



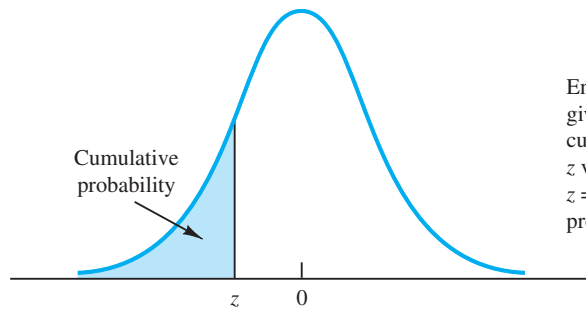
CENGAGE

Anderson Sweeney Williams Camm Cochran Fry Ohlmann

Essentials of Statistics for Business & Economics



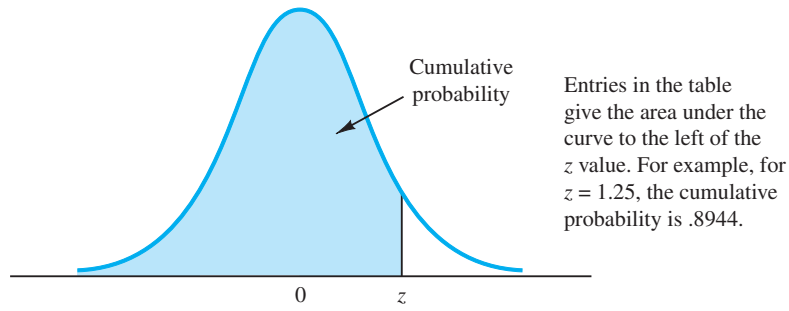
CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION



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



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



Essentials of Statistics for Business and Economics

9e

David R. Anderson
University of Cincinnati

Dennis J. Sweeney
University of Cincinnati

Thomas A. Williams
Rochester Institute
of Technology

Jeffrey D. Camm
Wake Forest University

James J. Cochran
University of Alabama

Michael J. Fry
University of Cincinnati

Jeffrey W. Ohlmann
University of Iowa



***Essentials of Statistics for Business
and Economics, 9e***

David R. Anderson
Dennis J. Sweeney
Thomas A. Williams
Jeffrey D. Camm
James J. Cochran
Michael J. Fry
Jeffrey W. Ohlmann

Senior Vice President, Higher Ed Product,
Content, and Market Development: Erin Joyner

Senior Product Team Manager: Joe Sabatino

Senior Product Manager: Aaron Arnsperger

Project Manager: John Rich

Content Manager: Conor Allen

Product Assistant: Renee Schnee

Marketing Manager: Chris Walz

Production Service: MPS Limited

Designer, Creative Studio: Chris Doughman

Text Designer: Beckmeyer Design

Cover Designer: Beckmeyer Design

Intellectual Property Analyst: Reba Frederics

Intellectual Property Project Manager: Nick
Barrows

© 2020, 2017 Cengage Learning, Inc.

Unless otherwise noted, all content is © Cengage.

WCN: 02-300

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced or distributed in any form or by any means, except as permitted by U.S. copyright law, without the prior written permission of the copyright owner.

For product information and technology assistance, contact us at
**Cengage Customer & Sales Support, 1-800-354-9706 or
support.cengage.com.**

For permission to use material from this text or product,
submit all requests online at **www.cengage.com/permissions.**

Library of Congress Control Number: 2018967096

ISBN: 978-0-357-04543-5

Cengage
20 Channel Center Street
Boston, MA 02210
USA

Cengage is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **www.cengage.com.**

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

To learn more about Cengage platforms and services, register or access your online learning solution, or purchase materials for your course, visit **www.cengage.com.**

Brief Contents

ABOUT THE AUTHORS xix
PREFACE xxiii

CHAPTER 1	Data and Statistics 1
CHAPTER 2	Descriptive Statistics: Tabular and Graphical Displays 33
CHAPTER 3	Descriptive Statistics: Numerical Measures 107
CHAPTER 4	Introduction to Probability 177
CHAPTER 5	Discrete Probability Distributions 223
CHAPTER 6	Continuous Probability Distributions 281
CHAPTER 7	Sampling and Sampling Distributions 319
CHAPTER 8	Interval Estimation 373
CHAPTER 9	Hypothesis Tests 417
CHAPTER 10	Inference About Means and Proportions with Two Populations 481
CHAPTER 11	Inferences About Population Variances 525
CHAPTER 12	Comparing Multiple Proportions, Test of Independence and Goodness of Fit 553
CHAPTER 13	Experimental Design and Analysis of Variance 597
CHAPTER 14	Simple Linear Regression 653
CHAPTER 15	Multiple Regression 731
APPENDIX A	References and Bibliography 800
APPENDIX B	Tables 802
APPENDIX C	Summation Notation 829
APPENDIX D	Answers to Even-Numbered Exercises (MindTap Reader)
APPENDIX E	Microsoft Excel 2016 and Tools for Statistical Analysis 831
APPENDIX F	Computing p -Values with JMP and Excel 839
INDEX	843

Contents

ABOUT THE AUTHORS xix

PREFACE xxiii

CHAPTER 1 Data and Statistics 1

Statistics in Practice: Bloomberg Businessweek 2

1.1 Applications in Business and Economics 3

Accounting 3

Finance 3

Marketing 4

Production 4

Economics 4

Information Systems 4

1.2 Data 5

Elements, Variables, and Observations 5

Scales of Measurement 5

Categorical and Quantitative Data 7

Cross-Sectional and Time Series Data 8

1.3 Data Sources 10

Existing Sources 10

Observational Study 11

Experiment 12

Time and Cost Issues 13

Data Acquisition Errors 13

1.4 Descriptive Statistics 13

1.5 Statistical Inference 15

1.6 Analytics 16

1.7 Big Data and Data Mining 17

1.8 Computers and Statistical Analysis 19

1.9 Ethical Guidelines for Statistical Practice 19

Summary 21

Glossary 21

Supplementary Exercises 22

Appendix 1.1 Opening and Saving DATA Files and Converting to
Stacked form with JMP 30

Appendix 1.2 Getting Started with R and RStudio
(MindTap Reader)

Appendix 1.3 Basic Data Manipulation in R
(MindTap Reader)

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 37

2.2 Summarizing Data for a Quantitative Variable 42

Frequency Distribution 42

Relative Frequency and Percent Frequency Distributions 44

Dot Plot 45

Histogram 45

Cumulative Distributions 47

Stem-and-Leaf Display 47

2.3 Summarizing Data for Two Variables Using Tables 57

Crosstabulation 57

Simpson's Paradox 59

2.4 Summarizing Data for Two Variables Using Graphical Displays 65

Scatter Diagram and Trendline 65

Side-by-Side and Stacked Bar Charts 66

2.5 Data Visualization: Best Practices in Creating Effective Graphical Displays 71

Creating Effective Graphical Displays 71

Choosing the Type of Graphical Display 72

Data Dashboards 73

Data Visualization in Practice: Cincinnati Zoo and Botanical Garden 75

Summary 77

Glossary 78

Key Formulas 79

Supplementary Exercises 80

Case Problem 1: Pelican Stores 85

Case Problem 2: Movie Theater Releases 86

Case Problem 3: Queen City 87

Case Problem 4: Cut-Rate Machining, Inc. 88

Appendix 2.1 Creating Tabular and Graphical Presentations with JMP 90

Appendix 2.2 Creating Tabular and Graphical Presentations with Excel 93

Appendix 2.3 Creating Tabular and Graphical Presentations with R (MindTap Reader)

CHAPTER 3	Descriptive Statistics: Numerical Measures	107
	Statistics in Practice: Small Fry Design	108
3.1	Measures of Location	109
	Mean	109
	Weighted Mean	111
	Median	112
	Geometric Mean	113
	Mode	115
	Percentiles	115
	Quartiles	116
3.2	Measures of Variability	122
	Range	123
	Interquartile Range	123
	Variance	123
	Standard Deviation	125
	Coefficient of Variation	126
3.3	Measures of Distribution Shape, Relative Location, and Detecting Outliers	129
	Distribution Shape	129
	z-Scores	130
	Chebyshev's Theorem	131
	Empirical Rule	132
	Detecting Outliers	134
3.4	Five-Number Summaries and Boxplots	137
	Five-Number Summary	138
	Boxplot	138
	Comparative Analysis Using Boxplots	139
3.5	Measures of Association Between Two Variables	142
	Covariance	142
	Interpretation of the Covariance	144
	Correlation Coefficient	146
	Interpretation of the Correlation Coefficient	147
3.6	Data Dashboards: Adding Numerical Measures to Improve Effectiveness	150
	Summary	153
	Glossary	154
	Key Formulas	155
	Supplementary Exercises	156
	Case Problem 1: Pelican Stores	162
	Case Problem 2: Movie Theater Releases	163
	Case Problem 3: Business Schools of Asia-Pacific	164
	Case Problem 4: Heavenly Chocolates Website Transactions	164

Case Problem 5: African Elephant Populations	166
Appendix 3.1 Descriptive Statistics with JMP	168
Appendix 3.2 Descriptive Statistics with Excel	171
Appendix 3.3 Descriptive Statistics with R (MindTap Reader)	

CHAPTER 4 Introduction to Probability 177

Statistics in Practice: National Aeronautics and Space Administration 178

4.1 Random Experiments, Counting Rules, and Assigning Probabilities	179
Counting Rules, Combinations, and Permutations	180
Assigning Probabilities	184
Probabilities for the KP&L Project	185
4.2 Events and Their Probabilities	189
4.3 Some Basic Relationships of Probability	193
Complement of an Event	193
Addition Law	194
4.4 Conditional Probability	199
Independent Events	202
Multiplication Law	202
4.5 Bayes' Theorem	207
Tabular Approach	210
Summary	212
Glossary	213
Key Formulas	214
Supplementary Exercises	214
Case Problem 1: Hamilton County Judges	219
Case Problem 2: Rob's Market	221

CHAPTER 5 Discrete Probability Distributions 223

Statistics in Practice: Voter Waiting Times in Elections 224

5.1 Random Variables	225
Discrete Random Variables	225
Continuous Random Variables	225
5.2 Developing Discrete Probability Distributions	228
5.3 Expected Value and Variance	233
Expected Value	233
Variance	233
5.4 Bivariate Distributions, Covariance, and Financial Portfolios	238
A Bivariate Empirical Discrete Probability Distribution	238
Financial Applications	241
Summary	244

5.5	Binomial Probability Distribution	247
	A Binomial Experiment	248
	Martin Clothing Store Problem	249
	Using Tables of Binomial Probabilities	253
	Expected Value and Variance for the Binomial Distribution	254
5.6	Poisson Probability Distribution	258
	An Example Involving Time Intervals	259
	An Example Involving Length or Distance Intervals	260
5.7	Hypergeometric Probability Distribution	262
	Summary	265
	Glossary	266
	Key Formulas	266
	Supplementary Exercises	268
	Case Problem 1: <i>Go Bananas!</i> Breakfast Cereal	272
	Case Problem 2: McNeil's Auto Mall	272
	Case Problem 3: Grievance Committee at Tuglar Corporation	273
	Appendix 5.1 Discrete Probability Distributions with JMP	275
	Appendix 5.2 Discrete Probability Distributions with Excel	278
	Appendix 5.3 Discrete Probability Distributions with R (MindTap Reader)	
	CHAPTER 6 Continuous Probability Distributions	281
	Statistics in Practice: Procter & Gamble	282
6.1	Uniform Probability Distribution	283
	Area as a Measure of Probability	284
6.2	Normal Probability Distribution	287
	Normal Curve	287
	Standard Normal Probability Distribution	289
	Computing Probabilities for Any Normal Probability Distribution	294
	Gear Tire Company Problem	294
6.3	Normal Approximation of Binomial Probabilities	299
6.4	Exponential Probability Distribution	302
	Computing Probabilities for the Exponential Distribution	302
	Relationship Between the Poisson and Exponential Distributions	303
	Summary	305
	Glossary	305
	Key Formulas	306
	Supplementary Exercises	306

Case Problem 1: Specialty Toys	309
Case Problem 2: Gebhardt Electronics	311
Appendix 6.1 Continuous Probability Distributions with JMP	312
Appendix 6.2 Continuous Probability Distributions with Excel	317
Appendix 6.3 Continuous Probability Distribution with R (MindTap Reader)	
CHAPTER 7 Sampling and Sampling Distributions	319
Statistics in Practice: Meadwestvaco Corporation	320
7.1 The Electronics Associates Sampling Problem	321
7.2 Selecting a Sample	322
Sampling from a Finite Population	322
Sampling from an Infinite Population	324
7.3 Point Estimation	327
Practical Advice	329
7.4 Introduction to Sampling Distributions	331
7.5 Sampling Distribution of \bar{x}	333
Expected Value of \bar{x}	334
Standard Deviation of \bar{x}	334
Form of the Sampling Distribution of \bar{x}	335
Sampling Distribution of \bar{x} for the EAI Problem	337
Practical Value of the Sampling Distribution of \bar{x}	338
Relationship Between the Sample Size and the Sampling Distribution of \bar{x}	339
7.6 Sampling Distribution of \bar{p}	343
Expected Value of \bar{p}	344
Standard Deviation of \bar{p}	344
Form of the Sampling Distribution of \bar{p}	345
Practical Value of the Sampling Distribution of \bar{p}	345
7.7 Properties of Point Estimators	349
Unbiased	349
Efficiency	350
Consistency	351
7.8 Other Sampling Methods	351
Stratified Random Sampling	352
Cluster Sampling	352
Systematic Sampling	353
Convenience Sampling	353
Judgment Sampling	354

7.9	Big Data and Standard Errors of Sampling Distributions	354
	Sampling Error	354
	Nonsampling Error	355
	Big Data	356
	Understanding What Big Data Is	356
	Implications of Big Data for Sampling Error	357
	Summary	360
	Glossary	361
	Key Formulas	362
	Supplementary Exercises	363
	Case Problem: Marion Dairies	366
	Appendix 7.1 The Expected Value and Standard Deviation of \bar{x}	367
	Appendix 7.2 Random Sampling with JMP	368
	Appendix 7.3 Random Sampling with Excel	371
	Appendix 7.4 Random Sampling with R (MindTap Reader)	
CHAPTER 8	Interval Estimation	373
	Statistics in Practice: Food Lion	374
8.1	Population Mean: σ Known	375
	Margin of Error and the Interval Estimate	375
	Practical Advice	379
8.2	Population Mean: σ Unknown	381
	Margin of Error and the Interval Estimate	382
	Practical Advice	385
	Using a Small Sample	385
	Summary of Interval Estimation Procedures	386
8.3	Determining the Sample Size	390
8.4	Population Proportion	393
	Determining the Sample Size	394
8.5	Big Data and Confidence Intervals	398
	Big Data and the Precision of Confidence Intervals	398
	Implications of Big Data for Confidence Intervals	399
	Summary	401
	Glossary	402
	Key Formulas	402
	Supplementary Exercises	403
	Case Problem 1: <i>Young Professional Magazine</i>	406
	Case Problem 2: <i>Gulf Real Estate Properties</i>	407
	Case Problem 3: <i>Metropolitan Research, Inc.</i>	409

Appendix 8.1 Interval Estimation with JMP	410
Appendix 8.2 Interval Estimation Using Excel	413
Appendix 8.3 Interval Estimation with R (MindTap Reader)	

CHAPTER 9 Hypothesis Tests 417

Statistics in Practice: John Morrell & Company	418
9.1 Developing Null and Alternative Hypotheses	419
The Alternative Hypothesis as a Research Hypothesis	419
The Null Hypothesis as an Assumption to Be Challenged	420
Summary of Forms for Null and Alternative Hypotheses	421
9.2 Type I and Type II Errors	422
9.3 Population Mean: σ Known	425
One-Tailed Test	425
Two-Tailed Test	430
Summary and Practical Advice	433
Relationship Between Interval Estimation and Hypothesis Testing	434
9.4 Population Mean: σ Unknown	439
One-Tailed Test	439
Two-Tailed Test	440
Summary and Practical Advice	441
9.5 Population Proportion	445
Summary	447
9.6 Hypothesis Testing and Decision Making	450
9.7 Calculating the Probability of Type II Errors	450
9.8 Determining the Sample Size for a Hypothesis Test About a Population Mean	455
9.9 Big Data and Hypothesis Testing	459
Big Data, Hypothesis Testing, and p Values	459
Implications of Big Data in Hypothesis Testing	460
Summary	462
Glossary	462
Key Formulas	463
Supplementary Exercises	463
Case Problem 1: Quality Associates, Inc.	467
Case Problem 2: Ethical Behavior of Business Students at Bayview University	469
Appendix 9.1 Hypothesis Testing with JMP	471
Appendix 9.2 Hypothesis Testing with Excel	475
Appendix 9.3 Hypothesis Testing with R (MindTap Reader)	

CHAPTER 10 Inference About Means and Proportions with Two Populations 481

Statistics in Practice: U.S. Food and Drug Administration 482

- 10.1 Inferences About the Difference Between Two Population Means: σ_1 and σ_2 Known 483
 - Interval Estimation of $\mu_1 - \mu_2$ 483
 - Hypothesis Tests About $\mu_1 - \mu_2$ 485
 - Practical Advice 487
- 10.2 Inferences About the Difference Between Two Population Means: σ_1 and σ_2 Unknown 489
 - Interval Estimation of $\mu_1 - \mu_2$ 489
 - Hypothesis Tests About $\mu_1 - \mu_2$ 491
 - Practical Advice 493
- 10.3 Inferences About the Difference Between Two Population Means: Matched Samples 497
- 10.4 Inferences About the Difference Between Two Population Proportions 503
 - Interval Estimation of $p_1 - p_2$ 503
 - Hypothesis Tests About $p_1 - p_2$ 505

Summary 509

Glossary 509

Key Formulas 509

Supplementary Exercises 511

Case Problem: Par, Inc. 514

Appendix 10.1 Inferences About Two Populations with JMP 515

Appendix 10.2 Inferences About Two Populations with Excel 519

Appendix 10.3 Inferences about Two Populations with R
(MindTap Reader)

CHAPTER 11 Inferences About Population Variances 525

Statistics in Practice: U.S. Government Accountability Office 526

- 11.1 Inferences About a Population Variance 527
 - Interval Estimation 527
 - Hypothesis Testing 531
- 11.2 Inferences About Two Population Variances 537

Summary 544

Key Formulas 544

Supplementary Exercises 544

Case Problem 1: Air Force Training Program 546

Case Problem 2: Meticulous Drill & Reamer 547

Appendix 11.1 Population Variances with JMP 549

Appendix 11.2 Population Variances with Excel 551
Appendix 11.3 Population Variances with R (MindTap Reader)

CHAPTER 12 Comparing Multiple Proportions, Test of Independence and Goodness of Fit 553

Statistics in Practice: United Way 554

12.1 Testing the Equality of Population Proportions for Three or More Populations 555

A Multiple Comparison Procedure 560

12.2 Test of Independence 565

12.3 Goodness of Fit Test 573

Multinomial Probability Distribution 573

Normal Probability Distribution 576

Summary 582

Glossary 582

Key Formulas 583

Supplementary Exercises 583

Case Problem 1: A Bipartisan Agenda for Change 587

Case Problem 2: Fuentes Salty Snacks, Inc. 588

Case Problem 3: Fresno Board Games 588

Appendix 12.1 Chi-Square Tests with JMP 590

Appendix 12.2 Chi-Square Tests with Excel 593

Appendix 12.3 Chi-Squared Tests with R (MindTap Reader)

CHAPTER 13 Experimental Design and Analysis of Variance 597

Statistics in Practice: Burke Marketing Services, Inc. 598

13.1 An Introduction to Experimental Design and Analysis of Variance 599

Data Collection 600

Assumptions for Analysis of Variance 601

Analysis of Variance: A Conceptual Overview 601

13.2 Analysis of Variance and the Completely Randomized Design 604

Between-Treatments Estimate of Population Variance 605

Within-Treatments Estimate of Population Variance 606

Comparing the Variance Estimates: The F Test 606

ANOVA Table 608

Computer Results for Analysis of Variance 609

Testing for the Equality of k Population Means:

An Observational Study 610

13.3	Multiple Comparison Procedures	615
	Fisher's LSD	615
	Type I Error Rates	617
13.4	Randomized Block Design	621
	Air Traffic Controller Stress Test	621
	ANOVA Procedure	623
	Computations and Conclusions	623
13.5	Factorial Experiment	627
	ANOVA Procedure	629
	Computations and Conclusions	629
	Summary	635
	Glossary	635
	Key Formulas	636
	Supplementary Exercises	638
	Case Problem 1: Wentworth Medical Center	643
	Case Problem 2: Compensation for Sales Professionals	644
	Case Problem 3: Touristopia Travel	644
	Appendix 13.1 Analysis of Variance with JMP	646
	Appendix 13.2 Analysis of Variance with Excel	649
	Appendix 13.3 Analysis Variance with R (MindTap Reader)	
CHAPTER 14 Simple Linear Regression 653		
	Statistics in Practice: Alliance Data Systems	654
14.1	Simple Linear Regression Model	655
	Regression Model and Regression Equation	655
	Estimated Regression Equation	656
14.2	Least Squares Method	658
14.3	Coefficient of Determination	668
	Correlation Coefficient	671
14.4	Model Assumptions	675
14.5	Testing for Significance	676
	Estimate of σ^2	676
	t Test	677
	Confidence Interval for β_1	679
	F Test	679
	Some Cautions About the Interpretation of Significance Tests	681
14.6	Using the Estimated Regression Equation for Estimation and Prediction	684
	Interval Estimation	685
	Confidence Interval for the Mean Value of y	685
	Prediction Interval for an Individual Value of y	686

- 14.7 Computer Solution 691
- 14.8 Residual Analysis: Validating Model Assumptions 694
 - Residual Plot Against x 695
 - Residual Plot Against \hat{y} 697
 - Standardized Residuals 698
 - Normal Probability Plot 699
- 14.9 Residual Analysis: Outliers and Influential Observations 703
 - Detecting Outliers 703
 - Detecting Influential Observations 704
- 14.10 Practical Advice: Big Data and Hypothesis Testing in Simple Linear Regression 710

Summary 711

Glossary 711

Key Formulas 712

Supplementary Exercises 714

Case Problem 1: Measuring Stock Market Risk 721

Case Problem 2: U.S. Department of Transportation 721

Case Problem 3: Selecting a Point-and-Shoot Digital Camera 722

Case Problem 4: Finding the Best Car Value 723

Case Problem 5: Buckeye Creek Amusement Park 724

Appendix 14.1 Calculus-Based Derivation of Least Squares Formulas 726

Appendix 14.2 A Test for Significance Using Correlation 727

Appendix 14.3 Simple Linear Regression with JMP 727

Appendix 14.4 Regression Analysis with Excel 728

Appendix 14.5 Simple Linear Regression with R (MindTap Reader)

CHAPTER 15 Multiple Regression 731

Statistics in Practice: 84.51° 732

15.1 Multiple Regression Model 733

Regression Model and Regression Equation 733

Estimated Multiple Regression Equation 733

15.2 Least Squares Method 734

An Example: Butler Trucking Company 735

Note on Interpretation of Coefficients 737

15.3 Multiple Coefficient of Determination 743

15.4 Model Assumptions 746

15.5 Testing for Significance 747

F Test 747

t Test 750

Multicollinearity 750

15.6	Using the Estimated Regression Equation for Estimation and Prediction	753
15.7	Categorical Independent Variables	755
	An Example: Johnson Filtration, Inc.	756
	Interpreting the Parameters	758
	More Complex Categorical Variables	760
15.8	Residual Analysis	764
	Detecting Outliers	766
	Studentized Deleted Residuals and Outliers	766
	Influential Observations	767
	Using Cook's Distance Measure to Identify Influential Observations	767
15.9	Logistic Regression	771
	Logistic Regression Equation	772
	Estimating the Logistic Regression Equation	773
	Testing for Significance	774
	Managerial Use	775
	Interpreting the Logistic Regression Equation	776
	Logit Transformation	778
15.10	Practical Advice: Big Data and Hypothesis Testing in Multiple Regression	782
	Summary	783
	Glossary	783
	Key Formulas	784
	Supplementary Exercises	786
	Case Problem 1: Consumer Research, Inc.	790
	Case Problem 2: Predicting Winnings for NASCAR Drivers	791
	Case Problem 3: Finding the Best Car Value	792
	Appendix 15.1 Multiple Linear Regression with JMP	794
	Appendix 15.2 Logistic Regression with JMP	796
	Appendix 15.3 Multiple Regression with Excel	797
	Appendix 15.4 Multiple Linear Regression with R (MindTap Reader)	
	Appendix 15.5 Logistics Regression with R (MindTap Reader)	
	APPENDIX A References and Bibliography	800
	APPENDIX B Tables	802

APPENDIX C	Summation Notation	829
APPENDIX D	Answers to Even-Numbered Exercises (MindTap Reader)	
APPENDIX E	Microsoft Excel 2016 and Tools for Statistical Analysis	831
APPENDIX F	Computing p -Values with JMP and Excel	839
INDEX		843

About the Authors

David R. Anderson. David R. Anderson is Professor Emeritus 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 Emeritus 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 Emeritus 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 the Inmar Presidential Chair and Associate Dean of Analytics in the School of Business at Wake Forest University. Born in Cincinnati, Ohio, he holds a B.S. from Xavier University (Ohio) and a Ph.D. from Clemson University. Prior to

joining the faculty at Wake Forest, he was on the faculty of the University of Cincinnati. He has also 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 40 papers in the general area of optimization applied to problems in operations management and marketing. He has published his research in *Science*, *Management Science*, *Operations Research*, *Interfaces*, and other professional journals. Dr. Camm was named the Dornoff Fellow of Teaching Excellence at the University of Cincinnati 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*. In 2017, he was named an INFORMS Fellow.

James J. Cochran. James J. Cochran is Professor of Applied Statistics and the Rogers-Spivey Faculty Fellow at the University of Alabama. 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 the University of Alabama since 2014 and has been a visiting scholar at Stanford University, Universidad de Talca, the University of South Africa, and Pole Universitaire Leonard de Vinci.

Professor Cochran has published over 40 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*, *Annals of Operations Research*, *European Journal of Operational Research*, *Journal of Combinatorial Optimization*, *Interfaces*, *Statistics and Probability Letters*, 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. He received the Founders Award in 2014 and the Karl E. Peace Award in 2015 from the American Statistical Association. In 2017 he received the American Statistical Association's Waller Distinguished Teaching Career Award and was named a Fellow of INFORMS, and in 2018 he received the INFORMS President's Award.

A strong advocate for effective statistics and operations research 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; Nairobi, Kenya; Buea, Cameroon; Kathmandu, Nepal; Osijek, Croatia; Havana, Cuba; Ulaanbaatar, Mongolia; and Chişinău, Moldova. He has served as an operations research consultant to numerous companies and not-for-profit organizations. He served as editor-in-chief of *INFORMS Transactions on Education* from 2006 to 2012 and is on the editorial board of *Interfaces*, *International Transactions in Operational Research*, and *Significance*.

Michael J. Fry. Michael J. Fry is Professor of Operations, Business Analytics, and Information Systems and Academic Director of the Center for Business Analytics in the Carl H. Lindner College of Business at the University of Cincinnati. Born in Killeen, Texas, he earned a BS from Texas A&M University and M.S.E. and Ph.D. degrees from the University of Michigan. He has been at the University of Cincinnati since 2002, where he was previously Department Head and has been named a Lindner Research Fellow. He has also been a visiting professor at the Samuel Curtis Johnson Graduate School of Management at Cornell University and the Sauder School of Business at the University of British Columbia.

Professor Fry has published more than 25 research papers in journals such as *Operations Research*, *M&SOM*, *Transportation Science*, *Naval Research Logistics*, *IIE Transactions*, *Critical Care Medicine* and *Interfaces*. His research interests are in applying quantitative

management methods to the areas of supply chain analytics, sports analytics, and public-policy operations. He has worked with many different organizations for his research, including Dell, Inc., Starbucks Coffee Company, Great American Insurance Group, the Cincinnati Fire Department, the State of Ohio Election Commission, the Cincinnati Bengals, and the Cincinnati Zoo & Botanical Garden. He was named a finalist for the Daniel H. Wagner Prize for Excellence in Operations Research Practice, and he has been recognized for both his research and teaching excellence at the University of Cincinnati.

Jeffrey W. Ohlmann. Jeffrey W. Ohlmann is Associate Professor of Management Sciences and Huneke Research Fellow in the Tippie College of Business at the University of Iowa. Born in Valentine, Nebraska, he earned a B.S. from the University of Nebraska, and MS and Ph.D. degrees from the University of Michigan. He has been at the University of Iowa since 2003.

Professor Ohlmann's research on the modeling and solution of decision-making problems has produced more than 20 research papers in journals such as *Operations Research*, *Mathematics of Operations Research*, *INFORMS Journal on Computing*, *Transportation Science*, the *European Journal of Operational Research*, and *Interfaces*. He has collaborated with companies such as Transfreight, LeanCor, Cargill, the Hamilton County Board of Elections, and three National Football League franchises. Because of the relevance of his work to industry, he was bestowed the George B. Dantzig Dissertation Award and was recognized as a finalist for the Daniel H. Wagner Prize for Excellence in Operations Research Practice.

Preface

This text is the 9th edition of *ESSENTIALS OF STATISTICS FOR BUSINESS AND ECONOMICS*. In this edition, we include procedures for statistical analysis using Excel 2016 and JMP Student Edition 14. In MindTap Reader, we also include instructions for using the exceptionally popular open-source language R to perform statistical analysis. We are excited to introduce two new coauthors, Michael J. Fry of the University of Cincinnati and Jeffrey W. Ohlmann of the University of Iowa. Both are accomplished teachers and researchers. More details on their backgrounds may be found in the About the Authors section.

The remainder of this preface describes the authors' objectives in writing *ESSENTIALS OF STATISTICS FOR BUSINESS AND ECONOMICS* and the major changes that were made in developing the 9th edition. The purpose of the text 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 understanding 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 JMP Student Edition 14 and Microsoft® Office Excel 2016 and emphasizes the role of computer software in the application of statistical analysis. JMP 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. JMP and Excel procedures are provided in appendices so that instructors have the flexibility of using as much computer emphasis as desired for the course. MindTap Reader includes appendices for using R for statistical analysis. R is an open-source programming language that is widely used in practice to perform statistical analysis. The use of R typically requires more training than the use of software such as JMP or Excel, but the software is extremely powerful. To ease students' introduction to the R language, we also use RStudio which provides an integrated development environment for R.

Changes in the 9th Edition

We appreciate the acceptance and positive response to the previous editions of *Essentials 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 substantial changes in the new edition are summarized here.

Content Revisions

- **Software.** In addition to step-by-step instructions in the software appendices for Excel 2016, we also provide instructions for JMP Student Edition 14 and R. This provides students exposure to and experience with the current versions of several of the most commonly used software for statistical analysis in business. Excel 2016 and

JMP appendices are contained within the textbook chapters, while R appendices are provided in MindTap Reader. In this latest edition, we no longer provide discussion of the use of Minitab.

- **Case Problems.** We have added 12 new case problems in this edition; the total number of cases is now 42. One new case on graphical display has been added to Chapter 2. Two new cases using discrete probability distributions have been added to Chapter 5, and one new case using continuous probability distributions has been added to Chapter 6. A new case on hypothesis testing has been added to Chapter 11, and two new cases on testing proportions have been added to Chapter 12. The 42 case problems in this book provide students the opportunity to work on more complex problems, analyze larger data sets, and prepare managerial reports based on the results of their analyses.
- **Examples and Exercises Based on Real Data.** In this edition, we have added headers to all Applications exercises to make the application of each problem more obvious. We continue to make a substantial effort to update our text examples and exercises with the most current real data and referenced sources of statistical information. We have added more than 160 new examples and exercises based on real data and referenced sources. Using data from sources also used by *The Wall Street Journal*, *USA Today*, *The Financial Times*, and others, we have drawn from actual studies and applications to develop explanations and create exercises that demonstrate the many uses of statistics in business and economics. We believe that the use of real data from interesting and relevant problems helps generate more student interest in the material and enables the student to learn about both statistical methodology and its application. The 9th edition contains more than 350 examples and exercises based on real data.

Features and Pedagogy

Authors Anderson, Sweeney, Williams, Camm, Cochran, Fry, and Ohlmann 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.

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 accompany this text. Data files are provided in Excel format and step-by-step instructions on how to open Excel files in JMP are provided in Appendix 1.1. Files

for use with R are provided in comma-separated-value (CSV) format for easy loading into the R environment. Step-by-step instructions for importing CSV files into R are provided in MindTap Reader Appendix R 1.2.

The data files can be accessed from WebAssign within the resources section, directly within the MindTap Reader by clicking on the DATAfile icon, or online directly at www.cengage.com/decisionsciences/anderson/sbe/14e.

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	Reidar Hagtvedt University of Alberta School of Business	Claudiney Pereira Tulane University
Kathleen Arano Fort Hays State University	Clifford B. Hawley West Virginia University	J. G. Pitt University of Toronto
Musa Ayar Uw-baraboo/Sauk County	Vance A. Hughey Western Nevada College	Scott A. Redenius Brandeis University
Kathleen Burke SUNY Cortland	Tony Hunnicutt Ouachita Technical College	Sandra Robertson Thomas Nelson Community College
YC Chang University of Notre Dame	Stacey M. Jones Albers School of Business and Economics, Seattle University	Sunil Sapra California State University, Los Angeles
David Chen Rosemont College and Saint Joseph's University	Dukpa Kim University of Virginia	Kyle Vann Scott Snead State Community College
Margaret E. Cochran Northwestern State University of Louisiana	Rajaram Krishnan Earlham College	Rodney E. Stanley Tennessee State University
Thomas A. Dahlstrom Eastern University	Robert J. Lemke Lake Forest College	Jennifer Strehler Oakton Community College
Anne Drougas Dominican University	Philip J. Mizzi Arizona State University	Ronald Stunda Valdosta State University
Fesseha Gebremikael Strayer University/Calhoun Commu- nity College	Mehdi Mohaghegh Norwich University	Cindy van Es Cornell University
Malcolm C. Gold University of Wisconsin— Marshfield/Wood County	Mihail Motzev Walla Walla University	Jennifer VanGilder Ursinus College
Joel Goldstein Western Connecticut State University	Somnath Mukhopadhyay The University of Texas at El Paso	Jacqueline Wroughton Northern Kentucky University
Jim Grant Lewis & Clark College	Kenneth E. Murphy Chapman University	Dmitry Yarushkin Grand View University
	Ogbonnaya John Nwoha Grambling State University	David Zimmer Western Kentucky University

We continue to owe 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	West Virginia University	Roger Myerson
University of Tennessee at Chattanooga	Jim Hightower	Northwestern University
Lari Arjomand	California State University, Fullerton	Richard O'Connell
Clayton College and State University	Alan Humphrey	Miami University of Ohio
Robert Balough	University of Rhode Island	Alan Olinsky
Clarion University	Ann Hussein	Bryant College
Philip Boudreaux	Philadelphia College of Textiles and Science	Ceyhun Ozgur
University of Louisiana	C. Thomas Innis	Valparaiso University
Mike Bourke	University of Cincinnati	Tom Pray
Houston Baptist University	Ben Isselhardt	Rochester Institute of Technology
James Brannon	Rochester Institute of Technology	Harold Rahmlow
University of Wisconsin— Oshkosh	Jeffery Jarrett	St. Joseph's University
John Bryant	University of Rhode Island	H. V. Ramakrishna
University of Pittsburgh	Ronald Klimberg	Penn State University at Great Valley
Peter Bryant	St. Joseph's University	Tom Ryan
University of Colorado	David A. Kravitz	Case Western Reserve University
Terri L. Byczkowski	George Mason University	Bill Seaver
University of Cincinnati	David Krueger	University of Tennessee
Robert Carver	St. Cloud State University	Alan Smith
Stonehill College	John Leschke	Robert Morris College
Richard Claycombe	University of Virginia	Willbann Terpening
McDaniel College	Martin S. Levy	Gonzaga University
Robert Cochran	University of Cincinnati	Ted Tsukahara
University of Wyoming	John S. Loucks	St. Mary's College of California
Robert Collins	St. Edward's University	David Tufte
Marquette University	David Lucking-Reiley	University of New Orleans
David W. Cravens	Vanderbilt University	Victor Ukpolo
Texas Christian University	Bala Maniam	Austin Peay State University
Tom Dahlstrom	Sam Houston State University	Ebenge Usip
Eastern College	Don Marx	Youngstown State University
Gopal Dorai	University of Alaska, Anchorage	Cindy Van Es
William Patterson University	Tom McCullough	Cornell University
Nicholas Farnum	University of California— Berkeley	Jack Vaughn
California State University—Fullerton	Ronald W. Michener	University of Texas-El Paso
Donald Gren	University of Virginia	Andrew Welki
Salt Lake Community College	Glenn Milligan	John Carroll University
Paul Guy	Ohio State University	
California State University—Chico	Mitchell Muesham	
Clifford Hawley	Sam Houston State University	

Ari Wijetunga Morehead State University	Mustafa Yilmaz Northeastern University	Yan Yu University of Cincinnati
J. E. Willis Louisiana State University	Gary Yoshimoto St. Cloud State 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 product manager, Aaron Arnsparger; our learning designer, Brandon Foltz; our content manager, Conor Allen; our project manager at MPS Limited, Manoj Kumar; and others at Cengage 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
Michael J. Fry
Jeffrey W. Ohlmann

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
- Observational Study
- Experiment
- Time and Cost Issues
- Data Acquisition Errors

1.4 DESCRIPTIVE STATISTICS

1.5 STATISTICAL INFERENCE

1.6 ANALYTICS

1.7 BIG DATA AND DATA MINING

1.8 COMPUTERS AND STATISTICAL ANALYSIS

1.9 ETHICAL GUIDELINES FOR STATISTICAL PRACTICE

SUMMARY 21

GLOSSARY 21

SUPPLEMENTARY EXERCISES 22

APPENDIX 1.1 OPENING AND SAVING DATA FILES AND
CONVERTING TO STACKED FORM WITH JMP

STATISTICS IN PRACTICE

Bloomberg Businessweek*

NEW YORK, NEW YORK

Bloomberg Businessweek is one of the most widely read business magazines in the world. 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* 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. Examples of articles and reports include the impact of businesses moving important work to cloud computing, the crisis facing the U.S. Postal Service, and why the debt crisis is even worse than we 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 advertisers. One North American subscriber survey indicated that 64% of *Bloomberg Businessweek* subscribers are involved with computer purchases at work. Such statistics alert *Bloomberg*

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



Bloomberg Businessweek uses statistical facts and summaries in many of its articles. AP Images/Weng lei-Imaginechina

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 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:

- Unemployment dropped to an 18-year low of 3.8% in May 2018 from 3.9% in April and after holding at 4.1% for the prior six months (*Wall Street Journal*, June 1, 2018).
- Tesla ended 2017 with around \$5.4 billion of liquidity. Analysts forecast it will burn through \$2.8 billion of cash this year (*Bloomberg Businessweek*, April 19, 2018).
- The biggest banks in America reported a good set of earnings for the first three months of 2018. Bank of America and Morgan Stanley made quarterly net profits of \$6.9 billion and \$2.7 billion, respectively (*The Economist*, April 21, 2018).
- According to a study from the Pew Research Center, 15% of U.S. adults say they have used online dating sites or mobile apps (*Wall Street Journal*, May 2, 2018).

- According to the U.S. Centers for Disease Control and Prevention, in the United States alone, at least 2 million illnesses and 23,000 deaths can be attributed each year to antibiotic-resistant bacteria (*Wall Street Journal*, February 13, 2018).

The numerical facts in the preceding statements—3.8%, 3.9%, 4.1%, \$5.4 billion, \$2.8 billion, \$6.9 billion, \$2.7 billion, 15%, 2 million, 23,000—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 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 uses of data in developing descriptive statistics and in making statistical inferences are described in Sections 1.4 and 1.5. The last four sections of Chapter 1 provide an introduction to business analytics and the role statistics plays in it, an introduction to big data and data mining, the role of the computer in statistical analysis, and a discussion of ethical guidelines for statistical practice.

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 average dividend yield for the S&P 500 companies for 2017 was 1.88%. Over the same period, the average dividend yield for Microsoft was 1.72% (Yahoo Finance). In this case, the statistical information on dividend yield indicates a lower 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 The Nielsen Company and IRI 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 \bar{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 \bar{x} -bar value, is plotted on an \bar{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 \bar{x} -bar values fall between the chart's upper and lower control limits. Properly interpreted, an \bar{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.

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.

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.

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 disputes.

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 Gross Domestic Product (GDP) (\$):** The total market value (\$) of all goods and services produced by the nation divided by the number of people in the nation; this is commonly used to compare economic productivity of the nations.
- **Fitch Rating:** The nation's sovereign credit rating as appraised by the Fitch Group¹; 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 first observation (Armenia) contains the following measurements: Member, 3615, BB-, and Stable. The second observation (Australia) contains the following measurements: Member, 49755, 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 the 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

¹The 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 & Poor's and Moody's.

TABLE 1.1 Data Set for 60 Nations in the World Trade Organization

Nation	WTO Status	Per Capita GDP (\$)	Fitch Rating	Fitch Outlook
Armenia	Member	3,615	BB–	Stable
Australia	Member	49,755	AAA	Stable
Austria	Member	44,758	AAA	Stable
Azerbaijan	Observer	3,879	BBB–	Stable
Bahrain	Member	22,579	BBB	Stable
Belgium	Member	41,271	AA	Stable
Brazil	Member	8,650	BBB	Stable
Bulgaria	Member	7,469	BBB–	Stable
Canada	Member	42,349	AAA	Stable
Cape Verde	Member	2,998	B+	Stable
Chile	Member	13,793	A+	Stable
China	Member	8,123	A+	Stable
Colombia	Member	5,806	BBB–	Stable
Costa Rica	Member	11,825	BB+	Stable
Croatia	Member	12,149	BBB–	Negative
Cyprus	Member	23,541	B	Negative
Czech Republic	Member	18,484	A+	Stable
Denmark	Member	53,579	AAA	Stable
Ecuador	Member	6,019	B–	Positive
Egypt	Member	3,478	B	Negative
El Salvador	Member	4,224	BB	Negative
Estonia	Member	17,737	A+	Stable
France	Member	36,857	AAA	Negative
Georgia	Member	3,866	BB–	Stable
Germany	Member	42,161	AAA	Stable
Hungary	Member	12,820	BB+	Stable
Iceland	Member	60,530	BBB	Stable
Ireland	Member	64,175	BBB+	Stable
Israel	Member	37,181	A	Stable
Italy	Member	30,669	A–	Negative
Japan	Member	38,972	A+	Negative
Kazakhstan	Observer	7,715	BBB+	Stable
Kenya	Member	1,455	B+	Stable
Latvia	Member	14,071	BBB	Positive
Lebanon	Observer	8,257	B	Stable
Lithuania	Member	14,913	BBB	Stable
Malaysia	Member	9,508	A–	Stable
Mexico	Member	8,209	BBB	Stable
Peru	Member	6,049	BBB	Stable
Philippines	Member	2,951	BB+	Stable
Poland	Member	12,414	A–	Positive
Portugal	Member	19,872	BB+	Negative
South Korea	Member	27,539	AA–	Stable
Romania	Member	9,523	BBB–	Stable
Russia	Member	8,748	BBB	Stable
Rwanda	Member	703	B	Stable
Serbia	Observer	5,426	BB–	Negative
Singapore	Member	52,962	AAA	Stable
Slovakia	Member	16,530	A+	Stable



Slovenia	Member	21,650	A–	Negative
South Africa	Member	5,275	BBB	Stable
Spain	Member	26,617	A–	Stable
Sweden	Member	51,845	AAA	Stable
Switzerland	Member	79,888	AAA	Stable
Thailand	Member	5,911	BBB	Stable
Turkey	Member	10,863	BBB–	Stable
United Kingdom	Member	40,412	AAA	Negative
Uruguay	Member	15,221	BB+	Positive
United States	Member	57,638	AAA	Stable
Zambia	Member	1,270	B+	Negative

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 numerical. 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 of 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

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

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

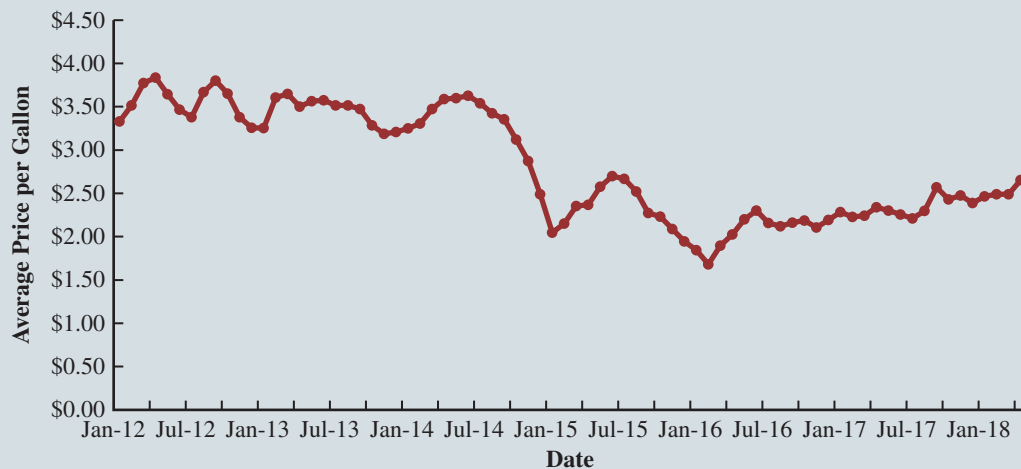
For purposes of statistical analysis, distinguishing between cross-sectional data and time series data is important. **Cross-sectional data** are data collected at the same or approximately the same point in time. The data in Table 1.1 are cross-sectional because they describe the five variables for 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 2012 and 2018. From January 2012 until June 2014, prices fluctuated between \$3.19 and \$3.84 per gallon before a long stretch of decreasing prices from July 2014 to January 2016. The lowest average price per gallon occurred in January 2016 (\$1.68). Since then, the average price appears to be on a gradual increasing trend.

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 2008 to 2018. Poor economic conditions caused a serious drop in the index during 2008 with the low point occurring in February 2009 (7062). After that, the index has been on a remarkable nine-year increase, reaching its peak (26,149) in January 2018.

The graph in Panel (B) shows the net income of McDonald's Inc. from 2008 to 2017. 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, decreased slightly in 2012, and peaked in 2013. After three years of relatively lower net income, their net income increased to \$5.19 billion in 2017.

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

FIGURE 1.1 U.S. Average Price per Gallon for Conventional Regular Gasoline

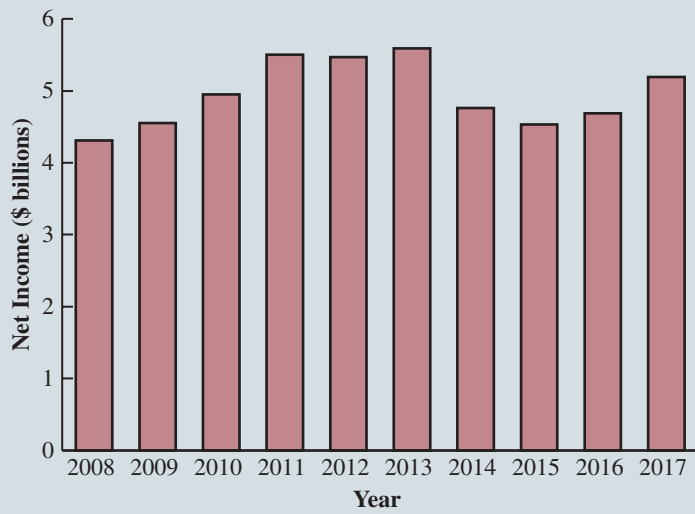


Source: Energy Information Administration, U.S. Department of Energy.

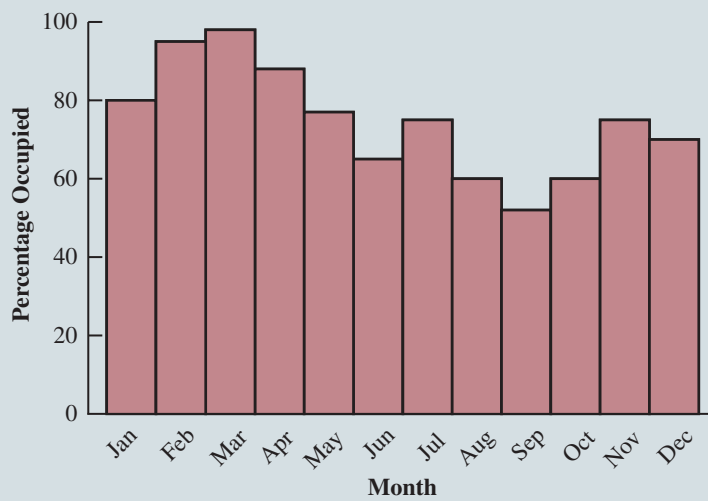
FIGURE 1.2 A Variety of Graphs of Time Series Data



(A) Dow Jones Industrial Average Index



(B) Net Income for McDonald's Inc.



(C) Occupancy Rate of South Florida Hotels