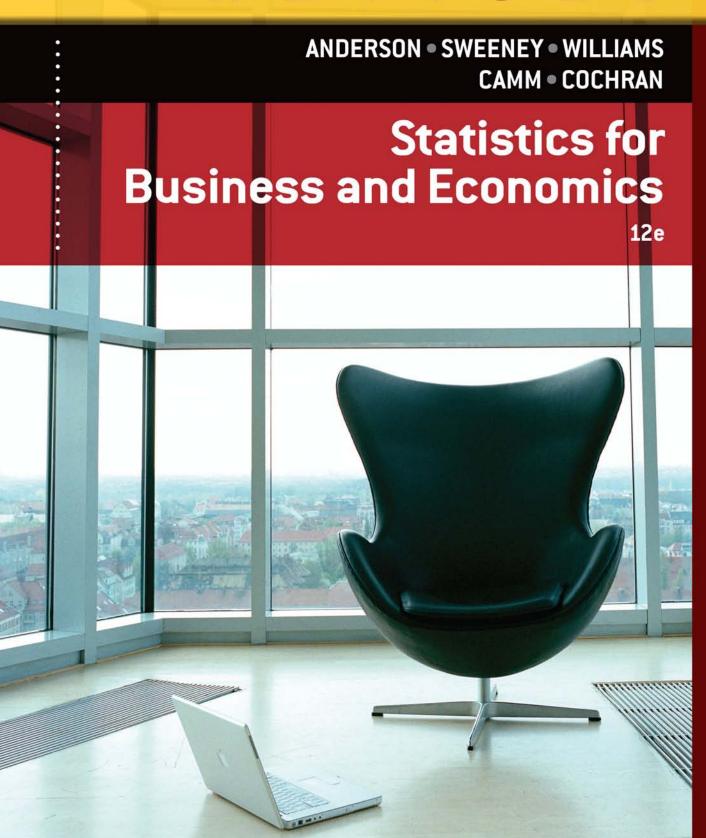
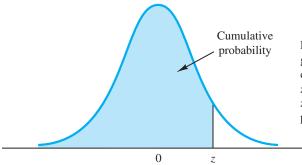
REVISED



CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION

	Cumulati					gi cu z	ntries in to the area area to the value. For a85, the cobability	ea under the left of the example e cumular	ne , for	
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	.0025	.0023	.0024	.0023	.0023	.0022	.0021	.0021	.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 .0 .5000 .5040 .508 .1 .5398 .5438 .547 .2 .5793 .5832 .587 .3 .6179 .6217 .625 .4 .6554 .6591 .662 .5 .6915 .6950 .698 .6 .7257 .7291 .732 .7 .7580 .7611 .764					
.1 .5398 .5438 .547 .2 .5793 .5832 .587 .3 .6179 .6217 .625 .4 .6554 .6591 .662 .5 .6915 .6950 .698 .6 .7257 .7291 .732 .7 .7580 .7611 .764	2 .03 .	04 .05	.06	.07	.08 .09
.2 .5793 .5832 .587 .3 .6179 .6217 .625 .4 .6554 .6591 .662 .5 .6915 .6950 .698 .6 .7257 .7291 .732 .7 .7580 .7611 .764	80 .5120 .5	160 .5199	.5239 .	5279 .:	5319 .535
.3 .6179 .6217 .625 .4 .6554 .6591 .662 .5 .6915 .6950 .698 .6 .7257 .7291 .732 .7 .7580 .7611 .764	78 .5517 .5	557 .5596	.5636 .	5675 .:	5714 .575
.4 .6554 .6591 .662 .5 .6915 .6950 .698 .6 .7257 .7291 .732 .7 .7580 .7611 .764	71 .5910 .5	948 .5987	.6026 .	6064 .	6103 .614
.5 .6915 .6950 .698 .6 .7257 .7291 .732 .7 .7580 .7611 .764	55 .6293 .6	.6368	.6406 .	6443	6480 .651
.6 .7257 .7291 .732 .7 .7580 .7611 .764	28 .6664 .6	700 .6736	.6772 .	6808 .0	.6844 .687
.7 .7580 .7611 .764	85 .7019 .7	.7088	.7123 .	7157 .	7190 .722
	24 .7357 .7	389 .7422	.7454 .	7486 .	7517 .754
0 7001 7010 703	42 .7673 .7	704 .7734	.7764 .	7794 .	7823 .785
.8 .7881 .7910 .793	39 .7967 .7	.8023		8078 .	8106 .813
.9 .8159 .8186 .821	12 .8238 .8	.8289	.8315 .	8340 .	8365 .838
1.0 .8413 .8438 .846	.8485 .8	508 .8531	.8554 .	8577 .	8599 .862
1.1 .8643 .8665 .868	.8708 .8	729 .8749	.8770 .	8790 .	8810 .883
1.2 .8849 .8869 .888	.8907 .8	925 .8944	.8962 .	8980 .	8997 .901
1.3 .9032 .9049 .906	.9082 .9	.9115	.9131 .	9147 .	9162 .917
1.4 .9192 .9207 .922	22 .9236 .9	251 .9265	.9279 .	9292 .	9306 .931
1.5 .9332 .9345 .935	.9370 .9	.9394	.9406 .	9418 .	9429 .944
1.6 .9452 .9463 .947	74 .9484 .9	495 .9505	.9515 .	9525 .	9535 .954
1.7 .9554 .9564 .957	73 .9582 .9.	591 .9599	.9608 .	9616 .	9625 .963
1.8 .9641 .9649 .965	56 .9664 .9	.9678	.9686 .	9693 .	9699 .970
1.9 .9713 .9719 .972	26 .9732 .9	738 .9744	.9750 .	9756 .	9761 .976
2.0 .9772 .9778 .978		793 .9798			9812 .981
2.1 .9821 .9826 .983	.9834 .9	.9842	.9846 .	9850 .	9854 .985
2.2 .9861 .9864 .986	58 .9871 .9	.9878	.9881 .	9884 .	9887 .989
2.3 .9893 .9896 .989	.9901 .9	904 .9906	.9909 .	9911 .	9913 .991
2.4 .9918 .9920 .992	22 .9925 .9	927 .9929	.9931 .	9932 .	9934 .993
2.5 .9938 .9940 .994	.9943 .9	945 .9946	.9948 .	9949 .	9951 .995
2.6 .9953 .9955 .995	56 .9957 .9	959 .9960	.9961 .	9962 .	9963 .996
2.7 .9965 .9966 .996	.9968 .9	969 .9970	.9971 .	9972 .	9973 .997
2.8 .9974 .9975 .997	76 .9977 .9	.9978	.9979 .	9979 .	9980 .998
2.9 .9981 .9982 .998	.9983 .9	984 .9984	.9985 .	9985 .9	9986 .998
3.0 .9987 .9987 .998	.9988 .9	.9989	.9989 .	9989 .	9990 .999

STATISTICS FOR BUSINESS AND ECONOMICS 12e REVISED

STATISTICS FOR BUSINESS AND ECONOMICS 12e REVISED

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



This is an electronic version of the print textbook. Due to electronic rights restrictions, some third party content may be suppressed. Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. 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.cengage.com/highered to search by ISBN#, author, title, or keyword for materials in your areas of interest.



Statistics for Business and Economics, Twelfth Edition, Revised

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

Product Director: Joe Sabatino

Product Manager: Aaron Arnsparger

Sr. Brand Manager: Kristen Hurd

Content Developer: Maggie Kubale

Content Project Manager: Jana Lewis

Media Developer: 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

Intellectual Property

Analyst: Jennifer Nonenmacher Project Manager: Anne Sheroff

© 2014, 2012 Cengage Learning

WCN: 02-200-208

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

Further permissions questions can be emailed to permissionrequest@cengage.com

ExamView[®] is a registered trademark of elnstruction 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.

Unless otherwise noted, screen shots are Microsoft Excel[®].

Library of Congress Control Number: 2014931833

ISBN-13: 978-1-285-84632-3

ISBN-10: 1-285-84632-X

Cengage Learning

200 First Stamford Place, 4th Floor Stamford, CT 06902 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

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

For your course and learning solutions, visit **www.cengage.com**

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**

Printed in Canada 1 2 3 4 5 6 7 18 17 16 15 14

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

Brief Contents

Preface xxv	
About the Author	ors xxxi
Chapter 1	Data and Statistics 1
Chapter 2	Descriptive Statistics: Tabular and
_	Graphical Displays 33
Chapter 3	Descriptive Statistics: Numerical Measures 99
Chapter 4	Introduction to Probability 169
Chapter 5	Discrete Probability Distributions 215
Chapter 6	Continuous Probability Distributions 265
Chapter 7	Sampling and Sampling Distributions 298
Chapter 8	Interval Estimation 342
Chapter 9	Hypothesis Tests 382
Chapter 10	Inference About Means and Proportions with Two Populations 441
Chapter 11	Inferences About Population Variances 482
Chapter 12	Comparing Multiple Proportions, Test of Independence and Goodness of Fit 507
Chapter 13	Experimental Design and Analysis of Variance 545
Chapter 14	Simple Linear Regression 598
Chapter 15	Multiple Regression 682
Chapter 16	Regression Analysis: Model Building 751
Chapter 17	Time Series Analysis and Forecasting 800
Chapter 18	Nonparametric Methods 870
Chapter 19	Statistical Methods for Quality Control 916
Chapter 20	Index Numbers 951
Chapter 21	Decision Analysis (On Website)
Chapter 22	Sample Survey (On Website)
Appendix A	References and Bibliography 971
Appendix B	Tables 974
Appendix C	Summation Notation 1001
Appendix D	Self-Test Solutions and Answers to Even-Numbered Exercises 1003
Appendix E	Microsoft Excel 2013 and Tools for Statistical Analysis 1064
Appendix F	Computing <i>p</i> -Values Using Minitab and Excel 1070
Index 1080	

Contents

	face xxv
ADO	ut the Authors xxxi
Cha	apter 1 Data and Statistics 1
Stati	stics 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
Sum	mary 21
	sary 21
	blementary Exercises 22
	endix: An Introduction to StatTools 29

Chapter 2 Descriptive Statistics: Tabular and Graphical Displays 33

Statistics in Practice: Colgate-Palmolive Company 342.1 Summarizing Data for a Categorical Variable 35Frequency Distribution 35

Relative Frequency and Percent Frequency Distributions 36 Bar Charts and Pie Charts 36

2.2 Summarizing Data for a Quantitative Variable 42

Frequency Distribution 42

Relative Frequency and Percent Frequency Distributions 43

Dot Plot 44

Histogram 44

Cumulative Distributions 46

Stem-and-Leaf Display 47

2.3 Summarizing Data for Two Variables Using Tables 55

Crosstabulation 55

Simpson's Paradox 58

2.4 Summarizing Data for Two Variables Using Graphical Displays 64

Scatter Diagram and Trendline 64

Side-by-Side and Stacked Bar Charts 65

2.5 Data Visualization: Best Practices in Creating Effective Graphical Displays 70

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

Summary 77

Glossary 78

Key Formulas 79

Supplementary Exercises 79

Case Problem 1: Pelican Stores 84

Case Problem 2: Motion Picture Industry 85

Appendix 2.1 Using Minitab for Tabular and Graphical Presentations 86

Appendix 2.2 Using Excel for Tabular and Graphical Presentations 88

Appendix 2.3 Using StatTools for Tabular and Graphical Presentations 98

Chapter 3 Descriptive Statistics: Numerical Measures 99

Statistics in Practice: Small Fry Design 100

3.1 Measures of Location 101

Mean 101

Weighted Mean 103

Median 104

Geometric Mean 106

Mode 107

Percentiles 108

Quartiles 109

Contents xi

3.2 Measures of Variability 116

Range 116

Interquartile Range 117

Variance 117

Standard Deviation 118

Coefficient of Variation 119

3.3 Measures of Distribution Shape, Relative Location, and Detecting Outliers 123

Distribution Shape 123

z-Scores 123

Chebyshev's Theorem 125

Empirical Rule 126

Detecting Outliers 127

3.4 Five-Number Summaries and Box Plots 130

Five-Number Summary 131

Box Plot 131

3.5 Measures of Association Between Two Variables 136

Covariance 136

Interpretation of the Covariance 138

Correlation Coefficient 140

Interpretation of the Correlation Coefficient 141

3.6 Data Dashboards: Adding Numerical Measures to Improve Effectiveness 145

Summary 149

Glossary 149

Key Formulas 150

Supplementary Exercises 152

Case Problem 1: Pelican Stores 157

Case Problem 2: Motion Picture Industry 158

Case Problem 3: Business Schools of Asia-Pacific 159

Case Problem 4: Heavenly Chocolates Website Transactions 161

Case Problem 5: African Elephant Populations 162

Appendix 3.1 Descriptive Statistics Using Minitab 163

Appendix 3.2 Descriptive Statistics Using Excel 165

Appendix 3.3 Descriptive Statistics Using StatTools 167

Chapter 4 Introduction to Probability 169

Statistics in Practice: Probability to the Rescue 170

4.1 Experiments, Counting Rules, and Assigning Probabilities 171

Counting Rules, Combinations, and Permutations 172

Assigning Probabilities 176

Probabilities for the KP&L Project 178

4.2 Events and Their Probabilities 181

4.3 Some Basic Relationships of Probability	19:
---	-----

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

Contents

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 Grear 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

7.7 Properties of Point Estimators 328

Unbiased 328 Efficiency 329 Consistency 330

7.8 Other Sampling Methods 331

Stratified Random Sampling 331 Cluster Sampling 331 Systematic Sampling 332 Convenience Sampling 332 Judgment Sampling 333

Summary 333

Glossary 334

Key Formulas 335

Supplementary Exercises 335

Appendix 7.1 The Expected Value and Standard Deviation of \bar{x} 337

Appendix 7.2 Random Sampling with Minitab 339

Appendix 7.3 Random Sampling with Excel 340

Appendix 7.4 Random Sampling with StatTools 341

Chapter 8 Interval Estimation 342

Statistics in Practice: Food Lion 343 8.1 Population Mean: *σ* Known 344

Margin of Error and the Interval Estimate 344 Practical Advice 348

8.2 Population Mean: σ Unknown 350

Margin of Error and the Interval Estimate 351
Practical Advice 354
Using a Small Sample 354
Summary of Interval Estimation Procedures 356

8.3 Determining the Sample Size 359

8.4 Population Proportion 362

Determining the Sample Size 364

Summary 367

Glossary 368

Key Formulas 369

Supplementary Exercises 369

Case Problem 1: Young Professional Magazine 372 Case Problem 2: Gulf Real Estate Properties 373 Case Problem 3: Metropolitan Research, Inc. 375 Appendix 8.1 Interval Estimation with Minitab 375 Contents

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: σ 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: σ 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: σ_1 and σ_2 Known 443

Interval Estimation of $\mu_1 - \mu_2$ 443

xvi Contents

Hypothesis Tests About $\mu_1 - \mu_2$ 445 Practical Advice 447

10.2 Inferences About the Difference Between Two Population Means: σ_1 and σ_2 Unknown 450

Interval Estimation of $\mu_1 - \mu_2$ 450 Hypothesis Tests About $\mu_1 - \mu_2$ 452 Practical Advice 454

- 10.3 Inferences About the Difference Between Two Population Means: Matched Samples 458
- 10.4 Inferences About the Difference Between Two Population Proportions 464

Interval Estimation of $p_1 - p_2$ 464 Hypothesis Tests About $p_1 - p_2$ 466

Summary 470

Glossary 471

Key Formulas 471

Supplementary Exercises 472 Case Problem: Par, Inc. 475

Appendix 10.1 Inferences About Two Populations Using Minitab 476 Appendix 10.2 Inferences About Two Populations Using Excel 478 Appendix 10.3 Inferences About Two Populations Using StatTools 479

Chapter 11 Inferences About Population Variances 482

Statistics in Practice: U.S. Government Accountability Office 483

11.1 Inferences About a Population Variance 484

Interval Estimation 484 Hypothesis Testing 488

11.2 Inferences About Two Population Variances 494

Summary 501

Key Formulas 501

Supplementary Exercises 501

Case Problem: Air Force Training Program 503

Appendix 11.1 Population Variances with Minitab 504

Appendix 11.2 Population Variances with Excel 505

Appendix 11.3 Single Population Standard Deviation with StatTools 505

Chapter 12 Comparing Multiple Proportions, Test of Independence and Goodness of Fit 507

Statistics in Practice: United Way 508

12.1 Testing the Equality of Population Proportions for Three or More Populations 509

A Multiple Comparison Procedure 514

Contents xvii

12.2 Test of Independence 519

12.3 Goodness of Fit Test 527

Multinomial Probability Distribution 527 Normal Probability Distribution 530

Summary 536

Glossary 536

Key Formulas 537

Supplementary Exercises 537

Case Problem: A Bipartisan Agenda for Change 540

Appendix 12.1 Chi-Square Tests Using Minitab 541

Appendix 12.2 Chi-Square Tests Using Excel 542

Appendix 12.3 Chi-Square Tests Using StatTools 544

Chapter 13 Experimental Design and Analysis of Variance 545

Statistics in Practice: Burke Marketing Services, Inc. 546

13.1 An Introduction to Experimental Design and Analysis of Variance 547

Data Collection 548

Assumptions for Analysis of Variance 549

Analysis of Variance: A Conceptual Overview 549

13.2 Analysis of Variance and the Completely Randomized Design 552

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

13.3 Multiple Comparison Procedures 563

Fisher's LSD 563

Type I Error Rates 566

13.4 Randomized Block Design 569

Air Traffic Controller Stress Test 570

ANOVA Procedure 571

Computations and Conclusions 572

13.5 Factorial Experiment 576

ANOVA Procedure 578

Computations and Conclusions 578

Summary 583

Glossary 584

Key Formulas 584

Supplementary Exercises 586

xviii Contents

Case Pr	oblem 1:	Wentworth	Medical	Center	591

Case Problem 2: Compensation for Sales Professionals 592

Appendix 13.1 Analysis of Variance with Minitab 592

Appendix 13.2 Analysis of Variance with Excel 594

Appendix 13.3 Analysis of a Completely Randomized Design Using StatTools 597

Chapter 14 Simple Linear Regression 598

Statistics in Practice: Alliance Data Systems 599

14.1 Simple Linear Regression Model 600

Regression Model and Regression Equation 600 Estimated Regression Equation 601

- 14.2 Least Squares Method 603
- 14.3 Coefficient of Determination 614

Correlation Coefficient 618

- 14.4 Model Assumptions 622
- 14.5 Testing for Significance 623

Estimate of σ^2 623

t Test 624

Confidence Interval for β_1 626

F Test 627

Some Cautions About the Interpretation of Significance Tests 629

14.6 Using the Estimated Regression Equation for Estimation and Prediction 632

or Estimation and Frediction

Interval Estimation 633

Confidence Interval for the Mean Value of y 634

Prediction Interval for an Individual Value of y 635

- 14.7 Computer Solution 640
- 14.8 Residual Analysis: Validating Model Assumptions 644

Residual Plot Against x 645

Residual Plot Against ŷ 646

Standardized Residuals 648

Normal Probability Plot 650

14.9 Residual Analysis: Outliers and Influential Observations 653

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

Contents xix

Case Problem 4: Finding the Best Car Value 674

Appendix 14.1 Calculus-Based Derivation of Least Squares Formulas 675

Appendix 14.2 A Test for Significance Using Correlation 677

Appendix 14.3 Regression Analysis with Minitab 678

Appendix 14.4 Regression Analysis with Excel 678

Appendix 14.5 Regression Analysis Using StatTools 681

Chapter 15 Multiple Regression 682

Statistics in Practice: dunnhumby 683

15.1 Multiple Regression Model 684

Regression Model and Regression Equation 684 Estimated Multiple Regression Equation 684

15.2 Least Squares Method 685

An Example: Butler Trucking Company 686 Note on Interpretation of Coefficients 688

15.3 Multiple Coefficient of Determination 694

15.4 Model Assumptions 698

15.5 Testing for Significance 699

F Test 699 *t* Test 702

Multicollinearity 703

15.6 Using the Estimated Regression Equation for Estimation and Prediction 706

15.7 Categorical Independent Variables 709

An Example: Johnson Filtration, Inc. 709 Interpreting the Parameters 711

More Complex Categorical Variables 713

15.8 Residual Analysis 717

Detecting Outliers 719

Studentized Deleted Residuals and Outliers 719

Influential Observations 720

Using Cook's Distance Measure to Identify

Influential Observations 720

15.9 Logistic Regression 724

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

Summary 736

Glossary 736

Key Formulas 737

Supp	lementary Exercises 739
Case	Problem 1: Consumer Research, Inc. 745
Case	Problem 2: Predicting Winnings for NASCAR Drivers 746
Case	Problem 3: Finding the Best Car Value 747
	ndix 15.1 Multiple Regression with Minitab 748
	ndix 15.2 Multiple Regression with Excel 748
	ndix 15.3 Logistic Regression with Minitab 750
	ndix 15.4 Multiple Regression Analysis Using StatTools 750
Cha	pter 16 Regression Analysis: Model Building 751
Statis	stics in Practice: Monsanto Company 752
16.1	General Linear Model 753
	Modeling Curvilinear Relationships 753
	Interaction 756
	Transformations Involving the Dependent Variable 760
	Nonlinear Models That Are Intrinsically Linear 763
16.2	Determining When to Add or Delete Variables 767
	General Case 769
16.2	Use of p-Values 770
16.3	
16.4	Variable Selection Procedures 777
	Stepwise Regression 777 Forward Selection 778
	Backward Elimination 779
	Best-Subsets Regression 779
	Making the Final Choice 780
16.5	-
16.6	Autocorrelation and the Durbin-Watson Test 788
Sumr	mary 792
	ary 792
	Formulas 792
•	lementary Exercises 793
	Problem 1: Analysis of PGA Tour Statistics 796
	Problem 2: Rating Wines from the Piedmont Region of Italy 797
	ndix 16.1 Variable Selection Procedures with Minitab 798
	ndix 16.2 Variable Selection Procedures Using StatTools 799
• • •	
Cha	pter 17 Time Series Analysis and Forecasting 800

Chapter 17 Time Series Analysis and Forecasting 800

Statistics in Practice: Nevada Occupational Health Clinic 801

17.1 Time Series Patterns 802

Horizontal Pattern 802

Contents xxi

Trend Pattern 804

Seasonal Pattern 804

Trend and Seasonal Pattern 805

Cyclical Pattern 805

Selecting a Forecasting Method 807

17.2 Forecast Accuracy 808

17.3 Moving Averages and Exponential Smoothing 813

Moving Averages 813

Weighted Moving Averages 816

Exponential Smoothing 816

17.4 Trend Projection 823

Linear Trend Regression 823

Holt's Linear Exponential Smoothing 828

Nonlinear Trend Regression 830

17.5 Seasonality and Trend 836

Seasonality Without Trend 836

Seasonality and Trend 838

Models Based on Monthly Data 841

17.6 Time Series Decomposition 845

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

Summary 855

Glossary 856

Key Formulas 857

Supplementary Exercises 857

Case Problem 1: Forecasting Food and Beverage Sales 861

Case Problem 2: Forecasting Lost Sales 862

Appendix 17.1 Forecasting with Minitab 864

Appendix 17.2 Forecasting with Excel 866

Appendix 17.3 Forecasting Using StatTools 867

Chapter 18 Nonparametric Methods 870

Statistics in Practice: West Shell Realtors 871

18.1 Sign Test 872

Hypothesis Test About a Population Median 872 Hypothesis Test with Matched Samples 877

- 18.2 Wilcoxon Signed-Rank Test 880
- 18.3 Mann-Whitney-Wilcoxon Test 885
- 18.4 Kruskal-Wallis Test 895

xxii Contents

18 5	Rank	Corre	lation	900
10.3	Канк	Corre	1211011	71111

Summary 905

Glossary 905

Key Formulas 906

Supplementary Exercises 907

Appendix 18.1 Nonparametric Methods with Minitab 910

Appendix 18.2 Nonparametric Methods with Excel 912

Appendix 18.3 Nonparametric Methods with StatTools 914

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

Contents xxiii

20.4 Some Important Price Indexes 959

Consumer Price Index 959 Producer Price Index 959 Dow Jones Averages 960

20.5 Deflating a Series by Price Indexes 961

20.6 Price Indexes: Other Considerations 964

Selection of Items 964 Selection of a Base Period 965 Quality Changes 965

20.7 Quantity Indexes 965

Summary 967 Glossary 967 Key Formulas 968 Supplementary Exercises 968

Chapter 21 Decision Analysis (On Website)

Statistics in Practice: Ohio Edison Company 21-2

21.1 Problem Formulation 21-3

Payoff Tables 21-4 Decision Trees 21-4

21.2 Decision Making with Probabilities 21-5

Expected Value Approach 21-5
Expected Value of Perfect Information 21-7

21.3 Decision Analysis with Sample Information 21-13

Decision Tree 21-14
Decision Strategy 21-15
Expected Value of Sample Information 21-18

21.4 Computing Branch Probabilities Using Bayes' Theorem 21-24

Summary 21-28 Glossary 21-29

Key Formulas 21-30

Supplementary Exercises 21-30

Case Problem: Lawsuit Defense Strategy 21-33 Appendix: An Introduction to PrecisionTree 21-34

Appendix: Self-Test Solutions and Answers to Even-Numbered

Exercises 21-39

Chapter 22 Sample Survey (On Website)

Statistics in Practice: Duke Energy 22-2

- 22.1 Terminology Used in Sample Surveys 22-2
- 22.2 Types of Surveys and Sampling Methods 22-3

xxiv Contents

22.3	Survey E	rrors 22-5
	Nonsam	apling Error 22-5
	Samplin	ng Error 22-5
22.4	Simple Ra	andom Sampling 22-6
	Populati	ion Mean 22-6
	Populati	ion Total 22-7
	Populati	ion Proportion 22-8
	Determi	ining the Sample Size 22-9
22.5	Stratified	Simple Random Sampling 22-12
	Populati	ion Mean 22-12
	Populati	ion Total 22-14
	Populati	ion Proportion 22-15
	Determi	ining the Sample Size 22-16
22.6	Cluster Sa	ampling 22-21
	Populati	ion Mean 22-23
	Populati	ion Total 22-24
	Populati	ion Proportion 22-25
	Determi	ining the Sample Size 22-26
22.7	Systemati	c Sampling 22-29
Sumr	mary 22-29	
Gloss	sary 22-30	
Key I	Formulas 2	22-30
Supp	lementary 1	Exercises 22-34
	•	Test Solutions and Answers to Even-Numbered
		rcises 22-37
Ann	endix A	References and Bibliography 971
11 P P	CHUIA 1	References and Bibliography 371
A	II D	
App	endix B	Tables 974
App	endix C	Summation Notation 1001
Ann	andiv D	Self-Test Solutions and Answers to Even-Numbered
App	eliuix D	Exercises 1003
		Exercises 1005
App	endix E	Microsoft Excel 2013 and Tools for Statistical
		Analysis 1064
Ann	endix F	Computing <i>p</i> -Values Using Minitab and Excel 1076
		Tong the same and
Inde	x 1080	

This text is the revised 12th edition of *STATISTICS FOR BUSINESS AND ECONOMICS*. With the 12th edition we welcomed 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 2013 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 Revised 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.

In addition to the major revisions described in the remainder of this section, this *revised* edition of the twelfth edition has been updated to incorporate Microsoft[®] Office Excel[®] 2013. Changes in functions and procedures and powerful new options such as Recommended Charts and Recommend PivotTables are discussed.

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.

Preface xxvii

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 Margaret E. Cochran Joel Goldstein
Aboueissa, University of Northwestern State Western Connecticut State
Southern Maine University of Louisiana University

Kathleen Arano Thomas A. Dahlstrom Jim Grant
Fort Hays State University Eastern University Lewis & Clark College

Musa Ayar Anne Drougas Reidar Hagtvedt
Uw-baraboo/Sauk County Dominican University University of Alberta
Fesseha Gebremikael School of Business

Kathleen Burke
SUNY Cortland
Strayer University/
Calhoun Community
Callege West Virginia University

YC Chang College

University of Notre
Dame

College

Malcolm C. Gold
University of

Western Nevada College

David Chen Wisconsin— Tony Hunnicutt
Rosemont College and Marshfield/ Ouachita Technical
Saint Joseph's University Wood County College

Copyright 2014 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it

xxviii Preface

Stacey M. Jones Kenneth E. Murphy Rodney E. Stanley Albers School of Business Tennessee State University Chapman University and Economics. Seattle Ogbonnaya John Nwoha Jennifer Strehler University Grambling State University Oakton Community Dukpa Kim College Claudiney Pereira University of Virginia Ronald Stunda **Tulane University** Rajaram Krishnan Valdosta State University J. G. Pitt Earlham College Cindy van Es University of Toronto Robert J. Lemke Cornell University Scott A. Redenius Lake Forest College Jennifer VanGilder Brandeis University Philip J. Mizzi Ursinus College Sandra Robertson Arizona State University Jacqueline Wroughton Thomas Nelson Mehdi Mohaghegh Northern Kentucky Community College Norwich University University Sunil Sapra Mihail Motzev Dmitry Yarushkin California State University, Walla Walla University Grand View University Los Angeles David Zimmer Somnath Mukhopadhyay Kyle Vann Scott The University of Texas Western Kentucky Snead State Community at El Paso University College

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	Richard Claycombe McDaniel College	Clifford Hawley West Virginia University
at Chattanooga Lari Arjomand Clayton College and State	Robert Cochran University of Wyoming Robert Collins	Jim Hightower California State University, Fullerton
University Robert Balough Clarion University	Marquette University David W. Cravens	Alan Humphrey University of Rhode Island
Philip Boudreaux University of Louisiana	Texas Christian University	Ann Hussein Philadelphia College of Textiles and Science
Mike Bourke Houston Baptist University	Tom Dahlstrom Eastern College Gopal Dorai	C. Thomas Innis University of Cincinnati
James Brannon University of Wisconsin— Oshkosh	William Patterson University Nicholas Farnum	Ben Isselhardt Rochester Institute of Technology
John Bryant University of Pittsburgh	California State University—Fullerton	Jeffery Jarrett University of Rhode
Peter Bryant University of Colorado	Donald Gren Salt Lake Community	Island Ronald Klimberg
Terri L. Byczkowski University of Cincinnati Robert Carver	College Paul Guy California State	St. Joseph's University David A. Kravitz George Mason
1000011 Cui (Ci	Cullivilla State	George Mason

University—Chico

University

Stonehill College

Preface xxix

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

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 content developer, Maggie Kubale; our content project manager, Jana Lewis; our Project Manager at MPS Limited, Lynn Lustberg; our media editor, Chris Valentine; and others at Cengage Learning 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-forprofit 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



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

- 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.



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

1.2 Data 5

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.



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 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 set of measurements for the first observation (Armenia) is

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 and Poor's and Moody's investor service.

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



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

Nation Status CDP (\$S) Deficit (\$1000s) Rating 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 Cape Verde Member 40,000 874,459 B+ Stable China Member 16,100 -14,558,218 A + Stable China Member 11,500 5,807,509 BBB — Stable Costa Rica Member										
Nation Status GDP (\$) Deficit (\$1000s) Rating Outlook Armenia Member 5,400 2,673,359 BB — Stable Austraia 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 Bulgaria Member 11,600 -29,796,166 BBB Stable Bulgaria Member 40,300 -16,11,380 AAA Stable Cape Verde Member 40,000 874,459 B+ Stable Chile Member 16,100 -14,558,218 A+ Stable Chile Member 16,100 -15,6709,311 A+ Stable Chile Member 11,500 5,807,509 BB+ Stable Costa Rica <	WTO		Per Canita	Trade	Fitch	Fitch				
Armenia Member 40,800 -33,304,157 AAA Stable Austria Member 41,700 12,796,558 AAA Stable Austria Member 41,700 12,796,558 AAA Stable Azerbaijam Observer 5,400 -16,747,320 BBB - Positive Belgium 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 40,300 -16,11,380 AAA Stable Cape Verde Member 4,000 874,459 B+ Stable Cape Verde Member 4,000 874,459 B+ Stable Chiia 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 Cyprus Member 25,900 -10,749,467 A+ Positive Demmark Member 40,200 -15,057,343 AAA Stable Ecuador Member 40,200 -15,057,343 AAA Stable Ecuador Member 6,500 28,486,933 BB Negative Egypt Member 6,500 28,486,933 BB Negative Efstonia Member 35,000 118,841,542 AAA Stable Estonia Member 37,900 -213,367,685 AAA Stable France Member 37,900 -213,367,685 AAA Stable Ireland Member 39,500 -50,93,323 BBB Negative Isalu Member 31,000 31,675,424 AA Negative Isalu Member 34,300 31,675,424 AA Negative Isalu Member 15,400 2,448,053 BBB Negative Isalu Member 15,400 2,448,053 BBB Negative Isalu Member 18,700 3,359,641 BBB Positive Isalu Member 15,600 1,288,112 BBB Stable Isalu Member 15,600 -3,420,064 A Stable Isalu Member 15,600 -3,420,064 A Stable Isalu Member 15,000 -3,420,064 A Stable Isalu Member 15,000 -3,420,064 A Stable Isalu Member 15,000 -3,439,964 BBB Stable Isalu Member 15,000 -3,439,964 BBB Stable Isalu Member 15,000 -3,420,064										
Australia Member 40,800 -33,304,157 AAA Stable Austria Member 41,700 12,796,558 AAA Stable Azerbaijam Observer 5,400 -16,747,320 BBBB Positive Bahrain Member 27,300 3,102,665 BBB Stable Belgium Member 37,600 -14,930,333 AA Negative Brazil Member 11,600 -29,796,166 BBB Stable Bulgaria Member 40,300 -1611,380 AAA Stable Canada Member 40,000 874,459 B+ Stable Cape Verde Member 16,100 -14,558,218 A+ Stable China Member 10,100 -1,561,199 BBB- Stable China Member 11,500 5,807,509 BB+ Stable Costa Rica Member 11,500 5,807,509 BB+ Stable Cyprus Member			* * * *	· · · · · · · · · · · · · · · · · · ·						
Austria Member 41,700 12,796,558 AAA Stable Azerbaijan Observer 5,400 -16,747,320 BBB Positive Bahrain Member 37,600 -14,930,833 AA+ Negative Brazil Member 11,600 -29,796,166 BBB Stable Brazil Member 11,600 -29,796,166 BBB Stable Canada Member 40,300 -1,611,380 AAA Stable Canada Member 4,000 874,459 B+ Stable Chile Member 4,000 874,459 B+ Stable Chile Member 16,100 -1,561,199 BBB- Stable Chile Member 10,100 -1,561,199 BBB- Stable Costa Rica Member 11,500 5,807,509 BBH Stable Cyprus Member 29,100 6,623,337 BBB Negative Cyprus Member <td< td=""><td></td><td></td><td></td><td></td><td></td><td colspan="2"></td></td<>										
Azerbaijan Observer 5,400 -16,747,320 BBB - BBB Positive Bahrain 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 Canada Member 40,300 -1,611,380 AAA Stable Cape Verde Member 40,000 874,459 B+ Stable Chile Member 16,100 -14,558,218 A+ Stable China Member 10,100 -156,705,311 A+ Stable Colombia Member 11,500 5,807,509 BB+ Stable Costa Rica Member 18,300 8,108,103 BBB- Negative Cyprus Member 29,100 6,623,337 BBB Negative Cyprus Member 8,300 1,993,819 B- Stable Ecuador Member										
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 Chile Member 10,100 -1,567,05,311 A+ Stable Chile Member 10,100 -5,807,509 BB+ Stable Costa Rica Member 11,500 5,807,509 BB+ Stable Cyprus Member 29,100 6,623,337 BBB Negative Cyprus Member 40,200 -15,057,343 AAA Stable Ecuador Member										
Belgium Member 37,600 -14,930,833 AA+ Negative Brazil Member 11,600 -29,796,166 BBB Stable Bulgaria 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 16,100 -1,561,199 BBB- Stable Colombia Member 10,100 -1,561,199 BBB- Stable Costa Rica Member 11,500 5,807,509 BB+ Stable Croatia Member 29,100 6,623,337 BBB Negative Cyprus 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 El Salvador Member	U									
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 Chile Member 10,100 -15,611,199 BBB- Stable Costa Rica Member 11,500 5,807,509 BB+ Stable Costa Rica Member 11,500 5,807,509 BBB- Negative Cyprus Member 29,100 6,623,337 BBB Negative Cyprus Member 29,100 6,623,337 BBB Negative Czech Republic Member 25,900 -10,749,467 A+ Positive Denmark Member 8,300 1,993,819 B- Stable El Salvador Mem	***									
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 10,100 -156,705,311 A + Stable Costa Rica Member 11,500 5,807,509 BBB - Stable Costa Rica Member 18,300 8,108,103 BBB - Negative Cyprus Member 29,100 6,623,337 BBB Negative Negative Cyprus 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 Estonia Member 7,600 5,019,363 BB Stable Estonia						_				
Canada Member 40,300 -1,611,380 AAA Stable Cape Verde Member 4,000 874,459 B+ Stable Chile Member 16,100 -14,588,218 A+ Stable China Member 10,100 -1,561,199 BBB- Stable Costa Rica Member 11,500 5,807,509 BB+ Stable Croatia Member 118,300 8,108,103 BBB- Negative Cyprus Member 29,100 6,623,337 BBB Negative Cyprus Member 40,200 -15,057,343 AAA Stable Ecuador Member 40,200 -15,057,343 BA Stable Egypt Member 6,500 28,486,933 BB Negative Estonia Member 7,600 5,019,363 BB Stable Estonia Member 35,000 118,841,542 AAA Stable France Member										
Cape Verde Member 4,000 874,459 B+ Stable Chile Member 16,100 -14,558,218 A+ Stable Chile Member 16,100 -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 Egypt Member 6,500 28,486,933 BB Negative El Salvador Member 7,600 5,019,363 BB Stable France Member 35,000 118,841,542 AAA Stable Georgia Member </td <td>-</td> <td></td> <td></td> <td></td> <td></td> <td colspan="2"></td>	-									
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 29,100 6,623,337 BBB Negative Cyprus Member 29,100 6,623,337 BBB Negative Cyprus Member 40,200 -15,057,343 AAA Stable Denmark Member 40,200 -15,057,343 AAA Stable Ecuador Member 8,300 1,993,819 B- Stable Esypt Member 7,600 5,019,363 BB Stable Estonia Member 7,600 5,019,363 BB Stable Estonia Member 35,000 118,841,542 AAA Stable France Member 5,4										
China Member Colombia 8,400 Member -156,705,311 D,199 A+ Stable BBB- Stable Stable Costa Rica Member 11,500 5,807,509 BB+ Stable Croatia 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 6,500 28,486,933 BB Stable Egypt Member 7,600 5,019,363 BB Stable Estonia Member 35,000 118,841,542 AAA Stable France Member 35,000 118,841,542 AAA Stable Georgia Member 37,900 -213,367,685 AAA Stable Hungary Member 39,500 -50,4939 BB+ Stable Negative </td <td>*</td> <td></td> <td></td> <td>· ·</td> <td>A+</td> <td></td>	*			· ·	A+					
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 40,200 -10,749,467 A+ Positive Denmark Member 40,200 -15,057,343 AAA Stable Ecuador 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 Estonia Member 35,000 118,841,542 AAA Stable Georgia Member 37,900 -213,367,685 AAA Stable Hungary Member 39,500 -504,939 BB+ Stable Ireland Memb										
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 6,500 28,486,933 BB Negative El Salvador Member 6,500 28,486,933 BB Negative Estonia 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 37,900 -213,367,685 AAA Stable Hungary Member 19,600 -9,421,301 BBB- Negative Israel Mem	- "									
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 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 37,900 —213,367,685 AAA Stable Hungary Member 19,600 —9,421,301 BBB — Negative Ireland Member 39,500 —59,093,323 BBB+ Negative Israel Member 31,000 6,722,291 A Stable Italy Me	Costa Rica	Member			BB+	Stable				
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 Israel Member 31,000 6,722,291 A Stable Italy Member	Croatia	Member				Negative				
Czech Republic Denmark Member 25,900 -10,749,467 A+ Positive Positive Positive Positive Benmark Member 40,200 -15,057,343 AAA Stable Ecuador Member 8,300 1,993,819 B- Stable Esypt Member 6,500 28,486,933 BB Negative El Salvador Member 7,600 5,019,363 BB Negative Estonia Member 20,200 802,234 A+ Stable France Member 35,000 118,841,542 AAA Stable Georgia 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 31,000 6,722,291 A Stable Israel Member 31,000 6,722,291 A Stable Kazakhst					BBB					
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 38,000 -504,939 BB + Stable Ireland Member 38,000 -59,093,323 BBB + Negative Israel Member 31,000 6,722,291 A Stable Italy Member 34,300 31,675,424 AA Negative Kenya Member										
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 31,000 6,722,291 A Stable Israel Member 31,000 33,568,668 A+ Negative Israel Member 34,300 31,675,424 AA Negative Kazakhstan Observer <	•									
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 31,000 6,722,291 A Stable Israel Member 31,000 6,722,291 A Stable Italy Member 34,300 31,675,424 AA Negative Kazakhstan Observer 13,000 -33,220,437 BBB Positive Kenya Member <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td></td<>										
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 31,000 6,722,291 A Stable Israel Member 30,100 33,568,668 A+ Negative Kazakhstan Observer 13,000 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				4 4						
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 -50,4939 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 1,400 2,448,053 BBB- Positive Latvia Member		Member			BB					
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 Labanon Observer										
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										
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 <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>										
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	_									
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 Portiugal Member	-									
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 4,100 15,667,209 BB+ Stable Poland Member						_				
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 Poland Member 4,100 15,667,209 BB+ Stable Portugal Member <t< td=""><td></td><td></td><td></td><td>′</td><td></td><td></td></t<>				′						
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 Poland Member 4,100 15,667,209 BB+ Stable Portugal Member 20,100 19,552,976 A- Stable Portugal Member	Israel	Member			A					
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 Portugal Member 20,100 19,552,976 A- Stable Portugal Member 31,700 -37,509,141 A+ Stable Romania Member	Italv	Member			A+	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 12,300 13,323,709 BBB- Stable Russia Observ	•	Member			AA	_				
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 Rwanda Member		Observer			BBB					
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 Rwanda Member 1,300 939,222 B Stable Serbia Observer </td <td>Kenya</td> <td>Member</td> <td></td> <td>9,174,198</td> <td>$\mathrm{B}+$</td> <td>Stable</td>	Kenya	Member		9,174,198	$\mathrm{B}+$	Stable				
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 Rwanda Member 1,300 939,222 B Stable Serbia Observer 10,700 8,275,693 BB- Stable	Latvia	Member			BBB-	Positive				
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	Lebanon	Observer			В	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	Lithuania	Member			BBB	Positive				
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	Malaysia	Member	15,600	-39,420,064	A-	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	•	Member	15,100		BBB					
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	Peru	Member			BBB					
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	Philippines									
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		Member			A-	Stable				
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	Portugal	Member			BBB-	Negative				
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	-	Member	31,700		A+	_				
Rwanda Member 1,300 939,222 B Stable Serbia Observer 10,700 8,275,693 BB- Stable	Romania	Member			BBB-	Stable				
Rwanda Member 1,300 939,222 B Stable Serbia Observer 10,700 8,275,693 BB- Stable	Russia	Observer		1 1		Positive				
Serbia Observer 10,700 8,275,693 BB- Stable	Rwanda	Member								
Seycnelles Observer 24,700 666,026 B Stable	Seychelles	Observer	24,700	666,026	В	Stable				
Singapore Member 59,900 -27,110,421 AAA Stable	•	Member			AAA					
Slovakia Member 23,400 -2,110,626 A+ Stable										
Slovenia Member 29,100 2,310,617 AA Negative	Slovenia	Member			AA-	Negative				
South Africa Member 11,000 3,321,801 BBB+ Stable	South Africa	Member	11,000	3,321,801	BBB+	Stable				

1.2 Data 7

Sweden Switzerland Thailand	Member Member Member	40,600 43,400 9,700	-10,903,251 -27,197,873 2,049,669	AAA AAA BBB	Stable Stable 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.

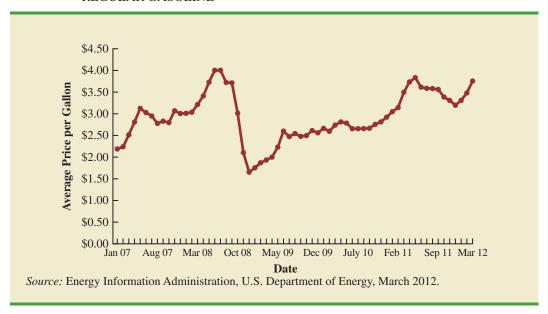
Cross-Sectional and Time Series Data

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. 1.2 Data 9

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



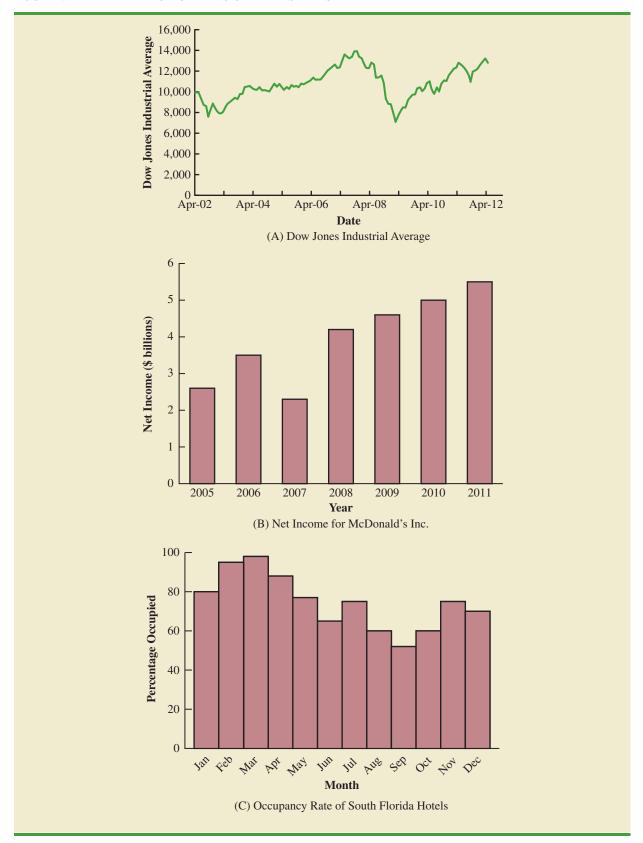
The graph in Panel (B) shows the net income of McDonald's Inc. from 2005 to 2011. The declining economic conditions in 2008 and 2009 were actually beneficial to McDonald's as the company's net income rose to all-time highs. The growth in McDonald's net income showed that the company was thriving during 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.
- 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.

FIGURE 1.2 A VARIETY OF GRAPHS OF TIME SERIES DATA



1.3 Data Sources



Data Sources

Data can be obtained from existing sources or from surveys and experimental studies designed to collect new data.

Existing Sources

In some cases, data needed for a particular application already exist. Companies maintain a variety of databases about their employees, customers, and business operations. Data on employee salaries, ages, and years of experience can usually be obtained from internal personnel records. Other internal records contain data on sales, advertising expenditures, distribution costs, inventory levels, and production quantities. Most companies also maintain detailed data about their customers. Table 1.2 shows some of the data commonly available from internal company records.

Organizations that specialize in collecting and maintaining data make available substantial amounts of business and economic data. Companies access these external data sources through leasing arrangements or by purchase. Dun & Bradstreet, Bloomberg, and Dow Jones & Company are three firms that provide extensive business database services to clients. ACNielsen and Information Resources, Inc., built successful businesses collecting and processing data that they sell to advertisers and product manufacturers.

Data are also available from a variety of industry associations and special interest organizations. The Travel Industry Association of America maintains travel-related information such as the number of tourists and travel expenditures by states. Such data would be of interest to firms and individuals in the travel industry. The Graduate Management Admission Council maintains data on test scores, student characteristics, and graduate management education programs. Most of the data from these types of sources are available to qualified users at a modest cost.

The Internet is an important source of data and statistical information. Almost all companies maintain websites that provide general information about the company as well as data on sales, number of employees, number of products, product prices, and product specifications. In addition, a number of companies now specialize in making information available over the Internet. As a result, one can obtain access to stock quotes, meal prices at restaurants, salary data, and an almost infinite variety of information.

Government agencies are another important source of existing data. For instance, the U.S. Department of Labor maintains considerable data on employment rates, wage rates, size of the labor force, and union membership. Table 1.3 lists selected governmental agencies

TABLE 1.2 EXAMPLES OF DATA AVAILABLE FROM INTERNAL COMPANY RECORDS

Source	Some of the Data Typically Available
Employee records	Name, address, social security number, salary, number of vacation days, number of sick days, and bonus
Production records	Part or product number, quantity produced, direct labor cost, and materials cost
Inventory records	Part or product number, number of units on hand, reorder level, economic order quantity, and discount schedule
Sales records	Product number, sales volume, sales volume by region, and sales volume by customer type
Credit records	Customer name, address, phone number, credit limit, and accounts receivable balance
Customer profile	Age, gender, income level, household size, address, and preferences

TABLE 1.3 EXAMPLES OF DATA AVAILABLE FROM SELECTED GOVERNMENT AGENCIES

Government Agency	Some of the Data Available
Census Bureau	Population data, number of households, and household income
Federal Reserve Board	Data on the money supply, installment credit, exchange rates, and discount rates
Office of Management and Budget	Data on revenue, expenditures, and debt of the federal government
Department of Commerce	Data on business activity, value of shipments by industry, level of profits by industry, and growing and declining industries
Bureau of Labor Statistics	Consumer spending, hourly earnings, unemployment rate, safety records, and international statistics

and some of the data they provide. Most government agencies that collect and process data also make the results available through a website. Figure 1.3 shows the homepage for the U.S. Bureau of Labor Statistics website.

Statistical Studies

Sometimes the data needed for a particular application are not available through existing sources. In such cases, the data can often be obtained by conducting a statistical study. Statistical studies can be classified as either *experimental* or *observational*.

In an experimental study, a variable of interest is first identified. Then one or more other variables are identified and controlled so that data can be obtained about how they influence the variable of interest. For example, a pharmaceutical firm might be interested in conducting an experiment to learn how a new drug affects blood pressure. Blood pressure is the variable of interest in the study. The dosage level of the new drug is another variable that

statistical study ever conducted is believed to be the 1954 Public Health Service experiment for the Salk polio vaccine. Nearly 2 million children in grades 1, 2, and 3 were selected from throughout the United States.

The largest experimental

FIGURE 1.3 U.S. BUREAU OF LABOR STATISTICS HOMEPAGE



1.3 Data Sources

is hoped to have a causal effect on blood pressure. To obtain data about the effect of the new drug, researchers select a sample of individuals. The dosage level of the new drug is controlled, as different groups of individuals are given different dosage levels. Before and after data on blood pressure are collected for each group. Statistical analysis of the experimental data can help determine how the new drug affects blood pressure.

Nonexperimental, or observational, statistical studies make no attempt to control the variables of interest. A survey is perhaps the most common type of observational study. For instance, in a personal interview survey, research questions are first identified. Then a questionnaire is designed and administered to a sample of individuals. Some restaurants use observational studies to obtain data about customer opinions on the quality of food, quality of service, atmosphere, and so on. A customer opinion questionnaire used by Chops City Grill in Naples, Florida, is shown in Figure 1.4. Note that the customers who fill out the questionnaire are asked to provide ratings for 12 variables, including overall experience, greeting by hostess, manager (table visit), overall service, and so on. The response categories of excellent, good, average, fair, and poor provide categorical data that enable Chops City Grill management to maintain high standards for the restaurant's food and service.

Studies of smokers and nonsmokers are observational studies because researchers do not determine or control who will smoke and who will not smoke.

Anyone wanting to use data and statistical analysis as aids to decision making must be aware of the time and cost required to obtain the data. The use of existing data sources is desirable when data must be obtained in a relatively short period of time. If important data are not readily available from an existing source, the additional time and cost involved in obtaining the data must be taken into account. In all cases, the decision maker should

FIGURE 1.4 CUSTOMER OPINION QUESTIONNAIRE USED BY CHOPS CITY GRILL RESTAURANT IN NAPLES, FLORIDA

			Y 6		<u></u>	
Date:				1	Server Name:	
Our custor survey card, so we cardesk or return by main	iii octici s	cive yo	priority. I our needs	Please . You	take a moment to fill out o may return this card to the	ur front
SERVICE SURVEY	Excellent	Good	Average	Fair	Poor	
Overall Experience						
Greeting by Hostess						
Manager (Table Visit)						
Overall Service						
Professionalism						
Menu Knowledge						
Friendliness						
Wine Selection						
Menu Selection						
Food Quality						
Food Presentation						
Value for \$ Spent						
What comments could	you give u	s to imp	prove our	restau	rant?	

consider the contribution of the statistical analysis to the decision-making process. The cost of data acquisition and the subsequent statistical analysis should not exceed the savings generated by using the information to make a better decision.

Data Acquisition Errors

Managers should always be aware of the possibility of data errors in statistical studies. Using erroneous data can be worse than not using any data at all. An error in data acquisition occurs whenever the data value obtained is not equal to the true or actual value that would be obtained with a correct procedure. Such errors can occur in a number of ways. For example, an interviewer might make a recording error, such as a transposition in writing the age of a 24-year-old person as 42, or the person answering an interview question might misinterpret the question and provide an incorrect response.

Experienced data analysts take great care in collecting and recording data to ensure that errors are not made. Special procedures can be used to check for internal consistency of the data. For instance, such procedures would indicate that the analyst should review the accuracy of data for a respondent shown to be 22 years of age but reporting 20 years of work experience. Data analysts also review data with unusually large and small values, called outliers, which are candidates for possible data errors. In Chapter 3 we present some of the methods statisticians use to identify outliers.

Errors often occur during data acquisition. Blindly using any data that happen to be available or using data that were acquired with little care can result in misleading information and bad decisions. Thus, taking steps to acquire accurate data can help ensure reliable and valuable decision-making information.



Descriptive Statistics

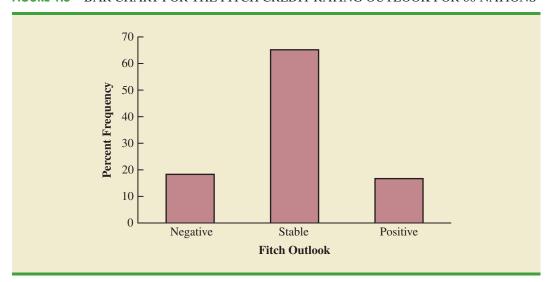
Most of the statistical information in newspapers, magazines, company reports, and other publications consists of data that are summarized and presented in a form that is easy for the reader to understand. Such summaries of data, which may be tabular, graphical, or numerical, are referred to as **descriptive statistics**.

Refer to the data set in Table 1.1 showing data for 60 nations that participate in the World Trade Organization. Methods of descriptive statistics can be used to summarize these data. For example, consider the variable Fitch Outlook that indicates the direction the nation's credit rating is likely to move over the next two years. The Fitch Outlook is recorded as being negative, stable, or positive. A tabular summary of the data showing the number of nations with each of the Fitch Outlook ratings is shown in Table 1.4. A graphical summary of the same data, called a bar chart, is shown in Figure 1.5. These types of summaries make the data easier to interpret. Referring to Table 1.4 and Figure 1.5, we can see that the majority of Fitch Outlook credit ratings are stable, with 65% of the nations

TABLE 1.4 FREQUENCIES AND PERCENT FREQUENCIES FOR THE FITCH CREDIT RATING OUTLOOK OF 60 NATIONS

10	16.7
39	65.0
11	18.3
	39

FIGURE 1.5 BAR CHART FOR THE FITCH CREDIT RATING OUTLOOK FOR 60 NATIONS

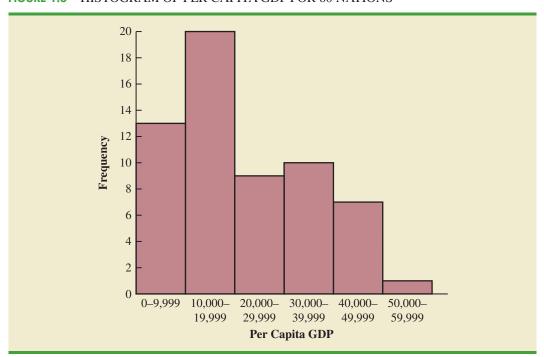


having this rating. Negative and positive outlook credit ratings are similar with slightly more nations having a negative outlook (18.3%) than a positive outlook (16.7%).

A graphical summary of the data for quantitative variable Per Capita GDP in Table 1.1, called a histogram, is provided in Figure 1.6. Using the histogram, it is easy to see that Per Capita GDP for the 60 nations ranges from \$0 to \$60,000, with the highest concentration between \$10,000 and \$20,000. Only one nation had a Per Capita GDP exceeding \$50,000.

In addition to tabular and graphical displays, numerical descriptive statistics are used to summarize data. The most common numerical measure is the average, or mean. Using

FIGURE 1.6 HISTOGRAM OF PER CAPITA GDP FOR 60 NATIONS



the data on Per Capita GDP for the 60 nations in Table 1.1, we can compute the average by adding Per Capita GDP for all 60 nations and dividing the total by 60. Doing so provides an average Per Capita GDP of \$21,387. This average provides a measure of the central tendency, or central location of the data.

There is a great deal of interest in effective methods for developing and presenting descriptive statistics. Chapters 2 and 3 devote attention to the tabular, graphical, and numerical methods of descriptive statistics.



Statistical Inference

Many situations require information about a large group of elements (individuals, companies, voters, households, products, customers, and so on). But, because of time, cost, and other considerations, data can be collected from only a small portion of the group. The larger group of elements in a particular study is called the **population**, and the smaller group is called the **sample**. Formally, we use the following definitions.

POPULATION

A population is the set of all elements of interest in a particular study.

SAMPLE

A sample is a subset of the population.

The U.S. government conducts a census every 10 years. Market research firms conduct sample surveys every day. The process of conducting a survey to collect data for the entire population is called a **census**. The process of conducting a survey to collect data for a sample is called a **sample survey**. As one of its major contributions, statistics uses data from a sample to make estimates and test hypotheses about the characteristics of a population through a process referred to as **statistical inference**.

As an example of statistical inference, let us consider the study conducted by Norris Electronics. Norris manufactures a high-intensity lightbulb used in a variety of electrical products. In an attempt to increase the useful life of the lightbulb, the product design group developed a new lightbulb filament. In this case, the population is defined as all lightbulbs that could be produced with the new filament. To evaluate the advantages of the new filament, 200 bulbs with the new filament were manufactured and tested. Data collected from this sample showed the number of hours each lightbulb operated before filament burnout. See Table 1.5.

Suppose Norris wants to use the sample data to make an inference about the average hours of useful life for the population of all lightbulbs that could be produced with the new filament. Adding the 200 values in Table 1.5 and dividing the total by 200 provides the sample average lifetime for the lightbulbs: 76 hours. We can use this sample result to estimate that the average lifetime for the lightbulbs in the population is 76 hours. Figure 1.7 provides a graphical summary of the statistical inference process for Norris Electronics.

Whenever statisticians use a sample to estimate a population characteristic of interest, they usually provide a statement of the quality, or precision, associated with the estimate. For the Norris example, the statistician might state that the point estimate of the average

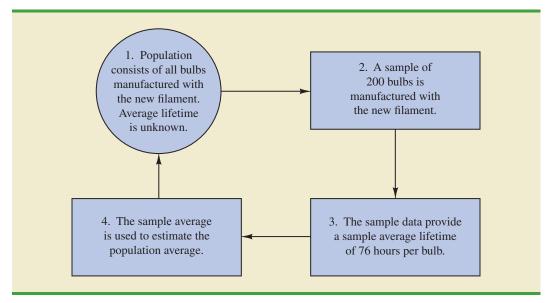
1.5 Statistical Inference

TABLE 1.5 HOURS UNTIL BURNOUT FOR A SAMPLE OF 200 LIGHTBULBS FOR THE NORRIS ELECTRONICS EXAMPLE



40=			0.7		=0	0.4		0.0	
107	73	68	97	76	79	94	59	98	57
54	65	71	70	84	88	62	61	79	98
66	62	79	86	68	74	61	82	65	98
62	116	65	88	64	79	78	79	77	86
74	85	73	80	68	78	89	72	58	69
92	78	88	77	103	88	63	68	88	81
75	90	62	89	71	71	74	70	74	70
65	81	75	62	94	71	85	84	83	63
81	62	79	83	93	61	65	62	92	65
83	70	70	81	77	72	84	67	59	58
78	66	66	94	77	63	66	75	68	76
90	78	71	101	78	43	59	67	61	71
96	75	64	76	72	77	74	65	82	86
66	86	96	89	81	71	85	99	59	92
68	72	77	60	87	84	75	77	51	45
85	67	87	80	84	93	69	76	89	75
83	68	72	67	92	89	82	96	77	102
74	91	76	83	66	68	61	73	72	76
73	77	79	94	63	59	62	71	81	65
73	63	63	89	82	64	85	92	64	73

FIGURE 1.7 THE PROCESS OF STATISTICAL INFERENCE FOR THE NORRIS ELECTRONICS EXAMPLE



lifetime for the population of new lightbulbs is 76 hours with a margin of error of ± 4 hours. Thus, an interval estimate of the average lifetime for all lightbulbs produced with the new filament is 72 hours to 80 hours. The statistician can also state how confident he or she is that the interval from 72 hours to 80 hours contains the population average.



Computers and Statistical Analysis

Minitab and Excel data sets and the Excel add-in StatTools are available on the website for this text. Statisticians frequently use computer software to perform the statistical computations required with large amounts of data. For example, computing the average lifetime for the 200 lightbulbs in the Norris Electronics example (see Table 1.5) would be quite tedious without a computer. To facilitate computer usage, many of the data sets in this book are available on the website that accompanies the text. The data files may be downloaded in either Minitab or Excel formats. In addition, the Excel add-in StatTools can be downloaded from the website. End-of-chapter appendixes cover the step-by-step procedures for using Minitab, Excel, and the Excel add-in StatTools to implement the statistical techniques presented in the chapter.

1.7

Data Mining

With the aid of magnetic card readers, bar code scanners, and point-of-sale terminals, most organizations obtain large amounts of data on a daily basis. And, even for a small local restaurant that uses touch screen monitors to enter orders and handle billing, the amount of data collected can be substantial. For large retail companies, the sheer volume of data collected is hard to conceptualize, and figuring out how to effectively use these data to improve profitability is a challenge. Mass retailers such as Walmart capture data on 20 to 30 million transactions every day, telecommunication companies such as France Telecom and AT&T generate over 300 million call records per day, and Visa processes 6800 payment transactions per second or approximately 600 million transactions per day. Storing and managing the transaction data is a substantial undertaking.

The term *data warehousing* is used to refer to the process of capturing, storing, and maintaining the data. Computing power and data collection tools have reached the point where it is now feasible to store and retrieve extremely large quantities of data in seconds. Analysis of the data in the warehouse may result in decisions that will lead to new strategies and higher profits for the organization.

The subject of **data mining** deals with methods for developing useful decision-making information from large databases. Using a combination of procedures from statistics, mathematics, and computer science, analysts "mine the data" in the warehouse to convert it into useful information, hence the name *data mining*. Dr. Kurt Thearling, a leading practitioner in the field, defines data mining as "the automated extraction of predictive information from (large) databases." The two key words in Dr. Thearling's definition are "automated" and "predictive." Data mining systems that are the most effective use automated procedures to extract information from the data using only the most general or even vague queries by the user. And data mining software automates the process of uncovering hidden predictive information that in the past required hands-on analysis.

The major applications of data mining have been made by companies with a strong consumer focus, such as retail businesses, financial organizations, and communication companies. Data mining has been successfully used to help retailers such as Amazon and Barnes & Noble determine one or more related products that customers who have already purchased a specific product are also likely to purchase. Then, when a customer logs on to the company's website and purchases a product, the website uses pop-ups to alert the customer about additional products that the customer is likely to purchase. In another application, data mining may be used to identify customers who are likely to spend more than \$20 on a particular shopping trip. These customers may then be identified as the ones to receive special e-mail or regular mail discount offers to encourage them to make their next shopping trip before the discount termination date.

Data mining is a technology that relies heavily on statistical methodology such as multiple regression, logistic regression, and correlation. But it takes a creative integration of all

Statistical methods play an important role in data mining, both in terms of discovering relationships in the data and predicting future outcomes. However, a thorough coverage of data mining and the use of statistics in data mining are outside the scope of this text.

these methods and computer science technologies involving artificial intelligence and machine learning to make data mining effective. A substantial investment in time and money is required to implement commercial data mining software packages developed by firms such as Oracle, Teradata, and SAS. The statistical concepts introduced in this text will be helpful in understanding the statistical methodology used by data mining software packages and enable you to better understand the statistical information that is developed.

Because statistical models play an important role in developing predictive models in data mining, many of the concerns that statisticians deal with in developing statistical models are also applicable. For instance, a concern in any statistical study involves the issue of model reliability. Finding a statistical model that works well for a particular sample of data does not necessarily mean that it can be reliably applied to other data. One of the common statistical approaches to evaluating model reliability is to divide the sample data set into two parts: a training data set and a test data set. If the model developed using the training data is able to accurately predict values in the test data, we say that the model is reliable. One advantage that data mining has over classical statistics is that the enormous amount of data available allows the data mining software to partition the data set so that a model developed for the training data set may be tested for reliability on other data. In this sense, the partitioning of the data set allows data mining to develop models and relationships and then quickly observe if they are repeatable and valid with new and different data. On the other hand, a warning for data mining applications is that with so much data available, there is a danger of overfitting the model to the point that misleading associations and cause/effect conclusions appear to exist. Careful interpretation of data mining results and additional testing will help avoid this pitfall.

1.8

Ethical Guidelines for Statistical Practice

Ethical behavior is something we should strive for in all that we do. Ethical issues arise in statistics because of the important role statistics plays in the collection, analysis, presentation, and interpretation of data. In a statistical study, unethical behavior can take a variety of forms including improper sampling, inappropriate analysis of the data, development of misleading graphs, use of inappropriate summary statistics, and/or a biased interpretation of the statistical results.

As you begin to do your own statistical work, we encourage you to be fair, thorough, objective, and neutral as you collect data, conduct analyses, make oral presentations, and present written reports containing information developed. As a consumer of statistics, you should also be aware of the possibility of unethical statistical behavior by others. When you see statistics in newspapers, on television, on the Internet, and so on, it is a good idea to view the information with some skepticism, always being aware of the source as well as the purpose and objectivity of the statistics provided.

The American Statistical Association, the nation's leading professional organization for statistics and statisticians, developed the report "Ethical Guidelines for Statistical Practice" to help statistical practitioners make and communicate ethical decisions and assist students in learning how to perform statistical work responsibly. The report contains 67 guidelines organized into eight topic areas: Professionalism; Responsibilities to Funders, Clients, and Employers; Responsibilities in Publications and Testimony; Responsibilities to Research Subjects; Responsibilities to Research Team Colleagues; Responsibilities to Other Statisticians or Statistical Practitioners; Responsibilities Regarding Allegations of Misconduct; and Responsibilities of Employers Including Organizations, Individuals, Attorneys, or Other Clients Employing Statistical Practitioners.

²American Statistical Association "Ethical Guidelines for Statistical Practice," 1999.