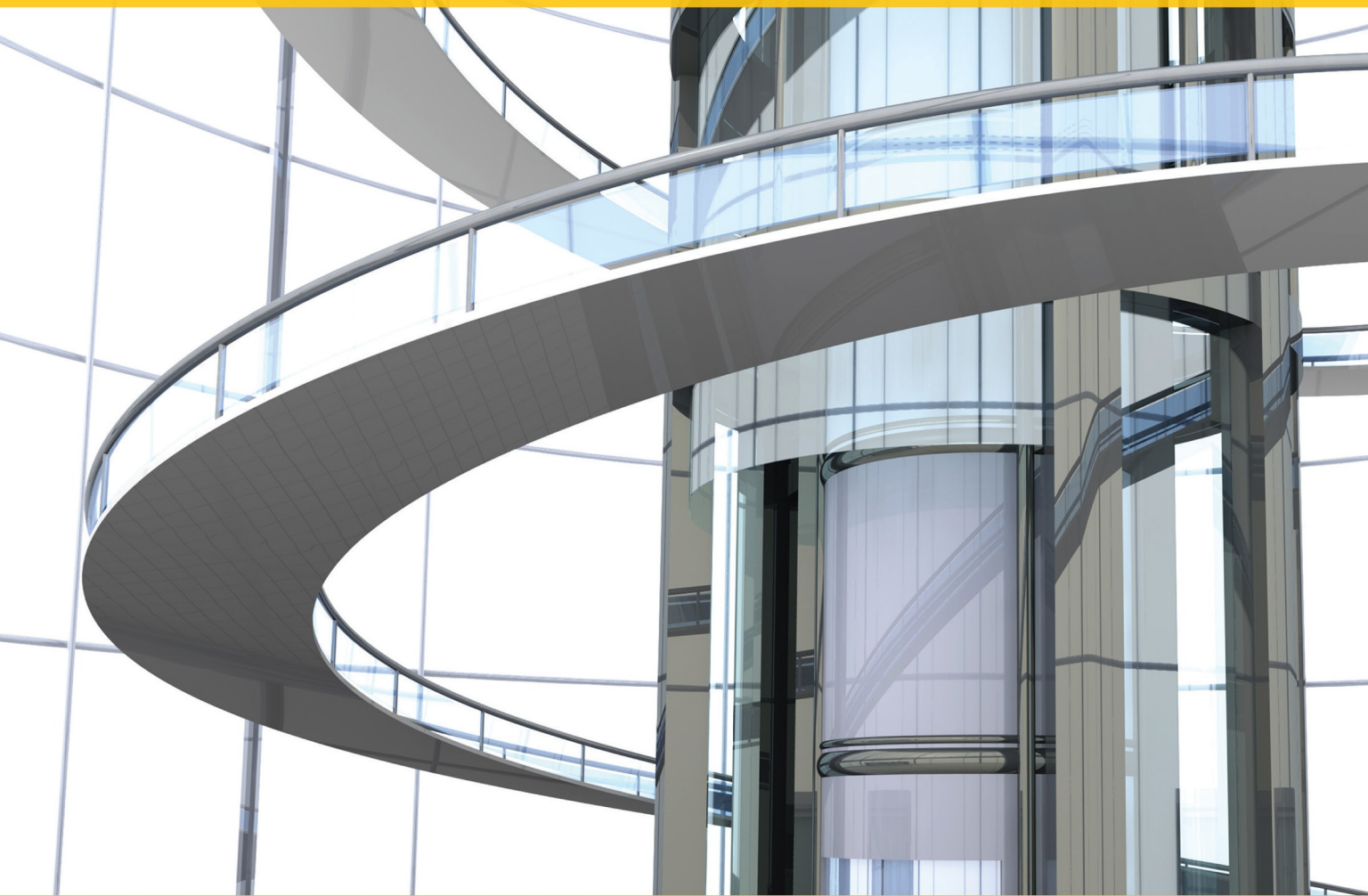


Anderson * Sweeney * Williams * Camm * Cochran



Essentials of Modern Business Statistics^{7e}

with Microsoft[®] Office Excel[®]

Want to turn your C into an A? Obviously, right?

But the right way to go about it isn't always so obvious. Go digital to get the grades. MindTap's customizable study tools and eTextbook give you everything you need all in one place.

Engage with your course content, enjoy the flexibility of studying anytime and anywhere, stay connected to assignment due dates and instructor notifications with the MindTap Mobile app...

and most of all...EARN BETTER GRADES.



TO GET STARTED VISIT
WWW.CENGAGE.COM/STUDENTS/MINDTAP



Microsoft® Office Excel® Functions

Function	Description
AVERAGE	Returns the arithmetic mean of a range its arguments.
BINOM.DIST	Returns the individual term binomial distribution probability.
CHISQ.DIST	Returns a probability from the chi-squared distribution.
CHISQ.DIST.RT	Returns the one-tailed probability of the chi-squared distribution.
CHISQ.INV	Returns the inverse of the left-tailed probability of the chi-squared distribution.
CHISQ.TEST	Returns the value from the chi-squared distribution for the statistic and the degrees of freedom.
CONFIDENCE.NORM	Returns the confidence interval for a population mean using the normal distribution.
CORREL	Returns the correlation coefficient between two data sets.
COUNT	Returns the number of cells in the range that contain numbers.
COUNTA	Returns the number of non-blank cells in the range.
COUNTIF	Returns the number of cells in a range that meet the specified criterion.
COVARIANCE.S	Returns the sample covariance.
EXPON.DIST	Returns a probability from the exponential distribution.
F.DIST.RT	Returns the right-tailed probability from the F distribution.
GEOMEAN	Returns the geometric mean of a range of cells.
HYPGEOM.DIST	Returns a probability from the hypergeometric distribution.
MAX	Returns the maximum value of the values in a range of cells.
MMEDIAN	Returns the median value of the values in a range of cells.
MIN	Returns the minimum value of the values in a range of cells.
MODE.SNGL	Returns the most-frequently occurring value in a range of cells.
NORM.S.DIST	Returns a probability from a standard normal distribution.
NORM.S.INV	Inverse of the standard normal distribution.
PERCENTILE.EXC	Returns the specified percentile of the values in a range of cells.
POISSON.DIST	Returns a probability from the poisson distribution.
POWER	Returns the result of a number raised to a power.
QUARTILE.EXC	Returns the specified quartile of the values in a range of cells.
RAND	Returns a real number from the uniform distribution between 0 and 1.
SQRT	Returns the positive square root of its argument.
STDEV.S	Returns the sample standard deviation of the values in a range of cells.
SUM	Returns the sum of the values in a range of cells.
SUMPRODUCT	Returns the sum of the products of the paired elements of the values in two ranges of cells.
T.DIST	Returns a left-tailed probability of the t distribution.
T.INV.2T	Returns the two-tailed inverse of the student's t-distribution.
VAR.S	Returns the sample variance of the values in a range of cells.



Essentials of Modern Business Statistics^{7e}

with Microsoft[®] Office Excel[®]

David R. Anderson
University of Cincinnati

Dennis J. Sweeney
University of Cincinnati

Thomas A. Williams
Rochester Institute
of Technology

Jeffrey D. Camm
Wake Forest University

James J. Cochran
University of Alabama



Australia • Brazil • Mexico • Singapore • United Kingdom • United States

**Essentials of Modern Business Statistics,
Seventh Edition****David R. Anderson, Dennis J. Sweeney,
Thomas A. Williams, Jeffrey D. Camm,
James J. Cochran**Senior Vice President, General Manager,
Social Science, Business and Humanities:
Erin Joyner

Product Director: Michael Schenk

Product Team Manager: Joe Sabatino

Sr. Product Manager: Aaron Arnsperger

Content Developer: Anne Merrill

Product Assistant: Renee Schnee

Digital Content Designer: Brandon Foltz

Sr. Marketing Manager: Nathan Anderson

Sr. Content Project Manager: Colleen Farmer

Manufacturing Planner: Ron Montgomery

Production Service: MPS Limited

Sr. Art Director: Michelle Kunkler

Internal Designer: Beckmeyer Design

Cover Designer: Beckmeyer Design

Cover Image: Spectral-Design/
Shutterstock.com

Intellectual Property:

Analyst: Brittani Morgan

Project Manager: Nicholas Barrows

© 2018, 2016 Cengage Learning®

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

For product information and technology assistance, contact us at
Cengage Learning Customer & Sales Support, 1-800-354-9706For 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.comMicrosoft Excel® is a registered trademark of Microsoft Corporation.
© 2016 Microsoft.

Unless otherwise noted, all items © Cengage Learning.

Library of Congress Control Number: 2017930365

Package ISBN: 978-1-337-29829-2

Book only ISBN: 978-1-337-29835-3

Cengage Learning20 Channel Center Street
Boston, MA 02210
USACengage Learning is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **www.cengage.com**.Cengage Learning products are represented in Canada by
Nelson Education, Ltd.To learn more about Cengage Learning Solutions, visit
www.cengage.comPurchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**

Brief Contents

Preface	xvii
Chapter 1	Data and Statistics 1
Chapter 2	Descriptive Statistics: Tabular and Graphical Displays 35
Chapter 3	Descriptive Statistics: Numerical Measures 108
Chapter 4	Introduction to Probability 180
Chapter 5	Discrete Probability Distributions 228
Chapter 6	Continuous Probability Distributions 285
Chapter 7	Sampling and Sampling Distributions 317
Chapter 8	Interval Estimation 363
Chapter 9	Hypothesis Tests 405
Chapter 10	Inference About Means and Proportions with Two Populations 455
Chapter 11	Inferences About Population Variances 502
Chapter 12	Tests of Goodness of Fit, Independence, and Multiple Proportions 530
Chapter 13	Experimental design and Analysis of Variance 564
Chapter 14	Simple Linear Regression 620
Chapter 15	Multiple Regression 706
Appendix A	References and Bibliography 762
Appendix B	Tables 764
Appendix C	Summation Notation 775
Appendix D	Self-Test Solutions and Answers to Even-Numbered Exercises (online) 777
Appendix E	Microsoft Excel 2016 and Tools for Statistical Analysis 778
Index	786

Contents

Preface xvii

Chapter 1 Data and Statistics 1

Statistics in Practice: *Bloomberg Businessweek* 2

1.1 Applications in Business and Economics 3

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

1.2 Data 5

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

1.3 Data Sources 11

- Existing Sources 11
- Observational Study 12
- Experiment 13
- Time and Cost Issues 13
- Data Acquisition Errors 13

1.4 Descriptive Statistics 14

1.5 Statistical Inference 16

1.6 Statistical Analysis Using Microsoft Excel 17

- Data Sets and Excel Worksheets 18
- Using Excel for Statistical Analysis 18

1.7 Analytics 21

1.8 Big Data and Data Mining 22

1.9 Ethical Guidelines for Statistical Practice 24

Summary 25

Glossary 26

Supplementary Exercises 27

Chapter 2 Descriptive Statistics: Tabular and Graphical Displays 35

Statistics in Practice: Colgate-Palmolive Company 36

2.1 Summarizing Data for a Categorical Variable 37

Frequency Distribution 37

Relative Frequency and Percent Frequency Distributions 38

Using Excel to Construct a Frequency Distribution, a Relative Frequency Distribution, and a Percent Frequency Distribution 39

Bar Charts and Pie Charts 40

Using Excel to Construct a Bar Chart and a Pie Chart 42

2.2 Summarizing Data for a Quantitative Variable 48

Frequency Distribution 48

Relative Frequency and Percent Frequency Distributions 50

Using Excel to Construct a Frequency Distribution 50

Dot Plot 52

Histogram 53

Using Excel's Recommended Charts Tool to Construct a Histogram 55

Cumulative Distributions 57

Stem-and-Leaf Display 58

2.3 Summarizing Data for Two Variables Using Tables 66

Crosstabulation 67

Using Excel's PivotTable Tool to Construct a Crosstabulation 69

Simpson's Paradox 71

2.4 Summarizing Data for Two Variables Using Graphical Displays 78

Scatter Diagram and Trendline 78

Using Excel to Construct a Scatter Diagram and a Trendline 80

Side-by-Side and Stacked Bar Charts 81

Using Excel's Recommended Charts Tool to Construct Side-by-Side and Stacked Bar Charts 83

2.5 Data Visualization: Best Practices in Creating Effective Graphical Displays 88

Creating Effective Graphical Displays 88

Choosing the Type of Graphical Display 89

Data Dashboards 90

Data Visualization in Practice: Cincinnati Zoo and Botanical Garden 92

Summary 94

Glossary 95

Key Formulas 96

Supplementary Exercises 97

Case Problem 1 Pelican Stores 102

Case Problem 2 Motion Picture Industry 103

Case Problem 3 Queen City 104

Case Problem 4 Cut-Rate Machining, Inc. 104

Chapter 3 Descriptive Statistics: Numerical Measures 108

Statistics in Practice: Small Fry Design 109

3.1 Measures of Location 110

Mean 110

Median 112

Mode 113

Using Excel to Compute the Mean, Median, and Mode 114

Weighted Mean 115

Geometric Mean 116

Using Excel to Compute the Geometric Mean 118

Percentiles 119

Quartiles 120

Using Excel to Compute Percentiles and Quartiles 121

3.2 Measures of Variability 127

Range 128

Interquartile Range 128

Variance 128

Standard Deviation 130

Using Excel to Compute the Sample Variance and Sample Standard Deviation 131

Coefficient of Variation 132

Using Excel's Descriptive Statistics Tool 132

3.3 Measures of Distribution Shape, Relative Location, and Detecting Outliers 137

Distribution Shape 137

z-Scores 137

Chebyshev's Theorem 140

Empirical Rule 140

Detecting Outliers 141

3.4 Five-Number Summaries and Box Plots 144

Five-Number Summary 145

Box Plot 145

Using Excel to Construct a Box Plot 146

Comparative Analysis Using Box Plots 147

Using Excel to Construct a Comparative Analysis Using Box Plots 147

3.5 Measures of Association Between Two Variables 151

Covariance 152

Interpretation of the Covariance 153

Correlation Coefficient 156

Interpretation of the Correlation Coefficient 157

	Using Excel to Compute the Sample Covariance and Sample Correlation Coefficient	158
3.6	Data Dashboards: Adding Numerical Measures to Improve Effectiveness	161
	Summary	164
	Glossary	165
	Key Formulas	166
	Supplementary Exercises	168
	Case Problem 1 Pelican Stores	173
	Case Problem 2 Motion Picture Industry	174
	Case Problem 3 Business Schools of Asia-Pacific	175
	Case Problem 4 Heavenly Chocolates Website Transactions	177
	Case Problem 5 African Elephant Populations	178
Chapter 4	Introduction to Probability	180
	Statistics in Practice: National Aeronautics and Space Administration	181
4.1	Experiments, Counting Rules, and Assigning Probabilities	182
	Counting Rules, Combinations, and Permutations	183
	Assigning Probabilities	187
	Probabilities for the KP&L Project	189
4.2	Events and Their Probabilities	192
4.3	Some Basic Relationships of Probability	196
	Complement of an Event	196
	Addition Law	197
4.4	Conditional Probability	203
	Independent Events	206
	Multiplication Law	206
4.5	Bayes' Theorem	211
	Tabular Approach	214
	Summary	216
	Glossary	217
	Key Formulas	218
	Supplementary Exercises	219
	Case Problem 1 Hamilton County Judges	224
	Case Problem 2 Rob's Market	226
Chapter 5	Discrete Probability Distributions	228
	Statistics in Practice: Citibank	229
5.1	Random Variables	230
	Discrete Random Variables	230
	Continuous Random Variables	230
5.2	Developing Discrete Probability Distributions	233

5.3 Expected Value and Variance 238

Expected Value 238

Variance 238

Using Excel to Compute the Expected Value, Variance,
and Standard Deviation 240**5.4 Bivariate Distributions, Covariance, and Financial Portfolios 244**

A Bivariate Empirical Discrete Probability Distribution 244

Financial Applications 247

Summary 251

5.5 Binomial Probability Distribution 254

A Binomial Experiment 254

Martin Clothing Store Problem 256

Using Excel to Compute Binomial Probabilities 260

Expected Value and Variance for the Binomial Distribution 262

5.6 Poisson Probability Distribution 265

An Example Involving Time Intervals 266

An Example Involving Length or Distance Intervals 267

Using Excel to Compute Poisson Probabilities 267

5.7 Hypergeometric Probability Distribution 270

Using Excel to Compute Hypergeometric Probabilities 272

Summary 274**Glossary 275****Key Formulas 276****Supplementary Exercises 277****Case Problem 1 *Go Bananas!* 281****Case Problem 2 *McNeil's Auto Mall* 282****Case Problem 3 *Grievance Committee at Tuglar Corporation* 283****Case Problem 4 *Sagittarius Casino* 283****Chapter 6 Continuous Probability Distributions 285****Statistics in Practice: Procter & Gamble 286****6.1 Uniform Probability Distribution 287**

Area as a Measure of Probability 288

6.2 Normal Probability Distribution 291

Normal Curve 291

Standard Normal Probability Distribution 293

Computing Probabilities for Any Normal Probability Distribution 298

Great Tire Company Problem 299

Using Excel to Compute Normal Probabilities 300

6.3 Exponential Probability Distribution 306

Computing Probabilities for the Exponential Distribution 307

Relationship Between the Poisson and Exponential Distributions 308

Using Excel to Compute Exponential Probabilities 308

Summary	311
Glossary	311
Key Formulas	312
Supplementary Exercises	312
Case Problem 1 Specialty Toys	315
Case Problem 2 Gebhardt Electronics	316
Chapter 7 Sampling and Sampling Distributions	317
Statistics in Practice: Meadwestvaco Corporation	318
7.1 The Electronics Associates Sampling Problem	319
7.2 Selecting a Sample	320
Sampling from a Finite Population	320
Sampling from an Infinite Population	323
7.3 Point Estimation	327
Practical Advice	329
7.4 Introduction to Sampling Distributions	331
7.5 Sampling Distribution of \bar{x}	334
Expected Value of \bar{x}	334
Standard Deviation of \bar{x}	335
Form of the Sampling Distribution of \bar{x}	336
Sampling Distribution of \bar{x} for the EAI Problem	338
Practical Value of the Sampling Distribution of \bar{x}	338
Relationship Between the Sample Size and the Sampling Distribution of \bar{x}	340
7.6 Sampling Distribution of \bar{p}	344
Expected Value of \bar{p}	344
Standard Deviation of \bar{p}	345
Form of the Sampling Distribution of \bar{p}	346
Practical Value of the Sampling Distribution of \bar{p}	347
7.7 Other Sampling Methods	350
Stratified Random Sampling	350
Cluster Sampling	350
Systematic Sampling	351
Convenience Sampling	351
Judgment Sampling	352
7.8 Practical Advice: Big Data and Errors in Sampling	352
Sampling Error	353
Nonsampling Error	354
Implications of Big Data	355
Summary	357
Glossary	358
Key Formulas	359
Supplementary Exercises	359
Case Problem 1 Marion Dairies	362

Chapter 8 Interval Estimation 363**Statistics in Practice: Food Lion 364****8.1 Population Mean: σ Known 365**

Margin of Error and the Interval Estimate 365

Using Excel 369

Practical Advice 370

8.2 Population Mean: σ Unknown 372

Margin of Error and the Interval Estimate 374

Using Excel 376

Practical Advice 378

Using a Small Sample 378

Summary of Interval Estimation Procedures 379

8.3 Determining the Sample Size 383**8.4 Population Proportion 386**

Using Excel 387

Determining the Sample Size 389

8.5 Practical Advice: Big Data and Interval Estimation 393

Big Data and the Precision of Confidence Intervals 393

Implications of Big Data 394

Summary 396**Glossary 396****Key Formulas 397****Supplementary Exercises 398****Case Problem 1 *Young Professional Magazine* 400****Case Problem 2 *Gulf Real Estate Properties* 402****Case Problem 3 *Metropolitan Research, Inc.* 403****Chapter 9 Hypothesis Tests 405****Statistics in Practice: John Morrell & Company 406****9.1 Developing Null and Alternative Hypotheses 407**

The Alternative Hypothesis as a Research Hypothesis 407

The Null Hypothesis as an Assumption to Be Challenged 408

Summary of Forms for Null and Alternative Hypotheses 409

9.2 Type I and Type II Errors 411**9.3 Population Mean: σ Known 413**

One-Tailed Test 413

Two-Tailed Test 419

Using Excel 422

Summary and Practical Advice 424

Relationship Between Interval Estimation and Hypothesis Testing 425

9.4 Population Mean: σ Unknown 429

One-Tailed Test 430

Two-Tailed Test 431

	Using Excel	432
	Summary and Practical Advice	434
9.5	Population Proportion	438
	Using Excel	440
	Summary	441
9.6	Practical Advice: Big Data and Hypothesis Testing	444
	Big Data and p -Values	444
	Implications of Big Data	445
	Summary	447
	Glossary	447
	Key Formulas	448
	Supplementary Exercises	448
	Case Problem 1 Quality Associates, Inc.	451
	Case Problem 2 Ethical Behavior of Business Students at Bayview University	452
Chapter 10	Inference About Means and Proportions with Two Populations	455
	Statistics in Practice: U.S. Food and Drug Administration	456
10.1	Inferences About the Difference Between Two Population Means: σ_1 and σ_2 Known	457
	Interval Estimation of $\mu_1 - \mu_2$	457
	Using Excel to Construct a Confidence Interval	459
	Hypothesis Tests About $\mu_1 - \mu_2$	461
	Using Excel to Conduct a Hypothesis Test	463
	Practical Advice	464
10.2	Inferences About the Difference Between Two Population Means: σ_1 and σ_2 Unknown	467
	Interval Estimation of $\mu_1 - \mu_2$	467
	Using Excel to Construct a Confidence Interval	469
	Hypothesis Tests About $\mu_1 - \mu_2$	471
	Using Excel to Conduct a Hypothesis Test	473
	Practical Advice	475
10.3	Inferences About the Difference Between Two Population Means: Matched Samples	478
	Using Excel to Conduct a Hypothesis Test	481
10.4	Inferences About the Difference Between Two Population Proportions	486
	Interval Estimation of $p_1 - p_2$	486
	Using Excel to Construct a Confidence Interval	488
	Hypothesis Tests About $p_1 - p_2$	489
	Using Excel to Conduct a Hypothesis Test	491
	Summary	495
	Glossary	496

Key Formulas 496

Supplementary Exercises 497

Case Problem Par, Inc. 501

Chapter 11 Inferences About Population Variances 502

Statistics in Practice: U.S. Government Accountability Office 503

11.1 Inferences About a Population Variance 504

Interval Estimation 504

Using Excel to Construct a Confidence Interval 508

Hypothesis Testing 509

Using Excel to Conduct a Hypothesis Test 512

11.2 Inferences About Two Population Variances 516

Using Excel to Conduct a Hypothesis Test 520

Summary 525

Key Formulas 525

Supplementary Exercises 525

Case Problem 1 Air Force Training Program 527

Case Problem 2 Meticulous Drill & Reamer 528

Chapter 12 Tests of Goodness of Fit, Independence, and Multiple Proportions 530

Statistics in Practice: United Way 531

12.1 Goodness of Fit Test 532

Multinomial Probability Distribution 532

Using Excel to Conduct a Goodness of Fit Test 536

12.2 Test of Independence 538

Using Excel to Conduct a Test of Independence 542

12.3 Testing for Equality of Three or More Population Proportions 547

A Multiple Comparison Procedure 551

Using Excel to Conduct a Test of Multiple Proportions 553

Summary 557

Glossary 558

Key Formulas 558

Supplementary Exercises 558

Case Problem 1 A Bipartisan Agenda for Change 561

Case Problem 2 Fuentes Salty Snacks, Inc. 562

Case Problem 3 Fresno Board Games 563

Chapter 13 Experimental Design and Analysis of Variance 564

Statistics in Practice: Burke Marketing Services, Inc. 565

13.1 An Introduction to Experimental Design and Analysis of Variance 567

	Data Collection	567
	Assumptions for Analysis of Variance	569
	Analysis of Variance: A Conceptual Overview	569
13.2	Analysis of Variance and the Completely Randomized Design	572
	Between-Treatments Estimate of Population Variance	573
	Within-Treatments Estimate of Population Variance	574
	Comparing the Variance Estimates: The F Test	574
	ANOVA Table	576
	Using Excel	578
	Testing for the Equality of k Population Means: An Observational Study	579
13.3	Multiple Comparison Procedures	584
	Fisher's LSD	584
	Type I Error Rates	587
13.4	Randomized Block Design	590
	Air Traffic Controller Stress Test	591
	ANOVA Procedure	592
	Computations and Conclusions	593
	Using Excel	595
13.5	Factorial Experiment	599
	ANOVA Procedure	600
	Computations and Conclusions	601
	Using Excel	604
	Summary	608
	Glossary	609
	Key Formulas	610
	Supplementary Exercises	612
	Case Problem 1 Wentworth Medical Center	616
	Case Problem 2 Compensation for Sales Professionals	617
	Case Problem 3 TourisTopia Travel	618
	Chapter 14 Simple Linear Regression	620
	Statistics in Practice: Alliance Data Systems	621
14.1	Simple Linear Regression Model	622
	Regression Model and Regression Equation	622
	Estimated Regression Equation	623
14.2	Least Squares Method	625
	Using Excel to Construct a Scatter Diagram, Display the Estimated Regression Line, and Display the Estimated Regression Equation	629
14.3	Coefficient of Determination	637
	Using Excel to Compute the Coefficient of Determination	641
	Correlation Coefficient	641
14.4	Model Assumptions	645

14.5	Testing for Significance	647
	Estimate of σ^2	647
	t Test	648
	Confidence Interval for β_1	650
	F Test	651
	Some Cautions About the Interpretation of Significance Tests	653
14.6	Using the Estimated Regression Equation for Estimation and Prediction	656
	Interval Estimation	657
	Confidence Interval for the Mean Value of y	658
	Prediction Interval for an Individual Value of y	659
14.7	Excel's Regression Tool	664
	Using Excel's Regression Tool for the Armand's Pizza Parlors Example	664
	Interpretation of Estimated Regression Equation Output	665
	Interpretation of ANOVA Output	666
	Interpretation of Regression Statistics Output	667
14.8	Residual Analysis: Validating Model Assumptions	670
	Residual Plot Against x	671
	Residual Plot Against \hat{y}	672
	Standardized Residuals	674
	Using Excel to Construct a Residual Plot	676
	Normal Probability Plot	679
14.9	Outliers and Influential Observations	682
	Detecting Outliers	682
	Detecting Influential Observations	684
14.10	Practical Advice: Big Data and Hypothesis Testing in Simple Linear Regression	689
	Summary	690
	Glossary	690
	Key Formulas	691
	Supplementary Exercises	694
	Case Problem 1 Measuring Stock Market Risk	698
	Case Problem 2 U.S. Department of Transportation	699
	Case Problem 3 Selecting a Point-and-Shoot Digital Camera	700
	Case Problem 4 Finding the Best Car Value	701
	Case Problem 5 Buckeye Creek Amusement Park	702
	Appendix 14.1 Calculus-Based Derivation of Least Squares Formulas	703
	Appendix 14.2 A Test for Significance Using Correlation	704
	Chapter 15 Multiple Regression	706
	Statistics in Practice: International Paper	707
	15.1 Multiple Regression Model	708

	Regression Model and Regression Equation	708
	Estimated Multiple Regression Equation	708
15.2	Least Squares Method	709
	An Example: Butler Trucking Company	710
	Using Excel's Regression Tool to Develop the Estimated Multiple Regression Equation	713
	Note on Interpretation of Coefficients	714
15.3	Multiple Coefficient of Determination	720
15.4	Model Assumptions	723
15.5	Testing for Significance	724
	<i>F</i> Test	725
	<i>t</i> Test	727
	Multicollinearity	728
15.6	Using the Estimated Regression Equation for Estimation and Prediction	731
15.7	Categorical Independent Variables	734
	An Example: Johnson Filtration, Inc.	734
	Interpreting the Parameters	737
	More Complex Categorical Variables	738
15.8	Residual Analysis	742
	Residual Plot Against \hat{y}	743
	Standardized Residual Plot Against \hat{y}	744
15.9	Practical Advice: Big Data and Hypothesis Testing in Multiple Regression	747
	Summary	748
	Glossary	748
	Key Formulas	749
	Supplementary Exercises	750
	Case Problem 1 Consumer Research, Inc.	755
	Case Problem 2 Predicting Winnings for NASCAR Drivers	756
	Case Problem 3 Finding the Best Car Value	758
	Appendix A References and Bibliography	762
	Appendix B Tables	764
	Appendix C Summation Notation	775
	Appendix D Self-Test Solutions and Answers to Even-Numbered Exercises (online)	777
	Appendix E Microsoft Excel 2016 and Tools for Statistical Analysis	778
	Index	786

Preface

This text is the seventh edition of *Essentials of Modern Business Statistics with Microsoft® Office Excel®*. With this edition we welcome two eminent scholars to our author team: Jeffrey D. Camm of Wake Forest University and James J. Cochran of the University of Alabama. 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 that follows this preface. We believe that the addition of Jeff and Jim as our coauthors will both maintain and improve the effectiveness of *Essentials of Modern Business Statistics with Microsoft Office Excel*.

The purpose of *Essentials of Modern Business Statistics with Microsoft® Office Excel®* 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 applications setting, with the statistical results providing insights for decision making and solutions to applied 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.

Use of Microsoft Excel for Statistical Analysis

Essentials of Modern Business Statistics with Microsoft® Office Excel® is first and foremost a statistics textbook that emphasizes statistical concepts and applications. But since most practical problems are too large to be solved using hand calculations, some type of statistical software package is required to solve these problems. There are several excellent statistical packages available today; however, because most students and potential employers value spreadsheet experience, many schools now use a spreadsheet package in their statistics courses. Microsoft Excel is the most widely used spreadsheet package in business as well as in colleges and universities. We have written *Essentials of Modern Business Statistics with Microsoft® Office Excel®* especially for statistics courses in which Microsoft Excel is used as the software package.

Excel has been integrated within each of the chapters and plays an integral part in providing an application orientation. Although we assume that readers using this text are familiar with Excel basics such as selecting cells, entering formulas, copying, and so on, we do not assume that readers are familiar with Excel 2016 or Excel's tools for statistical analysis. As a result, we have included Appendix E, which provides an introduction to Excel 2016 and tools for statistical analysis.

Throughout the text the discussion of using Excel to perform a statistical procedure appears in a subsection immediately following the discussion of the statistical procedure. We believe that this style enables us to fully integrate the use of Excel throughout

the text, but still maintain the primary emphasis on the statistical methodology being discussed. In each of these subsections, we provide a standard format for using Excel for statistical analysis. There are four primary tasks: Enter/Access Data, Enter Functions and Formulas, Apply Tools, and Editing Options. The Editing Options task is new with this edition. It primarily involves how to edit Excel output so that it is more suitable for presentations to users. We believe a consistent framework for applying Excel helps users to focus on the statistical methodology without getting bogged down in the details of using Excel.

In presenting worksheet figures, we often use a nested approach in which the worksheet shown in the background of the figure displays the formulas and the worksheet shown in the foreground shows the values computed using the formulas. Different colors and shades of colors are used to differentiate worksheet cells containing data, highlight cells containing Excel functions and formulas, and highlight material printed by Excel as a result of using one or more data analysis tools.

Changes in the Seventh Edition

We appreciate the acceptance and positive response to the previous editions of *Essentials of Modern Business Statistics with Microsoft® Office Excel®*. Accordingly, in making modifications for this new edition, we have maintained the presentation style and readability of those editions. The significant changes in the new edition are summarized here.

Users of the previous edition will notice that the chapters offered and topics covered in this edition differ from previous editions. While the topical coverage of the first nine chapters remains the same, the organization and coverage in some of the later chapters have expanded. We have eliminated the coverage of the advanced topic of Time Series and Quality Control in favor of the expanded coverage in Chapters 10, 11, 12 and 13. Chapter 10 now provides coverage of inferences of means and proportions for two populations, and chapter 11 is focused on inferences about population variances. Chapter 12 is a discussion of comparing multiple proportions, tests of independence and goodness of fit and chapter 13 covers experimental design and ANOVA. We believe you will find the expanded coverage in these chapters useful in your classes. Coverage of regression is now in chapters 14 and 15. These two chapters are revisions of the regression chapters from the 6th edition. In additions to these changes, we made the following revisions:

- **Microsoft Excel 2016.** Step-by-step instructions and screen captures show how to use the latest version of Excel to implement statistical procedures.
- **Data and Statistics—Chapter 1.** We have expanded our section on data mining to include a discussion of big data. We have added a new section on analytics. We have also placed greater emphasis on the distinction between observed and experimental data.
- **Descriptive Statistics: Tabular and Graphical Displays—Chapter 2.** Microsoft Excel now has the capability of creating box plots and comparative box plots. We have added to this chapter instruction on how to use this very useful new feature.
- **Interval Estimation—Chapter 8.** We have added a new section on the implications of big data (large data sets) on the interpretation of confidence intervals and importantly, the difference between statistical and practical significance.
- **Hypothesis Tests—Chapter 9.** Similar to our addition to Chapter 8, we have added a new section on the implications of big data (large data sets) on the interpretation of hypothesis tests and the difference between statistical and practical significance.
- **Simple Linear Regression—Chapter 14.** Similar to our addition to Chapter 8, we have added a new section on the implications of big data (large data sets) on

the interpretation of hypothesis tests in simple linear regression and the difference between statistical and practical significance.

- **New Case Problems.** We have added thirteen new case problems to this edition. The new case problems appear in the chapters on descriptive statistics and regression analysis. The case problems in the text provide students with the opportunity to analyze somewhat larger data sets and prepare managerial reports based on the results of their analysis.
- **New Examples and Exercises Based on Real Data.** We have added approximately 126 new examples and exercises based on real data and recently referenced sources of statistical information. Using data obtained from various data collection organizations, websites, and other sources such as *The Wall Street Journal*, *USA Today*, *Fortune*, and *Barron's*, we have drawn upon actual studies to develop explanations and to create exercises that demonstrate many uses of statistics in business and economics. We believe 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.
- **Updated and Improved End-of-Chapter Solutions and Solutions Manual.** Partnering with accomplished instructor Dawn Bulriss at Maricopa Community Colleges, we took a deep audit of the solutions manual. Every question and solution was reviewed and reworked, as necessary. The solutions now contain additional detail: improved rounding instructions; expanded explanations with a student-focus; and alternative answers using Excel and a statistical calculator. We believe this thorough review will enhance both the instructor and student learning experience in this digital age.

Features and Pedagogy

Authors Anderson, Sweeney, Williams, Camm, and Cochran have continued many of the features that appeared in previous editions.

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 those exercises are provided in Appendix D in the Student Resources online. 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 students are a key feature of this text. These annotations 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

Approximately 220 data files are available on the website that accompanies this text. The data sets are available in Excel 2016 format. DATAfile 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.

MindTap

MindTap, featuring all new Excel Online integration powered by Microsoft, is a complete digital solution for the business statistics course. It has enhancements that take students from learning basic statistical concepts to actively engaging in critical thinking applications, while learning valuable software skills for their future careers.

MindTap is a customizable digital course solution that includes an interactive eBook, autograded, algorithmic exercises from the textbook, Adaptive Test Prep, as well as interactive visualizations. All of these materials offer students better access to understand the materials within the course. For more information on MindTap, please contact your Cengage representative.

For Students

Online resources are available to help the student work more efficiently. The resources can be accessed through www.cengagebrain.com.

For Instructors

Instructor resources are available to adopters on the Instructor Companion Site, which can be found and accessed at www.cengage.com, including:

- **Solutions Manual:** The Solutions Manual, prepared by the authors, includes solutions for all problems in the text. It is available online as well as print.
- **Solutions to Case Problems:** These are also prepared by the authors and contain solutions to all case problems presented in the text.
- **PowerPoint Presentation Slides:** The presentation slides contain a teaching outline that incorporates figures to complement instructor lectures.
- **Test Bank:** Cengage Learning Testing Powered by Cognero is a flexible, online system that allows you to:
 - author, edit, and manage test bank content from multiple Cengage Learning solutions,
 - create multiple test versions in an instant, and
 - deliver tests from your LMS, your classroom, or wherever you want. The Test Bank is also available in Microsoft Word.

Acknowledgments

A special thanks goes to 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 product manager, Aaron Arnsparger; our content developer, Anne Merrill; our content project manager, Colleen Farmer; our project manager at MPS Limited, Gaurav Prabhu; digital content designer, Brandon Foltz; and others at Cengage for their editorial counsel and support during the preparation of this text.

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:

James Bang, Virginia Military Institute
Robert J. Banis, University of Missouri–St. Louis
Timothy M. Bergquist, Northwest Christian College
Gary Black, University of Southern Indiana
William Bleuel, Pepperdine University
Derrick Boone, Wake Forest University
Lawrence J. Bos, Cornerstone University
Dawn Bulriss, Maricopa Community Colleges
Joseph Cavanaugh, Wright State University–Lake Campus
Sheng-Kai Chang, Wayne State University
Robert Christopherson, SUNY-Plattsburgh
Michael Clark, University of Baltimore
Robert D. Collins, Marquette University
Ivona Contardo, Stellenbosch University
Sean Eom, Southeast Missouri State University
Samo Ghosh, Albright College
Philip A. Gibbs, Washington & Lee University
Daniel L. Gilbert, Tennessee Wesleyan College
Michael Gorman, University of Dayton
Erick Hofacker, University of Wisconsin, River Falls
David Juriga, St. Louis Community College
William Kasperski, Madonna University
Kuldeep Kumar, Bond Business School
Tenpao Lee, Niagara University
Ying Liao, Meredith College
Daniel Light, Northwest State College
Ralph Maliszewski, Waynesburg University
Saverio Manago, Salem State University
Patricia A. Mullins, University of Wisconsin–Madison
Jack Muryn, Cardinal Stritch University
Anthony Narsing, Macon State College
Robert M. Nauss, University of Missouri–St. Louis
Elizabeth L. Rankin, Centenary College of Louisiana
Surekha Rao, Indiana University, Northwest
Jim Robison, Sonoma State University
Farhad Saboori, Albright College
Susan Sandblom, Scottsdale Community College
Ahmad Saranjam, Bridgewater State University
Jeff Sarbaum, University of North Carolina at Greensboro
Robert Scott, Monmouth University

Toni Somers, Wayne State University
 Jordan H. Stein, University of Arizona
 Bruce Thompson, Milwaukee School of Engineering
 Ahmad Vessal, California State University, Northridge
 Dave Vinson, Pellissippi State
 Daniel B. Widdis, Naval Postgraduate School
 Peter G. Wagner, University of Dayton
 Sheng-Ping Yang, Black Hills State University

We would like to recognize the following individuals, who have helped us in the past and continue to influence our writing.

Glen Archibald, University of Mississippi
 Darl Bien, University of Denver
 Thomas W. Bolland, Ohio University
 Mike Bourke, Houston Baptist University
 Peter Bryant, University of Colorado
 Terri L. Byczkowski, University of Cincinnati
 Robert Carver, Stonehill College
 Ying Chien, University of Scranton
 Robert Cochran, University of Wyoming
 Murray Côté, University of Florida
 David W. Cravens, Texas Christian University
 Eddine Dahel, Monterey Institute of International Studies
 Tom Dahlstrom, Eastern College
 Terry Dielman, Texas Christian University
 Joan Donohue, University of South Carolina
 Jianjun Du, University of Houston–Victoria
 Thomas J. Dudley, Pepperdine University
 Swarna Dutt, University of West Georgia
 Ronald Ehresman, Baldwin-Wallace College
 Mohammed A. El-Saidi, Ferris State University
 Robert Escudero, Pepperdine University
 Stacy Everly, Delaware County Community College
 Soheila Kahkashan Fardanesh, Towson University
 Nicholas Farnum, California State University–Fullerton
 Abe Feinberg, California State University, Northridge
 Michael Ford, Rochester Institute of Technology
 Phil Fry, Boise State University
 V. Daniel Guide, Duquesne University
 Paul Guy, California State University–Chico
 Charles Harrington, University of Southern Indiana
 Carl H. Hess, Marymount University
 Woodrow W. Hughes, Jr., Converse College
 Alan Humphrey, University of Rhode Island
 Ann Hussein, Philadelphia College of Textiles and Science
 Ben Isselhardt, Rochester Institute of Technology
 Jeffery Jarrett, University of Rhode Island
 Barry Kadets, Bryant College
 Homayoun Khamooshi, George Washington University
 Kenneth Klassen, California State University Northridge
 David Krueger, St. Cloud State University
 June Lapidus, Roosevelt University

Martin S. Levy, University of Cincinnati
Daniel M. Light, Northwest State College
Ka-sing Man, Georgetown University
Don Marx, University of Alaska, Anchorage
Tom McCullough, University of California–Berkeley
Timothy McDaniel, Buena Vista University
Mario Miranda, The Ohio State University
Barry J. Monk, Macon State College
Mitchell Muesham, Sam Houston State University
Richard O’Connell, Miami University of Ohio
Alan Olinsky, Bryant College
Lynne Pastor, Carnegie Mellon University
Von Roderick Plessner, Northwest State University
Robert D. Potter, University of Central Florida
Tom Pray, Rochester Institute of Technology
Harold Rahmlow, St. Joseph’s University
Derrick Reagle, Fordham University
Avuthu Rami Reddy, University of Wisconsin–Platteville
Tom Ryan, Case Western Reserve University
Ahmad Saranjam, Bridgewater State College
Bill Seaver, University of Tennessee
Alan Smith, Robert Morris College
William Struning, Seton Hall University
Ahmad Syamil, Arkansas State University
David Tufte, University of New Orleans
Jack Vaughn, University of Texas–El Paso
Elizabeth Wark, Springfield College
Ari Wijetunga, Morehead State University
Nancy A. Williams, Loyola College in Maryland
J. E. Willis, Louisiana State University
Larry Woodward, University of Mary Hardin–Baylor
Mustafa Yilmaz, Northeastern University

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

CHAPTER 1

Data and Statistics

CONTENTS

STATISTICS IN PRACTICE:

BLOOMBERG BUSINESSWEEK

1.1 APPLICATIONS IN BUSINESS AND ECONOMICS

Accounting

Finance

Marketing

Production

Economics

Information Systems

1.2 DATA

Elements, Variables, and

Observations

Scales of Measurement

Categorical and Quantitative Data

Cross-Sectional and Time

Series Data

1.3 DATA SOURCES

Existing Sources

Observational Study

Experiment

Time and Cost Issues

Data Acquisition Errors

1.4 DESCRIPTIVE STATISTICS

1.5 STATISTICAL INFERENCE

1.6 STATISTICAL ANALYSIS

USING MICROSOFT EXCEL

Data Sets and Excel Worksheets

Using Excel for Statistical

Analysis

1.7 ANALYTICS

1.8 BIG DATA AND DATA MINING

1.9 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. Examples of articles and reports include the impact of businesses moving important work to cloud computing, the crisis facing the U.S. Postal Service, and why the debt crisis is even worse than we think. In addition, *Bloomberg Businessweek* provides a variety of statistics about the state of the economy, including production indexes, stock prices, mutual funds, and interest rates.

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



AP Photos/Weng lei - Imaginecthina

Bloomberg Businessweek uses statistical facts and summaries in many of its articles.

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.

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

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

- Uber Technologies Inc. is turning to the leveraged-loan market for the first time to raise as much as \$2 billion, a sign of the popular ride-sharing network's hunger for cash as it expands around the world (*The Wall Street Journal*, June 14, 2016).
- Against the U.S. dollar, the euro has lost nearly 30% of its value in the last year; the Australian dollar lost almost 20% (*The Economist*, April 25th–May 1st, 2015).

- VW Group's U.S. sales continue to slide, with total sales off by 13% from last January, to 36,930 vehicles (*Panorama*, March 2014).
- A poll of 1320 corporate recruiters indicated that 68% of the recruiters ranked communication skills as one of the top five most important skills for new hires (*Bloomberg Businessweek* April 13–April 19, 2015).
- Green Mountain sold 18 billion coffee pods in two years (*Harvard Business Review*, January-February, 2016).
- Most homeowners spend between about \$10,000 and roughly \$27,000 converting a basement, depending on the size of the space, according to estimates from HomeAdvisor, a website that connects homeowners with prescreened service professionals (*Consumer Reports*, February 9, 2016).
- A full 88% of consumers say they buy private label, primarily because of price, according to Market Track (*USA Today*, May 17, 2016).

The numerical facts in the preceding statements—\$2 billion, 30%, 20%, 13%, 36,930, 1320, 68%, 18 billion, \$10,000, \$27,000 and 88%—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 four sections of Chapter 1 provide the role of the computer in statistical analysis, an introduction to business analytics and the role statistics plays in it, an introduction to big data and data mining, and a discussion of ethical guidelines for statistical practice.

1.1

Applications in Business and Economics

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

Accounting

Public accounting firms use statistical sampling procedures when conducting audits for their clients. For instance, suppose an accounting firm wants to determine whether the amount of accounts receivable shown on a client's balance sheet fairly represents the actual amount of accounts receivable. Usually the large number of individual accounts receivable makes reviewing and validating every account too time-consuming and expensive. As common

practice in such situations, the audit staff selects a subset of the accounts called a sample. After reviewing the accuracy of the sampled accounts, the auditors draw a conclusion as to whether the accounts receivable amount shown on the client's balance sheet is acceptable.

Finance

Financial analysts use a variety of statistical information to guide their investment recommendations. In the case of stocks, analysts review financial data such as price/earnings ratios and dividend yields. By comparing the information for an individual stock with information about the stock market averages, an analyst can begin to draw a conclusion as to whether the stock is a good investment. For example, *The Wall Street Journal* (February 27, 2016) reported that the average dividend yield for the S&P 500 companies was 2.3%. Microsoft showed a dividend yield of 2.61%. 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 \bar{x} -bar chart can be used to monitor the average output. Suppose, for example, that a machine fills containers with 12 ounces of a soft drink. Periodically, a production worker selects a sample of containers and computes the average number of ounces in the sample. This average, or \bar{x} -bar value, is plotted on an \bar{x} -bar chart. A plotted value above the chart's upper control limit indicates overfilling, and a plotted value below the chart's lower control limit indicates underfilling. The process is termed "in control" and allowed to continue as long as the plotted \bar{x} -bar values fall between the chart's upper and lower control limits. Properly interpreted, an \bar{x} -bar chart can help determine when adjustments are necessary to correct a production process.

Economics

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

Information Systems

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

Applications of statistics such as those described in this section are an integral part of this text. Such examples provide an overview of the breadth of statistical applications. To supplement these examples, practitioners in the fields of business and economics provided chapter-opening Statistics in Practice articles that introduce the material covered in each chapter. The Statistics in Practice applications show the importance of statistics in a wide variety of business and economic situations.

1.2

Data

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

Elements, Variables, and Observations

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

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

- **WTO Status:** The nation's membership status in the World Trade Organization; this can be either as a member or an observer.
- **Per Capita GDP (\$):** The total market value (\$) of all goods and services produced by the nation divided by the number of people in the nation; this is commonly used to compare economic productivity of the nations.
- **Trade Deficit (\$1000s):** The difference between the total dollar value of the nation's imports and the 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 first observation contains the following measurements: Member,

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 
Nations

Data sets such as Nations are available on the companion site for this title.

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

Slovenia	Member	29,100	2,310,617	AA-	Negative
South Africa	Member	11,000	3,321,801	BBB+	Stable
Sweden	Member	40,600	-10,903,251	AAA	Stable
Switzerland	Member	43,400	-27,197,873	AAA	Stable
Thailand	Member	9,700	2,049,669	BBB	Stable
Turkey	Member	14,600	71,612,947	BB+	Positive
UK	Member	35,900	162,316,831	AAA	Negative
Uruguay	Member	15,400	2,662,628	BB	Positive
USA	Member	48,100	784,438,559	AAA	Stable
Zambia	Member	1,600	-1,805,198	B+	Stable

5400, 2,673,359, BB-, and Stable. The second observation contains the following measurements: Member, 40,800, -33,304,157, AAA, Stable, and so on. A data set with 60 elements contains 60 observations.

Scales of Measurement

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

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

The scale of measurement for a variable is considered an **ordinal scale** if the data exhibit the properties of nominal data and in addition, the order or rank of the data is meaningful. For example, referring to the data in Table 1.1, the scale of measurement for 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.

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

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

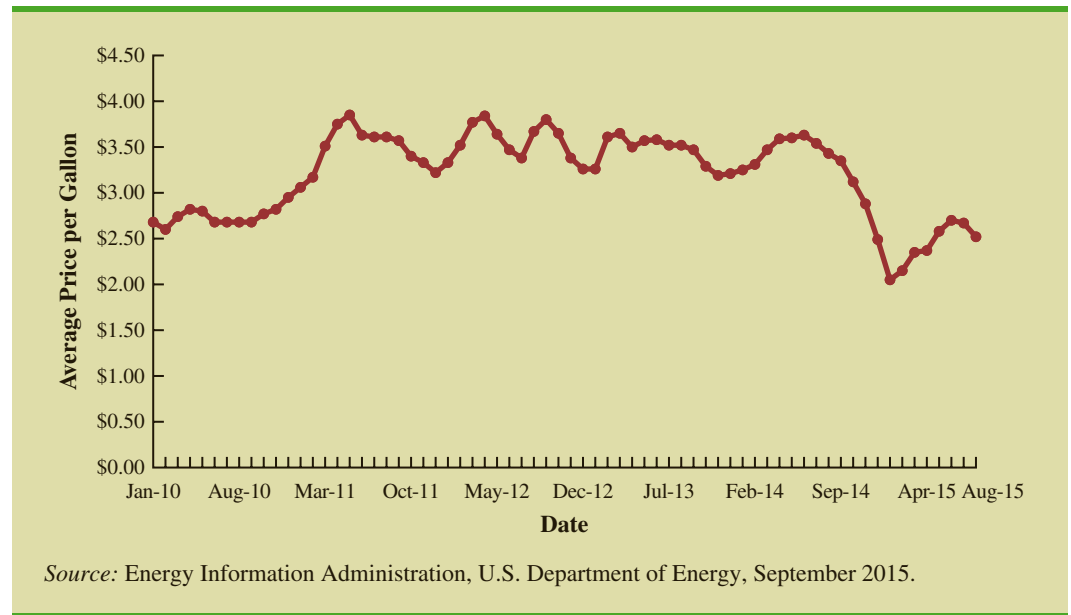
Arithmetic operations provide meaningful results for quantitative variables. For example, quantitative data may be added and then divided by the number of observations to compute the average value. This average is usually meaningful and easily interpreted. In general, more alternatives for statistical 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 2010 and 2015. Note that gasoline prices peaked in May 2011. Between June 2014 and January 2015, the average price per gallon dropped dramatically. In August 2015, the average price per gallon was \$2.52.

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 2005 to 2015. In September 2005, the popular stock market index was near 10,400. Over the next two years the index rose to almost 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 and rising to a high of over 18,000 in May 2015. By October 2015, the index had dropped substantially to just under 16,300.

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 2007 to 2015. 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, remained at about 5.5 billion from 2011 to 2013, decreased substantially in 2014, and dropped again in 2015. Analysts suspect that the drop in net income was due to loss of customers to newer competition such as Chipotle.

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

NOTES AND COMMENTS

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