparametric Methods with Excel</b>	<b>912</b>
<b>Appendix 18.3 Nonparametric Methods with StatTools</b>	<b>914</b>

## **Chapter 19 Statistical Methods for Quality Control 916**

### **Statistics in Practice: Dow Chemical Company 917**

#### **19.1 Philosophies and Frameworks 918**

Malcolm Baldrige National Quality Award	919
ISO 9000	919
Six Sigma	919
Quality in the Service Sector	922

#### **19.2 Statistical Process Control 922**

Control Charts	923
$\bar{x}$ Chart: Process Mean and Standard Deviation Known	924
$\bar{x}$ Chart: Process Mean and Standard Deviation Unknown	926
$R$ Chart	929
$p$ Chart	931
$np$ Chart	933
Interpretation of Control Charts	933

#### **19.3 Acceptance Sampling 936**

KALI, Inc.: An Example of Acceptance Sampling	937
Computing the Probability of Accepting a Lot	938
Selecting an Acceptance Sampling Plan	941
Multiple Sampling Plans	943

#### **Summary 944**

#### **Glossary 944**

#### **Key Formulas 945**

#### **Supplementary Exercises 946**

#### **Appendix 19.1 Control Charts with Minitab 948**

#### **Appendix 19.2 Control Charts Using StatTools 949**

## **Chapter 20 Index Numbers 951**

### **Statistics in Practice: U.S. Department of Labor, Bureau of Labor Statistics 952**

#### **20.1 Price Relatives 953**

#### **20.2 Aggregate Price Indexes 953**

#### **20.3 Computing an Aggregate Price Index from Price Relatives 957**

<b>20.4</b>	<b>Some Important Price Indexes</b>	<b>959</b>
	Consumer Price Index	959
	Producer Price Index	959
	Dow Jones Averages	960
<b>20.5</b>	<b>Deflating a Series by Price Indexes</b>	<b>961</b>
<b>20.6</b>	<b>Price Indexes: Other Considerations</b>	<b>964</b>
	Selection of Items	964
	Selection of a Base Period	965
	Quality Changes	965
<b>20.7</b>	<b>Quantity Indexes</b>	<b>965</b>
	<b>Summary</b>	<b>967</b>
	<b>Glossary</b>	<b>967</b>
	<b>Key Formulas</b>	<b>968</b>
	<b>Supplementary Exercises</b>	<b>968</b>

## **Chapter 21** **Decision Analysis** **On Website**

	<b>Statistics in Practice: Ohio Edison Company</b>	<b>21-2</b>
<b>21.1</b>	<b>Problem Formulation</b>	<b>21-3</b>
	Payoff Tables	21-4
	Decision Trees	21-4
<b>21.2</b>	<b>Decision Making with Probabilities</b>	<b>21-5</b>
	Expected Value Approach	21-5
	Expected Value of Perfect Information	21-7
<b>21.3</b>	<b>Decision Analysis with Sample Information</b>	<b>21-13</b>
	Decision Tree	21-14
	Decision Strategy	21-15
	Expected Value of Sample Information	21-18
<b>21.4</b>	<b>Computing Branch Probabilities Using Bayes' Theorem</b>	<b>21-24</b>
	<b>Summary</b>	<b>21-28</b>
	<b>Glossary</b>	<b>21-29</b>
	<b>Key Formulas</b>	<b>21-30</b>
	<b>Supplementary Exercises</b>	<b>21-30</b>
	<b>Case Problem: Lawsuit Defense Strategy</b>	<b>21-33</b>
	<b>Appendix: An Introduction to PrecisionTree</b>	<b>21-34</b>
	<b>Appendix: An Introduction to PrecisionTree</b>	<b>21-34</b>
	<b>Appendix: Self-Test Solutions and Answers to Even-Numbered Exercises</b>	<b>21-39</b>

## **Chapter 22** **Sample Survey** **On Website**

	<b>Statistics in Practice: Duke Energy</b>	<b>22-2</b>
<b>22.1</b>	<b>Terminology Used in Sample Surveys</b>	<b>22-2</b>
<b>22.2</b>	<b>Types of Surveys and Sampling Methods</b>	<b>22-3</b>

<b>22.3</b>	<b>Survey Errors</b>	<b>22-5</b>
	Nonsampling Error	22-5
	Sampling Error	22-5
<b>22.4</b>	<b>Simple Random Sampling</b>	<b>22-6</b>
	Population Mean	22-6
	Population Total	22-7
	Population Proportion	22-8
	Determining the Sample Size	22-9
<b>22.5</b>	<b>Stratified Simple Random Sampling</b>	<b>22-12</b>
	Population Mean	22-12
	Population Total	22-14
	Population Proportion	22-15
	Determining the Sample Size	22-16
<b>22.6</b>	<b>Cluster Sampling</b>	<b>22-21</b>
	Population Mean	22-23
	Population Total	22-24
	Population Proportion	22-25
	Determining the Sample Size	22-26
<b>22.7</b>	<b>Systematic Sampling</b>	<b>22-29</b>
	<b>Summary</b>	<b>22-29</b>
	<b>Glossary</b>	<b>22-30</b>
	<b>Key Formulas</b>	<b>22-30</b>
	<b>Supplementary Exercises</b>	<b>22-34</b>
	<b>Appendix: Self-Test Solutions and Answers to Even-Numbered Exercises</b>	<b>22-37</b>
<b>Appendix A</b>	<b>References and Bibliography</b>	<b>972</b>
<b>Appendix B</b>	<b>Tables</b>	<b>974</b>
<b>Appendix C</b>	<b>Summation Notation</b>	<b>1001</b>
<b>Appendix D</b>	<b>Self-Test Solutions and Answers to Even-Numbered Exercises</b>	<b>1003</b>
<b>Appendix E</b>	<b>Microsoft Excel 2010 and Tools for Statistical Analysis</b>	<b>1064</b>
<b>Appendix F</b>	<b>Computing <math>p</math>-Values Using Minitab and Excel</b>	<b>1076</b>
	<b>Index</b>	<b>1080</b>

# Preface

This text is the 12th edition of *STATISTICS FOR BUSINESS AND ECONOMICS*. With this edition we welcome two eminent scholars to our author team: Jeffrey D. Camm of the University of Cincinnati and James J. Cochran of Louisiana Tech University. Both Jeff and Jim are accomplished teachers, researchers, and practitioners in the fields of statistics and business analytics. Jim is a fellow of the American Statistical Association. You can read more about their accomplishments in the About the Authors section which follows this preface. We believe that the addition of Jeff and Jim as our coauthors will both maintain and improve the effectiveness of *Statistics for Business and Economics*.

The purpose of *Statistics for Business and Economics* is to give students, primarily those in the fields of business administration and economics, a conceptual introduction to the field of statistics and its many applications. The text is applications oriented and written with the needs of the nonmathematician in mind; the mathematical prerequisite is knowledge of algebra.

Applications of data analysis and statistical methodology are an integral part of the organization and presentation of the text material. The discussion and development of each technique is presented in an application setting, with the statistical results providing insights to decisions and solutions to problems.

Although the book is applications oriented, we have taken care to provide sound methodological development and to use notation that is generally accepted for the topic being covered. Hence, students will find that this text provides good preparation for the study of more advanced statistical material. A bibliography to guide further study is included as an appendix.

The text introduces the student to the software packages of Minitab 16 and Microsoft® Office Excel 2010 and emphasizes the role of computer software in the application of statistical analysis. Minitab is illustrated as it is one of the leading statistical software packages for both education and statistical practice. Excel is not a statistical software package, but the wide availability and use of Excel make it important for students to understand the statistical capabilities of this package. Minitab and Excel procedures are provided in appendixes so that instructors have the flexibility of using as much computer emphasis as desired for the course. StatTools, a commercial Excel add-in developed by Palisade Corporation, extends the range of statistical options for Excel users. We show how to download and install StatTools in an appendix to Chapter 1, and most chapters include a chapter appendix that shows the steps required to accomplish a statistical procedure using StatTools. We have made the use of StatTools optional so that instructors who want to teach using only the standard tools available in Excel can do so.

## Changes in the Twelfth Edition

We appreciate the acceptance and positive response to the previous editions of *Statistics for Business and Economics*. Accordingly, in making modifications for this new edition, we have maintained the presentation style and readability of those editions. There have been many changes made throughout the text to enhance its educational effectiveness. The most significant changes in the new edition are summarized here.

### Content Revisions

- **Descriptive Statistics—Chapters 2 and 3.** We have significantly revised these chapters to incorporate new material on data visualization, best practices, and much more. Chapter 2 has been reorganized to include new material on side-by-side and

stacked bar charts and a new section has been added on data visualization and best practices in creating effective displays. Chapter 3 now includes coverage of the geometric mean in the section on measures of location. The geometric mean has many applications in the computation of growth rates for financial assets, annual percentage rates, and so on. Chapter 3 also includes a new section on data dashboards and how summary statistics can be incorporated to enhance their effectiveness.

- **Discrete Probability Distributions—Chapter 5.** The introductory material in this chapter has been revised to explain better the role of probability distributions and to show how the material on assigning probabilities in Chapter 4 can be used to develop discrete probability distributions. We point out that the empirical discrete probability distribution is developed by using the relative frequency method to assign probabilities. At the request of many users, we have added a new section (Section 5.4) which covers bivariate discrete distributions and financial applications. We show how financial portfolios can be constructed and analyzed using these distributions.
- **Comparing Multiple Proportions, Tests of Independence, and Goodness of Fit—Chapter 12.** This chapter has undergone a major revision. We have added a new section on testing the equality of three or more population proportions. This section includes a procedure for making multiple comparison tests between all pairs of population proportions. The section on the test of independence has been rewritten to clarify that the test concerns the independence of two categorical variables. Revised appendixes with step-by-step instructions for Minitab, Excel, and StatTools are included.
- **New Case Problems.** We have added 8 new case problems to this edition; the total number of cases is 31. Three new descriptive statistics cases have been added to Chapters 2 and 3. Five new case problems involving regression appear in Chapters 14, 15, and 16. These case problems provide students with the opportunity to analyze larger data sets and prepare managerial reports based on the results of their analysis.
- **New Statistics in Practice Applications.** Each chapter begins with a Statistics in Practice vignette that describes an application of the statistical methodology to be covered in the chapter. New to this edition is a Statistics in Practice for Chapter 2 describing the use of data dashboards and data visualization at the Cincinnati Zoo. We have also added a new Statistics in Practice to Chapter 4 describing how a NASA team used probability to assist the rescue of 33 Chilean miners trapped by a cave-in.
- **New Examples and Exercises based on Real Data.** We continue to make a significant effort to update our text examples and exercises with the most current real data and referenced sources of statistical information. In this edition, we have added approximately 180 new examples and exercises based on real data and referenced sources. Using data from sources also used by *The Wall Street Journal*, *USA Today*, *Barron's*, and others, we have drawn from actual studies to develop explanations and to create exercises that demonstrate the many uses of statistics in business and economics. We believe that the use of real data helps generate more student interest in the material and enables the student to learn about both the statistical methodology and its application. The twelfth edition contains over 350 examples and exercises based on real data.

## Features and Pedagogy

Authors Anderson, Sweeney, Williams, Camm, and Cochran have continued many of the features that appeared in previous editions. Important ones for students are noted here.

### Methods Exercises and Applications Exercises

The end-of-section exercises are split into two parts, Methods and Applications. The Methods exercises require students to use the formulas and make the necessary computations.

The Applications exercises require students to use the chapter material in real-world situations. Thus, students first focus on the computational “nuts and bolts” and then move on to the subtleties of statistical application and interpretation.

## Self-Test Exercises

Certain exercises are identified as “Self-Test Exercises.” Completely worked-out solutions for these exercises are provided in Appendix D. Students can attempt the Self-Test Exercises and immediately check the solution to evaluate their understanding of the concepts presented in the chapter.

## Margin Annotations and Notes and Comments

Margin annotations that highlight key points and provide additional insights for the student are a key feature of this text. These annotations, which appear in the margins, are designed to provide emphasis and enhance understanding of the terms and concepts being presented in the text.

At the end of many sections, we provide Notes and Comments designed to give the student additional insights about the statistical methodology and its application. Notes and Comments include warnings about or limitations of the methodology, recommendations for application, brief descriptions of additional technical considerations, and other matters.

## Data Files Accompany the Text

Over 200 data files are available on the website that accompanies the text. The data sets are available in both Minitab and Excel formats. Webfile logos are used in the text to identify the data sets that are available on the website. Data sets for all case problems as well as data sets for larger exercises are included.

## Acknowledgments

We would like to acknowledge the work of our reviewers, who provided comments and suggestions of ways to continue to improve our text. Thanks to

AbouEl-Makarim Aboueissa, University of Southern Maine	Thomas A. Dahlstrom Eastern University	Clifford B. Hawley West Virginia University
Kathleen Arano Fort Hays State University	Anne Drougas Dominican University	Vance A. Hughey Western Nevada College
Musa Ayar Uw-baraboo/Sauk County	Fesseha Gebremikael Strayer University/ Calhoun Community College	Tony Hunnicutt Ouachita Technical College
Kathleen Burke SUNY Cortland	Malcolm C. Gold University of Wisconsin— Marshfield/Wood County	Stacey M. Jones Albers School of Business and Economics, Seattle University
YC Chang University of Notre Dame	Joel Goldstein Western Connecticut State University	Dukpa Kim University of Virginia
David Chen Rosemont College and Saint Joseph’s University	Jim Grant Lewis & Clark College	Rajaram Krishnan Earlham College
Margaret E. Cochran Northwestern State University of Louisiana	Reidar Hagtvedt University of Alberta School of Business	Robert J. Lemke Lake Forest College
		Philip J. Mizzi Arizona State University

Mehdi Mohaghegh Norwich University	Scott A. Redenius Brandeis University	Ronald Stunda Valdosta State University
Mihail Motzev Walla Walla University	Sandra Robertson Thomas Nelson Community College	Cindy van Es Cornell University
Somnath Mukhopadhyay The University of Texas at El Paso	Sunil Sapra California State University, Los Angeles	Jennifer VanGilder Ursinus College
Kenneth E. Murphy Chapman University	Kyle Vann Scott Snead State Community College	Jacqueline Wroughton Northern Kentucky University
Ogbonnaya John Nwoha Grambling State University	Rodney E. Stanley Tennessee State University	Dmitry Yarushkin Grand View University
Claudiney Pereira Tulane University	Jennifer Strehler Oakton Community College	David Zimmer Western Kentucky University
J. G. Pitt University of Toronto		

We continue to owe a debt to our many colleagues and friends for their helpful comments and suggestions in the development of this and earlier editions of our text. Among them are:

Mohammad Ahmadi University of Tennessee at Chattanooga	Robert Cochran University of Wyoming	Alan Humphrey University of Rhode Island
Lari Arjomand Clayton College and State University	Robert Collins Marquette University	Ann Hussein Philadelphia College of Textiles and Science
Robert Balough Clarion University	David W. Cravens Texas Christian University	C. Thomas Innis University of Cincinnati
Philip Boudreaux University of Louisiana	Tom Dahlstrom Eastern College	Ben Isselhardt Rochester Institute of Technology
Mike Bourke Houston Baptist University	Gopal Dorai William Patterson University	Jeffery Jarrett University of Rhode Island
James Brannon University of Wisconsin— Oshkosh	Nicholas Farnum California State University—Fullerton	Ronald Klimberg St. Joseph's University
John Bryant University of Pittsburgh	Donald Gren Salt Lake Community College	David A. Kravitz George Mason University
Peter Bryant University of Colorado	Paul Guy California State University—Chico	David Krueger St. Cloud State University
Terri L. Byczkowski University of Cincinnati	Clifford Hawley West Virginia University	John Leschke University of Virginia
Robert Carver Stonehill College	Jim Hightower California State University, Fullerton	Martin S. Levy University of Cincinnati
Richard Claycombe McDaniel College		John S. Loucks St. Edward's University



David Lucking-Reiley Vanderbilt University	Tom Pray Rochester Institute of Technology	Victor Ukpolo Austin Peay State University
Bala Maniam Sam Houston State University	Harold Rahmlow St. Joseph's University	Ebenge Usip Youngstown State University
Don Marx University of Alaska, Anchorage	H. V. Ramakrishna Penn State University at Great Valley	Cindy Van Es Cornell University
Tom McCullough University of California— Berkeley	Tom Ryan Case Western Reserve University	Jack Vaughn University of Texas-El Paso
Ronald W. Michener University of Virginia	Bill Seaver University of Tennessee	Andrew Welki John Carroll University
Glenn Milligan Ohio State University	Alan Smith Robert Morris College	Ari Wijetunga Morehead State University
Mitchell Muesham Sam Houston State University	Willbann Terpening Gonzaga University	J. E. Willis Louisiana State University
Roger Myerson Northwestern University	Ted Tsukahara St. Mary's College of California	Mustafa Yilmaz Northeastern University
Richard O'Connell Miami University of Ohio	Hroki Tsurumi Rutgers University	Gary Yoshimoto St. Cloud State University
Alan Olinsky Bryant College	David Tufte University of New Orleans	Yan Yu University of Cincinnati
Ceyhun Ozgur Valparaiso University		Charles Zimmerman Robert Morris College

We thank our associates from business and industry who supplied the Statistics in Practice features. We recognize them individually by a credit line in each of the articles. We are also indebted to our senior acquisitions editor, Charles McCormick Jr.; our developmental editor, Maggie Kubale; our content project manager, Tamborah Moore; our Project Manager at MPS Limited, Lynn Lustberg; our media editor, Chris Valentine; and others at Cengage South-Western for their editorial counsel and support during the preparation of this text.

*David R. Anderson  
Dennis J. Sweeney  
Thomas A. Williams  
Jeffrey D. Camm  
James J. Cochran*



# About the Authors

**David R. Anderson.** David R. Anderson is Professor of Quantitative Analysis in the College of Business Administration at the University of Cincinnati. Born in Grand Forks, North Dakota, he earned his B.S., M.S., and Ph.D. degrees from Purdue University. Professor Anderson has served as Head of the Department of Quantitative Analysis and Operations Management and as Associate Dean of the College of Business Administration at the University of Cincinnati. In addition, he was the coordinator of the College's first Executive Program.

At the University of Cincinnati, Professor Anderson has taught introductory statistics for business students as well as graduate-level courses in regression analysis, multivariate analysis, and management science. He has also taught statistical courses at the Department of Labor in Washington, D.C. He has been honored with nominations and awards for excellence in teaching and excellence in service to student organizations.

Professor Anderson has coauthored 10 textbooks in the areas of statistics, management science, linear programming, and production and operations management. He is an active consultant in the field of sampling and statistical methods.

**Dennis J. Sweeney.** Dennis J. Sweeney is Professor of Quantitative Analysis and Founder of the Center for Productivity Improvement at the University of Cincinnati. Born in Des Moines, Iowa, he earned a B.S.B.A. degree from Drake University and his M.B.A. and D.B.A. degrees from Indiana University, where he was an NDEA Fellow. Professor Sweeney has worked in the management science group at Procter & Gamble and spent a year as a visiting professor at Duke University. Professor Sweeney served as Head of the Department of Quantitative Analysis and as Associate Dean of the College of Business Administration at the University of Cincinnati.

Professor Sweeney has published more than 30 articles and monographs in the area of management science and statistics. The National Science Foundation, IBM, Procter & Gamble, Federated Department Stores, Kroger, and Cincinnati Gas & Electric have funded his research, which has been published in *Management Science*, *Operations Research*, *Mathematical Programming*, *Decision Sciences*, and other journals.

Professor Sweeney has coauthored 10 textbooks in the areas of statistics, management science, linear programming, and production and operations management.

**Thomas A. Williams.** Thomas A. Williams is Professor of Management Science in the College of Business at Rochester Institute of Technology. Born in Elmira, New York, he earned his B.S. degree at Clarkson University. He did his graduate work at Rensselaer Polytechnic Institute, where he received his M.S. and Ph.D. degrees.

Before joining the College of Business at RIT, Professor Williams served for seven years as a faculty member in the College of Business Administration at the University of Cincinnati, where he developed the undergraduate program in Information Systems and then served as its coordinator. At RIT he was the first chairman of the Decision Sciences Department. He teaches courses in management science and statistics, as well as graduate courses in regression and decision analysis.

Professor Williams is the coauthor of 11 textbooks in the areas of management science, statistics, production and operations management, and mathematics. He has been a consultant for numerous *Fortune* 500 companies and has worked on projects ranging from the use of data analysis to the development of large-scale regression models.

**Jeffrey D. Camm.** Jeffrey D. Camm is Professor of Quantitative Analysis, Head of the Department of Operations, Business Analytics, and Information Systems and College of Business Research Fellow in the Carl H. Lindner College of Business at the University of Cincinnati. Born in Cincinnati, Ohio, he holds a B.S. from Xavier University and a Ph.D. from Clemson University. He has been at the University of Cincinnati since 1984 and has been a visiting scholar at Stanford University and a visiting professor of business administration at the Tuck School of Business at Dartmouth College.

Dr. Camm has published over 30 papers in the general area of optimization applied to problems in operations management. He has published his research in *Science*, *Management Science*, *Operations Research*, *Interfaces*, and other professional journals. At the University of Cincinnati, he was named the Dornoff Fellow of Teaching Excellence and he was the 2006 recipient of the INFORMS Prize for the Teaching of Operations Research Practice. A firm believer in practicing what he preaches, he has served as an operations research consultant to numerous companies and government agencies. From 2005 to 2010 he served as editor-in-chief of *Interfaces* and is currently on the editorial board of *INFORMS Transactions on Education*.

**James J. Cochran.** James J. Cochran is the Bank of Ruston Endowed Research Professor of Quantitative Analysis at Louisiana Tech University. Born in Dayton, Ohio, he earned his B.S., M.S., and M.B.A. degrees from Wright State University and a Ph.D. from the University of Cincinnati. He has been at Louisiana Tech University since 2000 and has been a visiting scholar at Stanford University, Universidad de Talca, and the University of South Africa.

Professor Cochran has published over two dozen papers in the development and application of operations research and statistical methods. He has published his research in *Management Science*, *The American Statistician*, *Communications in Statistics—Theory and Methods*, *European Journal of Operational Research*, *Journal of Combinatorial Optimization*, and other professional journals. He was the 2008 recipient of the INFORMS Prize for the Teaching of Operations Research Practice and the 2010 recipient of the Mu Sigma Rho Statistical Education Award. Professor Cochran was elected to the International Statistics Institute in 2005 and named a Fellow of the American Statistical Association in 2011. A strong advocate for effective operations research and statistics education as a means of improving the quality of applications to real problems, Professor Cochran has organized and chaired teaching effectiveness workshops in Montevideo, Uruguay; Cape Town, South Africa; Cartagena, Colombia; Jaipur, India; Buenos Aires, Argentina; and Nairobi, Kenya. He has served as an operations research consultant to numerous companies and not-for-profit organizations. He currently serves as editor-in-chief of *INFORMS Transactions on Education* and is on the editorial board of *Interfaces*, *the Journal of the Chilean Institute of Operations Research*, and *ORiON*.

# CHAPTER 1

## Data and Statistics

---

### CONTENTS

#### STATISTICS IN PRACTICE:

*BLOOMBERG BUSINESSWEEK*

#### 1.1 APPLICATIONS IN BUSINESS AND ECONOMICS

Accounting

Finance

Marketing

Production

Economics

Information Systems

#### 1.2 DATA

Elements, Variables, and Observations

Scales of Measurement

Categorical and Quantitative Data

Cross-Sectional and Time

Series Data

#### 1.3 DATA SOURCES

Existing Sources

Statistical Studies

Data Acquisition Errors

#### 1.4 DESCRIPTIVE STATISTICS

#### 1.5 STATISTICAL INFERENCE

#### 1.6 COMPUTERS AND STATISTICAL ANALYSIS

#### 1.7 DATA MINING

#### 1.8 ETHICAL GUIDELINES FOR STATISTICAL PRACTICE

## STATISTICS *in* PRACTICE

### *BLOOMBERG BUSINESSWEEK\** NEW YORK, NEW YORK

With a global circulation of more than 1 million, *Bloomberg Businessweek* is one of the most widely read business magazines in the world. Bloomberg's 1700 reporters in 145 service bureaus around the world enable *Bloomberg Businessweek* to deliver a variety of articles of interest to the global business and economic community. Along with feature articles on current topics, the magazine contains articles on international business, economic analysis, information processing, and science and technology. Information in the feature articles and the regular sections helps readers stay abreast of current developments and assess the impact of those developments on business and economic conditions.

Most issues of *Bloomberg Businessweek*, formerly *BusinessWeek*, provide an in-depth report on a topic of current interest. Often, the in-depth reports contain statistical facts and summaries that help the reader understand the business and economic information. For example, the cover story for the March 3, 2011 issue discussed the impact of businesses moving their most important work to cloud computing; the May 30, 2011 issue included a report on the crisis facing the U.S. Postal Service; and the August 1, 2011 issue contained a report on why the debt crisis is even worse than you think. In addition, *Bloomberg Businessweek* provides a variety of statistics about the state of the economy, including production indexes, stock prices, mutual funds, and interest rates.

*Bloomberg Businessweek* also uses statistics and statistical information in managing its own business. For example, an annual survey of subscribers helps the company learn about subscriber demographics, reading habits, likely purchases, lifestyles, and so on. *Bloomberg Businessweek* managers use statistical summaries from the survey to provide better services to subscribers and

\*The authors are indebted to Charlene Trentham, Research Manager, for providing this Statistics in Practice.



*Bloomberg Businessweek* uses statistical facts and summaries in many of its articles. © Kyodo/Photoshot.

advertisers. One recent North American subscriber survey indicated that 90% of *Bloomberg Businessweek* subscribers use a personal computer at home and that 64% of *Bloomberg Businessweek* subscribers are involved with computer purchases at work. Such statistics alert *Bloomberg Businessweek* managers to subscriber interest in articles about new developments in computers. The results of the subscriber survey are also made available to potential advertisers. The high percentage of subscribers using personal computers at home and the high percentage of subscribers involved with computer purchases at work would be an incentive for a computer manufacturer to consider advertising in *Bloomberg Businessweek*.

In this chapter, we discuss the types of data available for statistical analysis and describe how the data are obtained. We introduce descriptive statistics and statistical inference as ways of converting data into meaningful and easily interpreted statistical information.

Frequently, we see the following types of statements in newspapers and magazines:

- United States Department of Labor reported that the unemployment rate fell to 8.2%, the lowest in over three years (*The Washington Post*, April 6, 2012).
- Each American consumes an average of 23.2 quarts of ice cream, ice milk, sherbet, ices, and other commercially produced frozen dairy products per year (makeicecream.com website, April 2, 2012).

- The median selling price of a vacation home is \$121,300 (@CNNMoney, March 29, 2012).
- The Wild Eagle rollercoaster at Dollywood in Pigeon Forge, Tennessee, reaches a maximum speed of 61 miles per hour (*USA Today* website, April 5, 2012).
- The number of registered users of Pinterest, a pinboard-style social photo sharing website, grew 85% between mid-January and mid-February (CNBC, March 29, 2012).
- The Pew Research Center reported that the United States median age of brides at the time of their first marriage is an all-time high of 26.5 years (*Significance*, February 2012).
- Canadians clocked an average of 45 hours online in the fourth quarter of 2011 (CBC News, March 2, 2012).
- The Federal Reserve reported that the average credit card debt is \$5,204 per person (PRWeb website, April 5, 2012).

The numerical facts in the preceding statements (8.2%, 23.2, \$121,300, 61, 85%, 26.5, 45, \$5,204) are called **statistics**. In this usage, the term statistics refers to numerical facts such as averages, medians, percentages, and maximums that help us understand a variety of business and economic situations. However, as you will see, the field, or subject, of statistics involves much more than numerical facts. In a broader sense, statistics is the art and science of collecting, analyzing, presenting, and interpreting data. Particularly in business and economics, the information provided by collecting, analyzing, presenting, and interpreting data gives managers and decision makers a better understanding of the business and economic environment and thus enables them to make more informed and better decisions. In this text, we emphasize the use of statistics for business and economic decision making.

Chapter 1 begins with some illustrations of the applications of statistics in business and economics. In Section 1.2 we define the term *data* and introduce the concept of a data set. This section also introduces key terms such as *variables* and *observations*, discusses the difference between quantitative and categorical data, and illustrates the uses of cross-sectional and time series data. Section 1.3 discusses how data can be obtained from existing sources or through survey and experimental studies designed to obtain new data. The important role that the Internet now plays in obtaining data is also highlighted. The uses of data in developing descriptive statistics and in making statistical inferences are described in Sections 1.4 and 1.5. The last three sections of Chapter 1 provide the role of the computer in statistical analysis, an introduction to data mining, and a discussion of ethical guidelines for statistical practice. A chapter-ending appendix includes an introduction to the add-in StatTools which can be used to extend the statistical options for users of Microsoft Excel.

## 1.1

# Applications in Business and Economics

In today's global business and economic environment, anyone can access vast amounts of statistical information. The most successful managers and decision makers understand the information and know how to use it effectively. In this section, we provide examples that illustrate some of the uses of statistics in business and economics.

## Accounting

Public accounting firms use statistical sampling procedures when conducting audits for their clients. For instance, suppose an accounting firm wants to determine whether the amount of accounts receivable shown on a client's balance sheet fairly represents the actual amount of accounts receivable. Usually the large number of individual accounts receivable makes

reviewing and validating every account too time-consuming and expensive. As common practice in such situations, the audit staff selects a subset of the accounts called a sample. After reviewing the accuracy of the sampled accounts, the auditors draw a conclusion as to whether the accounts receivable amount shown on the client's balance sheet is acceptable.

## Finance

Financial analysts use a variety of statistical information to guide their investment recommendations. In the case of stocks, analysts review financial data such as price/earnings ratios and dividend yields. By comparing the information for an individual stock with information about the stock market averages, an analyst can begin to draw a conclusion as to whether the stock is a good investment. For example, *The Wall Street Journal* (March 19, 2012) reported that the average dividend yield for the S&P 500 companies was 2.2%. Microsoft showed a dividend yield of 2.42%. In this case, the statistical information on dividend yield indicates a higher dividend yield for Microsoft than the average dividend yield for the S&P 500 companies. This and other information about Microsoft would help the analyst make an informed buy, sell, or hold recommendation for Microsoft stock.

## Marketing

Electronic scanners at retail checkout counters collect data for a variety of marketing research applications. For example, data suppliers such as ACNielsen and Information Resources, Inc., purchase point-of-sale scanner data from grocery stores, process the data, and then sell statistical summaries of the data to manufacturers. Manufacturers spend hundreds of thousands of dollars per product category to obtain this type of scanner data. Manufacturers also purchase data and statistical summaries on promotional activities such as special pricing and the use of in-store displays. Brand managers can review the scanner statistics and the promotional activity statistics to gain a better understanding of the relationship between promotional activities and sales. Such analyses often prove helpful in establishing future marketing strategies for the various products.

## Production

Today's emphasis on quality makes quality control an important application of statistics in production. A variety of statistical quality control charts are used to monitor the output of a production process. In particular, an  $x$ -bar chart can be used to monitor the average output. Suppose, for example, that a machine fills containers with 12 ounces of a soft drink. Periodically, a production worker selects a sample of containers and computes the average number of ounces in the sample. This average, or  $x$ -bar value, is plotted on an  $x$ -bar chart. A plotted value above the chart's upper control limit indicates overfilling, and a plotted value below the chart's lower control limit indicates underfilling. The process is termed "in control" and allowed to continue as long as the plotted  $x$ -bar values fall between the chart's upper and lower control limits. Properly interpreted, an  $x$ -bar chart can help determine when adjustments are necessary to correct a production process.

## Economics

Economists frequently provide forecasts about the future of the economy or some aspect of it. They use a variety of statistical information in making such forecasts. For instance, in forecasting inflation rates, economists use statistical information on such indicators as the Producer Price Index, the unemployment rate, and manufacturing capacity utilization. Often these statistical indicators are entered into computerized forecasting models that predict inflation rates.

Applications of statistics such as those described in this section are an integral part of this text. Such examples provide an overview of the breadth of statistical applications. To



supplement these examples, practitioners in the fields of business and economics provided chapter-opening Statistics in Practice articles that introduce the material covered in each chapter. The Statistics in Practice applications show the importance of statistics in a wide variety of business and economic situations.

## Information Systems

Information systems administrators are responsible for the day-to-day operation of an organization's computer networks. A variety of statistical information helps administrators assess the performance of computer networks, including local area networks (LANs), wide area networks (WANs), network segments, intranets, and other data communication systems. Statistics such as the mean number of users on the system, the proportion of time any component of the system is down, and the proportion of bandwidth utilized at various times of the day are examples of statistical information that help the system administrator better understand and manage the computer network.

### 1.2

## Data

**Data** are the facts and figures collected, analyzed, and summarized for presentation and interpretation. All the data collected in a particular study are referred to as the **data set** for the study. Table 1.1 shows a data set containing information for 60 nations that participate in the World Trade Organization. The World Trade Organization encourages the free flow of international trade and provides a forum for resolving trade dispute.

## Elements, Variables, and Observations

**Elements** are the entities on which data are collected. Each nation listed in Table 1.1 is an element with the nation or element name shown in the first column. With 60 nations, the data set contains 60 elements.

A **variable** is a characteristic of interest for the elements. The data set in Table 1.1 includes the following five variables:

- **WTO Status:** The nation's membership status in the World Trade Organization; this can be either as a member or an observer.
- **Per Capita GDP (\$):** The total output of the nation divided by the number of people in the nation; this is commonly used to compare economic productivity of the nations.
- **Trade Deficit (\$1000s):** The difference between total dollar value of the nation's imports and total dollar value of the nation's exports.
- **Fitch Rating:** The nation's sovereign credit rating as appraised by the Fitch Group<sup>1</sup>; the credit ratings range from a high of AAA to a low of F and can be modified by + or -.
- **Fitch Outlook:** An indication of the direction the credit rating is likely to move over the upcoming two years; the outlook can be negative, stable, or positive.

Measurements collected on each variable for every element in a study provide the data. The set of measurements obtained for a particular element is called an **observation**. Referring to Table 1.1, we see that the set of measurements for the first observation (Armenia) is

<sup>1</sup>The Fitch Group is one of three nationally recognized statistical rating organizations designated by the U.S. Securities and Exchange Commission. The other two are Standard and Poor's and Moody's investor service.

**TABLE 1.1** DATA SET FOR 60 NATIONS IN THE WORLD TRADE ORGANIZATION

Nation	WTO Status	Per Capita GDP (\$)	Trade Deficit (\$1000s)	Fitch Rating	Fitch Outlook
Armenia	Member	5,400	2,673,359	BB-	Stable
Australia	Member	40,800	-33,304,157	AAA	Stable
Austria	Member	41,700	12,796,558	AAA	Stable
Azerbaijan	Observer	5,400	-16,747,320	BBB-	Positive
Bahrain	Member	27,300	3,102,665	BBB	Stable
Belgium	Member	37,600	-14,930,833	AA+	Negative
Brazil	Member	11,600	-29,796,166	BBB	Stable
Bulgaria	Member	13,500	4,049,237	BBB-	Positive
Canada	Member	40,300	-1,611,380	AAA	Stable
Cape Verde	Member	4,000	874,459	B+	Stable
Chile	Member	16,100	-14,558,218	A+	Stable
China	Member	8,400	-156,705,311	A+	Stable
Colombia	Member	10,100	-1,561,199	BBB-	Stable
Costa Rica	Member	11,500	5,807,509	BB+	Stable
Croatia	Member	18,300	8,108,103	BBB-	Negative
Cyprus	Member	29,100	6,623,337	BBB	Negative
Czech Republic	Member	25,900	-10,749,467	A+	Positive
Denmark	Member	40,200	-15,057,343	AAA	Stable
Ecuador	Member	8,300	1,993,819	B-	Stable
Egypt	Member	6,500	28,486,933	BB	Negative
El Salvador	Member	7,600	5,019,363	BB	Stable
Estonia	Member	20,200	802,234	A+	Stable
France	Member	35,000	118,841,542	AAA	Stable
Georgia	Member	5,400	4,398,153	B+	Positive
Germany	Member	37,900	-213,367,685	AAA	Stable
Hungary	Member	19,600	-9,421,301	BBB-	Negative
Iceland	Member	38,000	-504,939	BB+	Stable
Ireland	Member	39,500	-59,093,323	BBB+	Negative
Israel	Member	31,000	6,722,291	A	Stable
Italy	Member	30,100	33,568,668	A+	Negative
Japan	Member	34,300	31,675,424	AA	Negative
Kazakhstan	Observer	13,000	-33,220,437	BBB	Positive
Kenya	Member	1,700	9,174,198	B+	Stable
Latvia	Member	15,400	2,448,053	BBB-	Positive
Lebanon	Observer	15,600	13,715,550	B	Stable
Lithuania	Member	18,700	3,359,641	BBB	Positive
Malaysia	Member	15,600	-39,420,064	A-	Stable
Mexico	Member	15,100	1,288,112	BBB	Stable
Peru	Member	10,000	-7,888,993	BBB	Stable
Philippines	Member	4,100	15,667,209	BB+	Stable
Poland	Member	20,100	19,552,976	A-	Stable
Portugal	Member	23,200	21,060,508	BBB-	Negative
South Korea	Member	31,700	-37,509,141	A+	Stable
Romania	Member	12,300	13,323,709	BBB-	Stable
Russia	Observer	16,700	-151,400,000	BBB	Positive
Rwanda	Member	1,300	939,222	B	Stable
Serbia	Observer	10,700	8,275,693	BB-	Stable
Seychelles	Observer	24,700	666,026	B	Stable
Singapore	Member	59,900	-27,110,421	AAA	Stable
Slovakia	Member	23,400	-2,110,626	A+	Stable
Slovenia	Member	29,100	2,310,617	AA-	Negative
South Africa	Member	11,000	3,321,801	BBB+	Stable



Data sets such as Nations are available on the website for this text.

Sweden	Member	40,600	−10,903,251	AAA	Stable
Switzerland	Member	43,400	−27,197,873	AAA	Stable
Thailand	Member	9,700	2,049,669	BBB	Stable
Turkey	Member	14,600	71,612,947	BB+	Positive
UK	Member	35,900	162,316,831	AAA	Negative
Uruguay	Member	15,400	2,662,628	BB	Positive
USA	Member	48,100	784,438,559	AAA	Stable
Zambia	Member	1,600	−1,805,198	B+	Stable

Member, 5,400, 2,673,359, BB−, and Stable. The set of measurements for the second observation (Australia) is Member, 40,800, −33,304,157, AAA, and Stable, and so on. A data set with 60 elements contains 60 observations.

## Scales of Measurement

Data collection requires one of the following scales of measurement: nominal, ordinal, interval, or ratio. The scale of measurement determines the amount of information contained in the data and indicates the most appropriate data summarization and statistical analyses.

When the data for a variable consist of labels or names used to identify an attribute of the element, the scale of measurement is considered a **nominal scale**. For example, referring to the data in Table 1.1, the scale of measurement for the WTO Status variable is nominal because the data “member” and “observer” are labels used to identify the status category for the nation. In cases where the scale of measurement is nominal, a numerical code as well as a nonnumerical label may be used. For example, to facilitate data collection and to prepare the data for entry into a computer database, we might use a numerical code for WTO Status variable by letting 1 denote a member nation in the World Trade Organization and 2 denote an observer nation. The scale of measurement is nominal even though the data appear as numerical values.

The scale of measurement for a variable is considered an **ordinal scale** if the data exhibit the properties of nominal data and in addition, the order or rank of the data is meaningful. For example, referring to the data in Table 1.1, the scale of measurement for the Fitch Rating is ordinal because the rating labels which range from AAA to F can be rank ordered from best credit rating AAA to poorest credit rating F. The rating letters provide the labels similar to nominal data, but in addition, the data can also be ranked or ordered based on the credit rating, which makes the measurement scale ordinal. Ordinal data can also be recorded by a numerical code, for example, your class rank in school.

The scale of measurement for a variable is an **interval scale** if the data have all the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure. Interval data are always numeric. College admission SAT scores are an example of interval-scaled data. For example, three students with SAT math scores of 620, 550, and 470 can be ranked or ordered in terms of best performance to poorest performance in math. In addition, the differences between the scores are meaningful. For instance, student 1 scored  $620 - 550 = 70$  points more than student 2, while student 2 scored  $550 - 470 = 80$  points more than student 3.

The scale of measurement for a variable is a **ratio scale** if the data have all the properties of interval data and the ratio of two values is meaningful. Variables such as distance, height, weight, and time use the ratio scale of measurement. This scale requires that a zero value be included to indicate that nothing exists for the variable at the zero point.

For example, consider the cost of an automobile. A zero value for the cost would indicate that the automobile has no cost and is free. In addition, if we compare the cost of \$30,000 for one automobile to the cost of \$15,000 for a second automobile, the ratio property shows that the first automobile is  $\$30,000/\$15,000 = 2$  times, or twice, the cost of the second automobile.

## Categorical and Quantitative Data

Data can be classified as either categorical or quantitative. Data that can be grouped by specific categories are referred to as **categorical data**. Categorical data use either the nominal or ordinal scale of measurement. Data that use numeric values to indicate how much or how many are referred to as **quantitative data**. Quantitative data are obtained using either the interval or ratio scale of measurement.

A **categorical variable** is a variable with categorical data, and a **quantitative variable** is a variable with quantitative data. The statistical analysis appropriate for a particular variable depends upon whether the variable is categorical or quantitative. If the variable is categorical, the statistical analysis is limited. We can summarize categorical data by counting the number of observations in each category or by computing the proportion of the observations in each category. However, even when the categorical data are identified by a numerical code, arithmetic operations such as addition, subtraction, multiplication, and division do not provide meaningful results. Section 2.1 discusses ways for summarizing categorical data.

Arithmetic operations provide meaningful results for quantitative variables. For example, quantitative data may be added and then divided by the number of observations to compute the average value. This average is usually meaningful and easily interpreted. In general, more alternatives for statistical analysis are possible when data are quantitative. Section 2.2 and Chapter 3 provide ways of summarizing quantitative data.

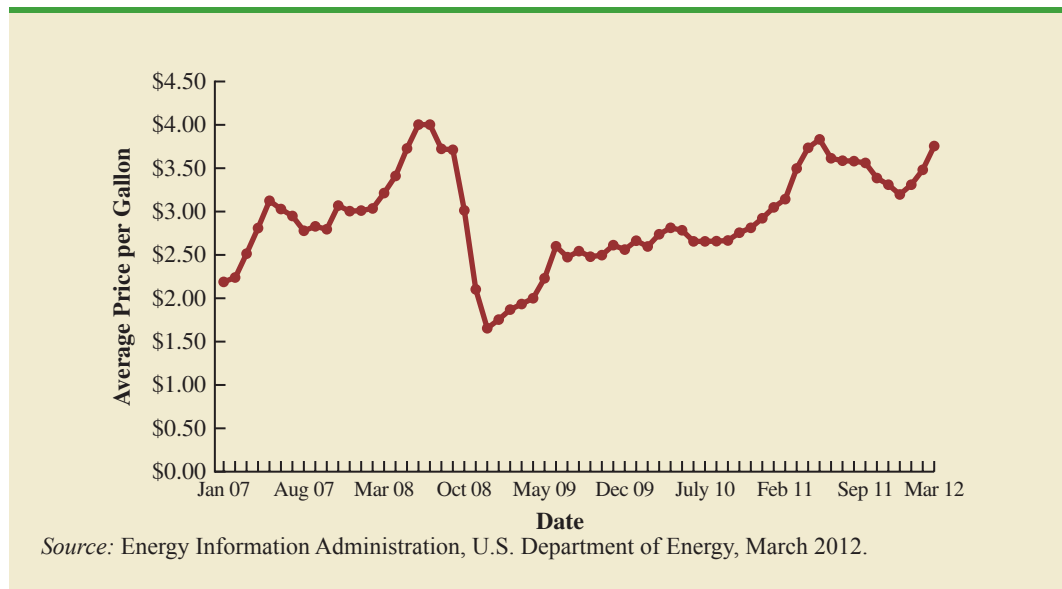
## Cross-Sectional and Time Series Data

For purposes of statistical analysis, distinguishing between cross-sectional data and time series data is important. **Cross-sectional data** are data collected at the same or approximately the same point in time. The data in Table 1.1 are cross-sectional because they describe the five variables for the 60 World Trade Organization nations at the same point in time. **Time series data** are data collected over several time periods. For example, the time series in Figure 1.1 shows the U.S. average price per gallon of conventional regular gasoline between 2007 and 2012. Note that gasoline prices peaked in the summer of 2008 and then dropped sharply in the fall of 2008. Since 2008, the average price per gallon has continued to climb steadily, approaching an all-time high again in 2012.

Graphs of time series data are frequently found in business and economic publications. Such graphs help analysts understand what happened in the past, identify any trends over time, and project future values for the time series. The graphs of time series data can take on a variety of forms, as shown in Figure 1.2. With a little study, these graphs are usually easy to understand and interpret. For example, Panel (A) in Figure 1.2 is a graph that shows the Dow Jones Industrial Average Index from 2002 to 2012. In April 2002, the popular stock market index was near 10,000. Over the next five years the index rose to its all-time high of slightly over 14,000 in October 2007. However, notice the sharp decline in the time series after the high in 2007. By March 2009, poor economic conditions had caused the Dow Jones Industrial Average Index to return to the 7000 level. This was a scary and discouraging period for investors. However, by late 2009, the index was showing a recovery by reaching 10,000. The index has climbed steadily and was above 13,000 in early 2012.

*The statistical method appropriate for summarizing data depends upon whether the data are categorical or quantitative.*

**FIGURE 1.1** U.S. AVERAGE PRICE PER GALLON FOR CONVENTIONAL REGULAR GASOLINE



The graph in Panel (B) shows the net income of McDonald's Inc. from 2005 to 2011. The declining economic conditions in 2008 and 2009 were actually beneficial to McDonald's as the company's net income rose to all-time highs. The growth in McDonald's net income showed that the company was thriving during the economic downturn as people were cutting back on the more expensive sit-down restaurants and seeking less-expensive alternatives offered by McDonald's. McDonald's net income continued to new all-time highs in 2010 and 2011.

Panel (C) shows the time series for the occupancy rate of hotels in South Florida over a one-year period. The highest occupancy rates, 95% and 98%, occur during the months of February and March when the climate of South Florida is attractive to tourists. In fact, January to April of each year is typically the high-occupancy season for South Florida hotels. On the other hand, note the low occupancy rates during the months of August to October, with the lowest occupancy rate of 50% occurring in September. High temperatures and the hurricane season are the primary reasons for the drop in hotel occupancy during this period.

## NOTES AND COMMENTS

1. An observation is the set of measurements obtained for each element in a data set. Hence, the number of observations is always the same as the number of elements. The number of measurements obtained for each element equals the number of variables. Hence, the total number of data items can be determined by multiplying the number of observations by the number of variables.
2. Quantitative data may be discrete or continuous. Quantitative data that measure how many (e.g., number of calls received in 5 minutes) are discrete. Quantitative data that measure how much (e.g., weight or time) are continuous because no separation occurs between the possible data values.