

Camm Cochran Fry Ohlmann Anderson Sweeney Williams

Business Analytics

DESCRIPTIVE • PREDICTIVE • PRESCRIPTIVE

Business Analytics

Third Edition



Jeffrey D. Camm
Wake Forest University

James J. Cochran
University of Alabama

Michael J. Fry
University of Cincinnati

Jeffrey W. Ohlmann
University of Iowa

David R. Anderson
University of Cincinnati

Dennis J. Sweeney
University of Cincinnati

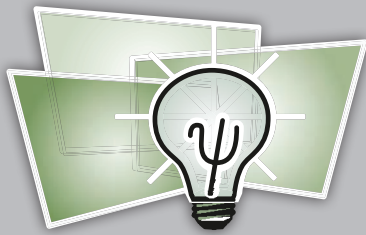
Thomas A. Williams
Rochester Institute
of Technology



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

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Copyright 2019 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.



FRONTLINE solvers

Analytic Solver®

www.solver.com/aspe

Your new textbook, *Business Analytics, 3e*, uses this software in the eBook (MindTap Reader). Here's how to get it for your course.

For Instructors: Setting Up the Course Code

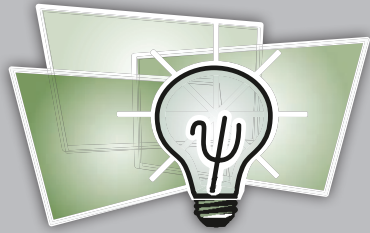
To set up a course code for your course, please email Frontline Systems at **academic@solver.com**, or call **775-831-0300**, press 0, and ask for the Academic Coordinator. Course codes MUST be set up each time a course is taught.

A course code is free, and it can usually be issued within 24 to 48 hours (often the same day). It will enable your students to download and install Analytic Solver, use the software free for 2 weeks, and continue to use the software for 20 weeks (a typical semester) for a small charge, using the course code to obtain a steep discount. It will enable Frontline Systems to assist students with installation, and provide technical support to you during the course.

Please give the course code, plus the instructions on the reverse side, to your students. If you're evaluating the book for adoption, you can use the course code yourself to download and install the software as described on the reverse.

**Instructions for Students: See reverse.
Installing Analytic Solver**





FRONTLINE solvers

For Students:

Installing Analytic Solver®

- 1) To download and install Analytic Solver (also called Analytic Solver® Basic) from Frontline Systems to work with Microsoft® Excel® for Windows®, please visit:

www.solver.com/student

- 2) **Fill out the registration form on this page**, supplying your name, school, email address (key information will be sent to this address), course code (obtain this from your instructor), and textbook code (enter **CCFOEBA3**).
- 3) On the download page, **click the Download Now** button, and save the downloaded file (**SolverSetup.exe**).
- 4) **Close any Excel® windows** you have open. Make sure you're still connected to the Internet.
- 5) **Run SolverSetup.exe to install the software.** Be patient – all necessary files will be downloaded and installed.

If you have problems downloading or installing, your best options in order are to (i) visit www.solver.com and click the **Live Chat** link at the top; (ii) email support@solver.com and watch for a reply; (iii) call **775-831-0300** and press 4 (tech support). Say that you have Analytic Solver for Education, and have your course code and textbook code available.

If you have problems setting up or solving your model, or interpreting the results, please ask your instructor for assistance. Frontline Systems cannot help you with homework problems.

*If you have this textbook but you aren't enrolled in a course, visit www.solver.com or call **775-831-0300** and press 0 for advice on courses and software licenses.*

*If you have a Mac, your options are to (i) use a version of Analytic Solver through a web browser at <https://analyticsolver.com>, or (ii) install "dual-boot" or VM software, Microsoft Windows®, and Excel® for Windows®. **Excel® for Mac will NOT work.***

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.

Important Notice: Media content referenced within the product description or the product text may not be available in the eBook version.

Business Analytics, Third Edition
Jeffrey D. Camm, James J. Cochran,
Michael J. Fry, Jeffrey W. Ohlmann,
David R. Anderson, Dennis J. Sweeney,
Thomas A. Williams

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

Vice President, B&E, 4LTR, and Support
Programs: Mike Schenk

Senior Product Team Manager: Joe Sabatino

Senior Product Manager: Aaron Arnsperger

Senior Digital Content Designer: Brandon Foltz

Content Developer: Anne Merrill

Content Project Manager: D. Jean Buttrom

Product Assistant: Renee Schnee

Senior Marketing Manager: Nathan Anderson

Production Service: SPi Global

Senior Art Director: Michelle Kunkler

Cover and Text Designer: Beckmeyer Design

Cover Image: iStockPhoto.com/tawanlubfah

Intellectual Property Analyst: Reba Frederics

Intellectual Property Project Manager:
Nick Barrows

© 2019, 2017 Cengage Learning, Inc.

Unless otherwise noted, all content is © Cengage.

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

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

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

Screenshots are ©Microsoft Corporation unless otherwise noted.

Library of Congress Control Number: 2017963667

ISBN: 978-1-337-40642-0

Cengage
20 Channel Center Street
Boston, MA 02210
USA

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

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

To learn more about Cengage platforms and services,
visit **www.cengage.com.**

To register or access your online learning solution or purchase materials
for your course, visit **www.cengagebrain.com.**

Printed in the United States of America
Print Number: 01 Print Year: 2018

Brief Contents

ABOUT THE AUTHORS XIX
PREFACE XXIII

CHAPTER 1	Introduction 2
CHAPTER 2	Descriptive Statistics 18
CHAPTER 3	Data Visualization 82
CHAPTER 4	Descriptive Data Mining 138
CHAPTER 5	Probability: An Introduction to Modeling Uncertainty 166
CHAPTER 6	Statistical Inference 220
CHAPTER 7	Linear Regression 294
CHAPTER 8	Time Series Analysis and Forecasting 372
CHAPTER 9	Predictive Data Mining 422
CHAPTER 10	Spreadsheet Models 464
CHAPTER 11	Monte Carlo Simulation 500
CHAPTER 12	Linear Optimization Models 556
CHAPTER 13	Integer Linear Optimization Models 606
CHAPTER 14	Nonlinear Optimization Models 646
CHAPTER 15	Decision Analysis 678
APPENDIX A	Basics of Excel 724
APPENDIX B	Database Basics with Microsoft Access 736
APPENDIX C	Solutions to Even-Numbered Questions (MindTap Reader)
REFERENCES	774
INDEX	776



Fit your coursework into your hectic life.

Make the most of your time by learning your way. Access the resources you need to succeed wherever, whenever.



Study with digital flashcards, listen to audio textbooks, and take quizzes.



Review your current course grade and compare your progress with your peers.



Get the free MindTap Mobile App and learn wherever you are.

Break Limitations. Create your own potential, and be unstoppable with MindTap.

MINDTAP. POWERED BY YOU.

cengage.com/mindtap

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Copyright 2019 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.



Contents

ABOUT THE AUTHORS XIX

PREFACE XXIII

CHAPTER 1 Introduction 2

1.1 Decision Making 4

1.2 Business Analytics Defined 5

1.3 A Categorization of Analytical Methods and Models 6

 Descriptive Analytics 6

 Predictive Analytics 6

 Prescriptive Analytics 7

1.4 Big Data 7

 Volume 9

 Velocity 9

 Variety 9

 Veracity 9

1.5 Business Analytics in Practice 11

 Financial Analytics 11

 Human Resource (HR) Analytics 12

 Marketing Analytics 12

 Health Care Analytics 12

 Supply-Chain Analytics 13

 Analytics for Government and Nonprofits 13

 Sports Analytics 13

 Web Analytics 14

Summary 14

Glossary 15

CHAPTER 2 Descriptive Statistics 18

Analytics in Action: U.S. Census Bureau 19

2.1 Overview of Using Data: Definitions and Goals 19

2.2 Types of Data 21

 Population and Sample Data 21

 Quantitative and Categorical Data 21

 Cross-Sectional and Time Series Data 21

 Sources of Data 21

2.3 Modifying Data in Excel 24

 Sorting and Filtering Data in Excel 24

 Conditional Formatting of Data in Excel 27

2.4 Creating Distributions from Data 29

 Frequency Distributions for Categorical Data 29

 Relative Frequency and Percent Frequency Distributions 30

 Frequency Distributions for Quantitative Data 31

 Histograms 34

 Cumulative Distributions 37

2.5	Measures of Location	39
	Mean (Arithmetic Mean)	39
	Median	40
	Mode	41
	Geometric Mean	41
2.6	Measures of Variability	44
	Range	44
	Variance	45
	Standard Deviation	46
	Coefficient of Variation	47
2.7	Analyzing Distributions	47
	Percentiles	48
	Quartiles	49
	z-Scores	49
	Empirical Rule	50
	Identifying Outliers	52
	Box Plots	52
2.8	Measures of Association Between Two Variables	55
	Scatter Charts	55
	Covariance	57
	Correlation Coefficient	60
2.9	Data Cleansing	61
	Missing Data	61
	Blakely Tires	63
	Identification of Erroneous Outliers and Other Erroneous Values	65
	Variable Representation	67
	Summary	68
	Glossary	69
	Problems	71
	Case Problem: Heavenly Chocolates Web Site Transactions	79
	Appendix 2.1 Creating Box Plots with Analytic Solver (MindTap Reader)	
CHAPTER 3	Data Visualization	82
	Analytics in Action: Cincinnati Zoo & Botanical Garden	83
3.1	Overview of Data Visualization	85
	Effective Design Techniques	85
3.2	Tables	88
	Table Design Principles	89
	Crosstabulation	90
	PivotTables in Excel	93
	Recommended PivotTables in Excel	97
3.3	Charts	99
	Scatter Charts	99
	Recommended Charts in Excel	101

Line Charts	102
Bar Charts and Column Charts	106
A Note on Pie Charts and Three-Dimensional Charts	107
Bubble Charts	109
Heat Maps	110
Additional Charts for Multiple Variables	112
PivotCharts in Excel	115
3.4 Advanced Data Visualization	117
Advanced Charts	117
Geographic Information Systems Charts	120
3.5 Data Dashboards	122
Principles of Effective Data Dashboards	123
Applications of Data Dashboards	123
Summary	125
Glossary	125
Problems	126
Case Problem: All-Time Movie Box-Office Data	136
Appendix 3.1 Creating a Scatter-Chart Matrix and a Parallel-Coordinates Plot with Analytic Solver (MindTap Reader)	

CHAPTER 4 Descriptive Data Mining 138

Analytics in Action: Advice from a Machine	139
4.1 Cluster Analysis	140
Measuring Similarity Between Observations	140
Hierarchical Clustering	143
k-Means Clustering	146
Hierarchical Clustering versus k-Means Clustering	147
4.2 Association Rules	148
Evaluating Association Rules	150
4.3 Text Mining	151
Voice of the Customer at Triad Airline	151
Preprocessing Text Data for Analysis	153
Movie Reviews	154
Summary	155
Glossary	155
Problems	156
Case Problem: Know Thy Customer	164

Available in the MindTap Reader:

Appendix 4.1 Hierarchical Clustering with Analytic Solver
Appendix 4.2 k-Means Clustering with Analytic Solver
Appendix 4.3 Association Rules with Analytic Solver
Appendix 4.4 Text Mining with Analytic Solver
Appendix 4.5 Opening and Saving Excel files in JMP Pro
Appendix 4.6 Hierarchical Clustering with JMP Pro

Appendix 4.7 *k*-Means Clustering with JMP Pro

Appendix 4.8 Association Rules with JMP Pro

Appendix 4.9 Text Mining with JMP Pro

CHAPTER 5 Probability: An Introduction to Modeling Uncertainty 166

Analytics in Action: National Aeronautics and Space Administration 167

5.1 Events and Probabilities 168

5.2 Some Basic Relationships of Probability 169

Complement of an Event 169

Addition Law 170

5.3 Conditional Probability 172

Independent Events 177

Multiplication Law 177

Bayes' Theorem 178

5.4 Random Variables 180

Discrete Random Variables 180

Continuous Random Variables 181

5.5 Discrete Probability Distributions 182

Custom Discrete Probability Distribution 182

Expected Value and Variance 184

Discrete Uniform Probability Distribution 187

Binomial Probability Distribution 188

Poisson Probability Distribution 191

5.6 Continuous Probability Distributions 194

Uniform Probability Distribution 194

Triangular Probability Distribution 196

Normal Probability Distribution 198

Exponential Probability Distribution 203

Summary 207

Glossary 207

Problems 209

Case Problem: Hamilton County Judges 218

CHAPTER 6 Statistical Inference 220

Analytics in Action: John Morrell & Company 221

6.1 Selecting a Sample 223

Sampling from a Finite Population 223

Sampling from an Infinite Population 224

6.2 Point Estimation 227

Practical Advice 229

6.3 Sampling Distributions 229

Sampling Distribution of \bar{x} 232

Sampling Distribution of \bar{p} 237

6.4	Interval Estimation	240
	Interval Estimation of the Population Mean	240
	Interval Estimation of the Population Proportion	247
6.5	Hypothesis Tests	250
	Developing Null and Alternative Hypotheses	250
	Type I and Type II Errors	253
	Hypothesis Test of the Population Mean	254
	Hypothesis Test of the Population Proportion	265
6.6	Big Data, Statistical Inference, and Practical Significance	268
	Sampling Error	268
	Nonsampling Error	269
	Big Data	270
	Understanding What Big Data Is	271
	Big Data and Sampling Error	272
	Big Data and the Precision of Confidence Intervals	273
	Implications of Big Data for Confidence Intervals	274
	Big Data, Hypothesis Testing, and p Values	275
	Implications of Big Data in Hypothesis Testing	277
	Summary	278
	Glossary	279
	Problems	281
	Case Problem 1: Young Professional Magazine	291
	Case Problem 2: Quality Associates, Inc	292
CHAPTER 7	Linear Regression	294
	Analytics in Action: Alliance Data Systems	295
7.1	Simple Linear Regression Model	296
	Regression Model	296
	Estimated Regression Equation	296
7.2	Least Squares Method	298
	Least Squares Estimates of the Regression Parameters	300
	Using Excel's Chart Tools to Compute the Estimated Regression Equation	302
7.3	Assessing the Fit of the Simple Linear Regression Model	304
	The Sums of Squares	304
	The Coefficient of Determination	306
	Using Excel's Chart Tools to Compute the Coefficient of Determination	307
7.4	The Multiple Regression Model	308
	Regression Model	308
	Estimated Multiple Regression Equation	308
	Least Squares Method and Multiple Regression	309
	Butler Trucking Company and Multiple Regression	310
	Using Excel's Regression Tool to Develop the Estimated Multiple Regression Equation	310

7.5	Inference and Regression	313
	Conditions Necessary for Valid Inference in the Least Squares Regression Model	314
	Testing Individual Regression Parameters	318
	Addressing Nonsignificant Independent Variables	321
	Multicollinearity	322
7.6	Categorical Independent Variables	325
	Butler Trucking Company and Rush Hour	325
	Interpreting the Parameters	327
	More Complex Categorical Variables	328
7.7	Modeling Nonlinear Relationships	330
	Quadratic Regression Models	331
	Piecewise Linear Regression Models	335
	Interaction Between Independent Variables	337
7.8	Model Fitting	342
	Variable Selection Procedures	342
	Overfitting	343
7.9	Big Data and Regression	344
	Inference and Very Large Samples	344
	Model Selection	348
7.10	Prediction with Regression	349
	Summary	351
	Glossary	352
	Problems	354
	Case Problem: Alumni Giving	369
	Appendix 7.1 Regression with Analytic Solver (MindTap Reader)	
CHAPTER 8	Time Series Analysis and Forecasting	372
	Analytics in Action: ACCO Brands	373
8.1	Time Series Patterns	375
	Horizontal Pattern	375
	Trend Pattern	377
	Seasonal Pattern	378
	Trend and Seasonal Pattern	379
	Cyclical Pattern	382
	Identifying Time Series Patterns	382
8.2	Forecast Accuracy	382
8.3	Moving Averages and Exponential Smoothing	386
	Moving Averages	387
	Exponential Smoothing	391
8.4	Using Regression Analysis for Forecasting	395
	Linear Trend Projection	395
	Seasonality Without Trend	397
	Seasonality with Trend	398
	Using Regression Analysis as a Causal Forecasting Method	401

Combining Causal Variables with Trend and Seasonality Effects	404
Considerations in Using Regression in Forecasting	405
8.5 Determining the Best Forecasting Model to Use	405
Summary	406
Glossary	406
Problems	407
Case Problem: Forecasting Food and Beverage Sales	415
Appendix 8.1 Using the Excel Forecast Sheet	416
Appendix 8.2 Forecasting with Analytic Solver (MindTap Reader)	

CHAPTER 9 Predictive Data Mining 422

Analytics in Action: Orbitz	423
9.1 Data Sampling, Preparation, and Partitioning	424
9.2 Performance Measures	425
Evaluating the Classification of Categorical Outcomes	425
Evaluating the Estimation of Continuous Outcomes	431
9.3 Logistic Regression	432
9.4 <i>k</i> -Nearest Neighbors	436
Classifying Categorical Outcomes with <i>k</i> -Nearest Neighbors	436
Estimating Continuous Outcomes with <i>k</i> -Nearest Neighbors	438
9.5 Classification and Regression Trees	439
Classifying Categorical Outcomes with a Classification Tree	439
Estimating Continuous Outcomes with a Regression Tree	445
Ensemble Methods	446
Summary	449
Glossary	450
Problems	452
Case Problem: Grey Code Corporation	462
Available in the MindTap Reader:	
Appendix 9.1 Data Partitioning with Analytic Solver	
Appendix 9.2 Logistic Regression Classification with Analytic Solver	
Appendix 9.3 <i>k</i> -Nearest Neighbor Classification and Estimation with Analytic Solver	
Appendix 9.4 Single Classification and Regression Trees with Analytic Solver	
Appendix 9.5 Random Forests of Classification or Regression Trees with Analytic Solver	
Appendix 9.6 Data Partitioning with JMP Pro	
Appendix 9.7 Logistic Regression Classification with JMP Pro	
Appendix 9.8 <i>k</i> -Nearest Neighbor Classification and Estimation with JMP Pro	
Appendix 9.9 Single Classification and Regression Trees with JMP Pro	
Appendix 9.10 Random Forests of Classification and Regression Trees with JMP Pro	

CHAPTER 10	Spreadsheet Models	464
	Analytics in Action: Procter & Gamble	465
10.1	Building Good Spreadsheet Models	466
	Influence Diagrams	466
	Building a Mathematical Model	466
	Spreadsheet Design and Implementing the Model in a Spreadsheet	468
10.2	What-If Analysis	471
	Data Tables	471
	Goal Seek	473
	Scenario Manager	475
10.3	Some Useful Excel Functions for Modeling	480
	SUM and SUMPRODUCT	481
	IF and COUNTIF	483
	VLOOKUP	485
10.4	Auditing Spreadsheet Models	487
	Trace Precedents and Dependents	487
	Show Formulas	487
	Evaluate Formulas	489
	Error Checking	489
	Watch Window	490
10.5	Predictive and Prescriptive Spreadsheet Models	491
	Summary	492
	Glossary	492
	Problems	493
	Case Problem: Retirement Plan	499
CHAPTER 11	Monte Carlo Simulation	500
	Analytics in Action: Polio Eradication	501
11.1	Risk Analysis for Sanotronics LLC	502
	Base-Case Scenario	502
	Worst-Case Scenario	503
	Best-Case Scenario	503
	Sanotronics Spreadsheet Model	503
	Use of Probability Distributions to Represent Random Variables	504
	Generating Values for Random Variables with Excel	506
	Executing Simulation Trials with Excel	510
	Measuring and Analyzing Simulation Output	510
11.2	Simulation Modeling for Land Shark Inc.	514
	Spreadsheet Model for Land Shark	515
	Generating Values for Land Shark's Random Variables	517
	Executing Simulation Trials and Analyzing Output	519
	Generating Bid Amounts with Fitted Distributions	522
11.3	Simulation with Dependent Random Variables	527
	Spreadsheet Model for Press Teag Worldwide	527

11.4	Simulation Considerations	532
	Verification and Validation	532
	Advantages and Disadvantages of Using Simulation	532
	Summary	533
	Glossary	534
	Problems	534
	Case Problem: Four Corners	547
	Appendix 11.1 Common Probability Distributions for Simulation	549

Available in the MindTap Reader:

Appendix 11.2	Land Shark Inc. Simulation with Analytic Solver
Appendix 11.3	Distribution Fitting with Analytic Solver
Appendix 11.4	Correlating Random Variables with Analytic Solver
Appendix 11.5	Simulation Optimization with Analytic Solver

CHAPTER 12 Linear Optimization Models 556

	Analytics in Action: General Electric	557
12.1	A Simple Maximization Problem	558
	Problem Formulation	559
	Mathematical Model for the Par, Inc. Problem	561
12.2	Solving the Par, Inc. Problem	561
	The Geometry of the Par, Inc. Problem	562
	Solving Linear Programs with Excel Solver	564
12.3	A Simple Minimization Problem	568
	Problem Formulation	568
	Solution for the M&D Chemicals Problem	568
12.4	Special Cases of Linear Program Outcomes	570
	Alternative Optimal Solutions	571
	Infeasibility	572
	Unbounded	573
12.5	Sensitivity Analysis	575
	Interpreting Excel Solver Sensitivity Report	575
12.6	General Linear Programming Notation and More Examples	577
	Investment Portfolio Selection	578
	Transportation Planning	580
	Advertising Campaign Planning	584
12.7	Generating an Alternative Optimal Solution for a Linear Program	589
	Summary	591
	Glossary	592
	Problems	593
	Case Problem: Investment Strategy	604
	Appendix 12.1 Solving Linear Optimization Models Using Analytic Solver (MindTap Reader)	

CHAPTER 13 Integer Linear Optimization Models 606

Analytics in Action: Petrobras 607

13.1 Types of Integer Linear Optimization Models 607

13.2 Eastborne Realty, an Example of Integer Optimization 608
The Geometry of Linear All-Integer Optimization 60913.3 Solving Integer Optimization Problems with Excel Solver 611
A Cautionary Note About Sensitivity Analysis 614

13.4 Applications Involving Binary Variables 616

Capital Budgeting 616

Fixed Cost 618

Bank Location 621

Product Design and Market Share Optimization 623

13.5 Modeling Flexibility Provided by Binary Variables 626

Multiple-Choice and Mutually Exclusive Constraints 626

 k Out of n Alternatives Constraint 627

Conditional and Corequisite Constraints 627

13.6 Generating Alternatives in Binary Optimization 628

Summary 630

Glossary 631

Problems 632

Case Problem: Applecore Children's Clothing 643

Appendix 13.1 Solving Integer Linear Optimization Problems Using
Analytic Solver (MindTap Reader)**CHAPTER 14 Nonlinear Optimization Models 646**

Analytics in Action: InterContinental Hotels 647

14.1 A Production Application: Par, Inc. Revisited 647

An Unconstrained Problem 647

A Constrained Problem 648

Solving Nonlinear Optimization Models Using Excel Solver 650

Sensitivity Analysis and Shadow Prices in Nonlinear Models 651

14.2 Local and Global Optima 652

Overcoming Local Optima with Excel Solver 655

14.3 A Location Problem 657

14.4 Markowitz Portfolio Model 658

14.5 Forecasting Adoption of a New Product 663

Summary 666

Glossary 667

Problems 667

Case Problem: Portfolio Optimization with Transaction Costs 675

Appendix 14.1 Solving Nonlinear Optimization Problems with Analytic
Solver (MindTap Reader)

CHAPTER 15 Decision Analysis 678

Analytics in Action: Phytopharm 679

15.1 Problem Formulation 680

Payoff Tables 681

Decision Trees 681

15.2 Decision Analysis without Probabilities 682

Optimistic Approach 682

Conservative Approach 683

Minimax Regret Approach 683

15.3 Decision Analysis with Probabilities 685

Expected Value Approach 685

Risk Analysis 687

Sensitivity Analysis 688

15.4 Decision Analysis with Sample Information 689

Expected Value of Sample Information 694

Expected Value of Perfect Information 694

15.5 Computing Branch Probabilities with Bayes' Theorem 695

15.6 Utility Theory 698

Utility and Decision Analysis 699

Utility Functions 703

Exponential Utility Function 706

Summary 708

Glossary 708

Problems 710

Case Problem: Property Purchase Strategy 721

Appendix 15.1 Using Analytic Solver to Create Decision Trees
(MindTap Reader)**APPENDIX A** Basics of Excel 724**APPENDIX B** Database Basics with Microsoft Access 736**APPENDIX C** Solutions to Even-Numbered Questions (MindTap Reader)**REFERENCES** 774**INDEX** 776

About the Authors

Jeffrey D. Camm. Jeffrey D. Camm is the Inmar Presidential Chair and Associate Dean of Analytics in the School of Business at Wake Forest University. Born in Cincinnati, Ohio, he holds a B.S. from Xavier University (Ohio) and a Ph.D. from Clemson University. Prior to joining the faculty at Wake Forest, he was on the faculty of the University of Cincinnati. He has also been a visiting scholar at Stanford University and a visiting professor of business administration at the Tuck School of Business at Dartmouth College.

Dr. Camm has published over 35 papers in the general area of optimization applied to problems in operations management and marketing. He has published his research in *Science*, *Management Science*, *Operations Research*, *Interfaces*, and other professional journals. Dr. Camm was named the Dornoff Fellow of Teaching Excellence at the University of Cincinnati and he was the 2006 recipient of the INFORMS Prize for the Teaching of Operations Research Practice. A firm believer in practicing what he preaches, he has served as an analytics consultant to numerous companies and government agencies. From 2005 to 2010 he served as editor-in-chief of *Interfaces*. In 2016, Dr. Camm was awarded the Kimball Medal for service to the operations research profession and in 2017 he was named an INFORMS Fellow.

James J. Cochran. James J. Cochran is Associate Dean for Research, Professor of Applied Statistics, and the Rogers-Spivey Faculty Fellow at the University of Alabama. Born in Dayton, Ohio, he earned his B.S., M.S., and M.B.A. degrees from Wright State University and a Ph.D. from the University of Cincinnati. He has been at the University of Alabama since 2014 and has been a visiting scholar at Stanford University, Universidad de Talca, the University of South Africa, and Pole Universitaire Leonard de Vinci.

Professor Cochran has published over three 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*, *Annals of Operations Research*, *European Journal of Operational Research*, *Journal of Combinatorial Optimization*, *Interfaces*, *Statistics and Probability Letters*, and other professional journals. He was the 2008 recipient of the INFORMS Prize for the Teaching of Operations Research Practice and the 2010 recipient of the Mu Sigma Rho Statistical Education Award. Professor Cochran was elected to the International Statistics Institute in 2005 and named a Fellow of the American Statistical Association in 2011 and a Fellow of the Institute for Operations Research and the Management Sciences (INFORMS) in 2017. He also received the Founders Award in 2014, the Karl E. Peace Award in 2015, and the Waller Distinguished Teaching Career Award in 2017 from the American Statistical Association. 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; Nairobi, Kenya; Buea, Cameroon; Suva, Fiji; Kathmandu, Nepal; Osijek, Croatia; Havana, Cuba; Ulaanbaatar, Mongolia; and Chişinău, Moldova. He has served as an operations research consultant to numerous companies and not-for-profit organizations. He served as editor-in-chief of *INFORMS Transactions on Education* from 2006 to 2012 and is on the editorial board of *Interfaces*, *International Transactions in Operational Research*, and *Significance*.

Michael J. Fry. Michael J. Fry is Professor and Head of the Department of Operations, Business Analytics, and Information Systems in the Carl H. Lindner College of Business at the University of Cincinnati. Born in Killeen, Texas, he earned a B.S. from Texas A&M University, and M.S.E. and Ph.D. degrees from the University of Michigan. He has been

at the University of Cincinnati since 2002, where he has been named a Lindner Research Fellow and has served as Assistant Director and Interim Director of the Center for Business Analytics. He has also been a visiting professor at Cornell University and at the University of British Columbia.

Professor Fry has published over 20 research papers in journals such as *Operations Research*, *Manufacturing & Service Operations Management*, *Transportation Science*, *Naval Research Logistics*, *IIE Transactions*, and *Interfaces*. His research interests are in applying quantitative management methods to the areas of supply chain analytics, sports analytics, and public-policy operations. He has worked with many different organizations for his research, including Dell, Inc., Copeland Corporation, Starbucks Coffee Company, Great American Insurance Group, the Cincinnati Fire Department, the State of Ohio Election Commission, the Cincinnati Bengals, and the Cincinnati Zoo. In 2008, he was named a finalist for the Daniel H. Wagner Prize for Excellence in Operations Research Practice, and he has been recognized for both his research and teaching excellence at the University of Cincinnati.

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

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

David R. Anderson. David R. Anderson is Professor Emeritus of Quantitative Analysis in the Carl H. Lindner College of Business 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. In addition, he was the coordinator of the College's first Executive Program.

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

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

Dennis J. Sweeney. Dennis J. Sweeney is Professor Emeritus of Quantitative Analysis and Founder of the Center for Productivity Improvement at the University of Cincinnati. Born in Des Moines, Iowa, he earned a B.S.B.A. degree from Drake University and his M.B.A. and D.B.A. degrees from Indiana University, where he was an NDEA Fellow. During 1978–1979, Professor Sweeney worked in the management science group at Procter & Gamble; during 1981–1982, he was 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 areas of management science and statistics. The National Science Foundation, IBM, Procter & Gamble, Federated Department Stores, Kroger, and Cincinnati Gas & Electric have funded his research, which has been published in *Management Science*, *Operations Research*, *Mathematical Programming*, *Decision Sciences*, and other journals.

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

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

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

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

Preface

Business Analytics 3E is designed to introduce the concept of business analytics to undergraduate and graduate students. This textbook contains one of the first collections of materials that are essential to the growing field of business analytics. In Chapter 1 we present an overview of business analytics and our approach to the material in this textbook. In simple terms, business analytics helps business professionals make better decisions based on data. We discuss models for summarizing, visualizing, and understanding useful information from historical data in Chapters 2 through 6. Chapters 7 through 9 introduce methods for both gaining insights from historical data and predicting possible future outcomes. Chapter 10 covers the use of spreadsheets for examining data and building decision models. In Chapter 11, we demonstrate how to explicitly introduce uncertainty into spreadsheet models through the use of Monte Carlo simulation. In Chapters 12 through 14 we discuss optimization models to help decision makers choose the best decision based on the available data. Chapter 15 is an overview of decision analysis approaches for incorporating a decision maker's views about risk into decision making. In Appendix A we present optional material for students who need to learn the basics of using Microsoft Excel. The use of databases and manipulating data in Microsoft Access is discussed in Appendix B.

This textbook can be used by students who have previously taken a course on basic statistical methods as well as students who have not had a prior course in statistics. *Business Analytics* 3E is also amenable to a two-course sequence in business statistics and analytics. All statistical concepts contained in this textbook are presented from a business analytics perspective using practical business examples. Chapters 2, 5, 6, and 7 provide an introduction to basic statistical concepts that form the foundation for more advanced analytics methods. Chapters 3, 4, and 9 cover additional topics of data visualization and data mining that are not traditionally part of most introductory business statistics courses, but they are exceedingly important and commonly used in current business environments. Chapter 10 and Appendix A provide the foundational knowledge students need to use Microsoft Excel for analytics applications. Chapters 11 through 15 build upon this spreadsheet knowledge to present additional topics that are used by many organizations that are leaders in the use of prescriptive analytics to improve decision making.

Updates in the Third Edition

The third edition of *Business Analytics* is a major revision. We have heavily modified our data mining chapters to allow instructors to choose their preferred means of teaching this material in terms of software usage. Chapters 4 and 9 now both contain conceptual homework problems that can be solved by students without using any software. Additionally, we now include online appendices on both Analytic Solver and JMP Pro as software for teaching data mining so that instructors can choose their favored way of teaching this material. Chapter 4 also now includes a section on text mining, a fast-growing topic in business analytics. We have moved our chapter on Monte Carlo simulation to Chapter 11, and we have completely rewritten this chapter to greatly expand the material that can be covered using only native Excel. Other changes in this edition include additional content on big-data concepts, data cleansing, new data visualization topics in Excel, and additional homework problems.

- **Software Updates for Data Mining Chapters.** Chapters 4 and 9 have received extensive updates. The end-of-chapter problems are now written so that they can be solved using any data-mining software. To allow instructors to choose different software for use with these chapters, we have created online appendices for both Analytic Solver and JMP Pro. Analytic Solver has undergone major changes since the previous edition of this textbook. Therefore, we have reworked all examples, problems, and cases using Analytic Solver Basic V2017, the version of this software now available to students. We have created new appendices for Chapters 4 and 9 that introduce the use of JMP Pro 13 for data mining. JMP Pro is a powerful software that is still easy to learn and easy to use. We have also added five homework problems to Chapters 4 and 9 that can

be solved without using any software. This allows instructors to cover the basics of data mining without using any additional software. The online appendices for Chapters 4 and 9 also include Analytic Solver and JMP Pro specific instructions for how to solve the end-of-chapter problems using Analytic Solver or JMP Pro. Problem and case solutions using both Analytic Solver and JMP Pro are also available to instructors.

- **New Section on Text Mining.** Chapter 4 now includes a section on text mining. With the proliferation of unstructured data, generating insights from text is becoming increasingly important. We have added two problems related to text mining to Chapter 4. We also include online appendices for using either Analytic Solver or JMP Pro to perform basic text mining functions.
- **Revision of Monte Carlo Simulation Chapter.** Our chapter on simulation models has been heavily revised. In the body of the chapter, we construct simulation models solely using native Excel functionality. This pedagogical design choice is based on the authors' own experiences and motivated by the following factors: (1) it primarily avoids software incompatibility issues for students with different operating systems (Apple OS versus Microsoft Windows); and (2) it separates simulation concepts from software steps so that students realize that they do not need a specific software package to utilize simulation in their future careers.

To support our approach, we have added many more topics and examples that can be taught using native Excel functions. Our coverage now guides the instructor and student through many different types of simulation models and output analysis using only native Excel. However, if instructors wish to utilize specialized Monte Carlo simulation software, the examples and problems in the chapter can all be solved with specialized software. To demonstrate, we include an updated online appendix for using Analytic Solver to create simulation models and perform output analysis.

We have also moved our chapter on simulation models to Chapter 11, prior to the chapters on optimization models. We believe this presents a better ordering of topics as it follows immediately after Chapter 10 that covers good design techniques for Excel spreadsheet models. In particular, we have added a new section in Chapter 10 on using the Scenario Manager tool in Excel that creates a natural bridge to the coverage of simulation models in Chapter 11. The end-of-chapter problems and case in Chapter 11 can all be solved using native Excel. Problem and case solutions for Chapter 11 using both native Excel and Analytic Solver are available to instructors.

- **Additional Material on Big-Data Topics.** We have added new sections in Chapters 6 and 7 to enhance our coverage of topics related to big data. In Chapter 6, we introduce the concept of big data, and we discuss some additional challenges and implications of applying statistical inference when you have very large sample sizes. In Chapter 7, we expand on these concepts by discussing the estimation and use of regression models with very large sample sizes.
- **New Data Analysis and Data Visualization Tools in Excel.** Excel 2016 introduces several new tools for data analysis and data visualization. Chapter 2 now covers how to create a box plot in native Excel. In Chapter 3 we have added coverage of how to create more advanced data visualization tools in native Excel such as treemaps and geographic information system (GIS) charts.
- **New Section on Data Cleansing.** Chapter 2 now includes a section on data cleansing. This section introduces concepts related to missing data, outliers, and variable representation. These are exceptionally important concepts that face all analytics professionals when dealing with real data that can have missing values and errors.
- **Excel Forecast Sheet.** As we did in the second edition, Chapter 8 includes an appendix for using the Forecast Sheet tool in Excel 2016. Excel's Forecast Sheet tool implements a time series forecasting model known as the Holt-Winters additive seasonal smoothing model.
- **New End-of-Chapter Problems.** The third edition of this textbook includes new problems in Chapters 2, 3, 4, 6, 9, 10, 11, 13, and 14. As we have done in past editions, Excel solution files are available to instructors for problems that require the use of Excel.

- **Software Appendices Moved Online.** Chapter appendices that deal with the use of Analytic Solver or JMP Pro have been moved online as part of the MindTap Reader eBook. This preserves the flow of material in the textbook and allows instructors to easily cover all material using only native Excel if that is preferred. The online appendices offer extensive coverage of Analytic Solver for concepts covered in Chapters 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, and 15. Online appendices for Chapters 4 and 9 cover the usage of JMP Pro for data mining. Contact your Cengage representative for more information on how your students can access the MindTap Reader.

Continued Features and Pedagogy

In the third edition of this textbook, we continue to offer all of the features that have been successful in the first two editions. Some of the specific features that we use in this textbook are listed below.

- **Integration of Microsoft Excel:** Excel has been thoroughly integrated throughout this textbook. For many methodologies, we provide instructions for how to perform calculations both by hand and with Excel. In other cases where realistic models are practical only with the use of a spreadsheet, we focus on the use of Excel to describe the methods to be used.
- **Notes and Comments:** At the end of many sections, we provide Notes and Comments to give the student additional insights about the methods presented in that section. These insights include comments on the limitations of the presented methods, recommendations for applications, and other matters. Additionally, margin notes are used throughout the textbook to provide additional insights and tips related to the specific material being discussed.
- **Analytics in Action:** Each chapter contains an Analytics in Action article. These articles present interesting examples of the use of business analytics in practice. The examples are drawn from many different organizations in a variety of areas including healthcare, finance, manufacturing, marketing, and others.
- **DATAfiles and MODELfiles:** All data sets used as examples and in student exercises are also provided online on the companion site as files available for download by the student. DATAfiles are Excel files that contain data needed for the examples and problems given in the textbook. MODELfiles contain additional modeling features such as extensive use of Excel formulas or the use of Excel Solver, Analytic Solver, or JMP Pro.
- **Problems and Cases:** With the exception of Chapter 1, each chapter contains an extensive selection of problems to help the student master the material presented in that chapter. The problems vary in difficulty and most relate to specific examples of the use of business analytics in practice. Answers to even-numbered problems are provided in an online supplement for student access. With the exception of Chapter 1, each chapter also includes an in-depth case study that connects many of the different methods introduced in the chapter. The case studies are designed to be more open-ended than the chapter problems, but enough detail is provided to give the student some direction in solving the cases.

MindTap

MindTap is a customizable digital course solution that includes an interactive eBook, auto-graded exercises from the textbook, algorithmic practice problems with solutions feedback, Exploring Analytics visualizations, Adaptive Test Prep, and more! All of these materials offer students better access to resources 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.

- **Analytic Solver:** Instructions to download an educational version of Frontline Systems' Analytic Solver Basic V2017 are included with the purchase of this textbook. These instructions can be found within the inside front cover of this text, within MindTap in the 'Course Materials' folder, and online on the text companion site at www.cengagebrain.com.
Note that there is now a small charge for one-semester access to Analytic Solver. For more information on pricing and available discounts for student users, please contact Frontline Systems directly at 775-831-0300 and press 0 for advice on courses and software licenses.
- **JMP Pro:** Most universities have site licenses of SAS Institute's JMP Pro software on both Mac and Windows. These are typically offered through your university's software licensing administrator. Faculty may contact the JMP Academic team to find out if their universities have a license or to request a complementary instructor copy at www.jmp.com/contact-academic. For institutions without a site license, students may rent a 6 or 12-month license for JMP at www.onthefhub.com/jmp.

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. Excel solution files are available to instructors for those problems that require the use of Excel. Solutions for Chapters 4 and 9 are available using both Analytic Solver and JMP Pro for data mining problems. Solutions for Chapter 11 are available using both native Excel and Analytic Solver for simulation problems.
- **Solutions to Case Problems:** These are also prepared by the authors and contain solutions to all case problems presented in the text. Case solutions for Chapters 4 and 9 are provided using both Analytic Solver and JMP Pro. Case solutions for Chapter 11 are available using both native Excel and Analytic Solver.
- **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 Learning Management System (LMS), your classroom, or wherever you want. The Test Bank is also available in Microsoft Word.

Acknowledgments

We would like to acknowledge the work of reviewers and users who have provided comments and suggestions for improvement of this text. Thanks to:

Matthew D. Bailey
Bucknell University

Phillip Beaver
University of Denver

M. Khurram S. Bhutta
Ohio University

Paolo Catasti
Virginia Commonwealth University

Q B. Chung
Villanova University

Elizabeth A. Denny
University of Kentucky

Mike Taein Eom
University of Portland

Yvette Njan Essounga
Fayetteville State University

Lawrence V. Fulton
Texas State University

Tom Groleau
Carthage College

James F. Hoelscher
Lincoln Memorial University

Eric Huggins
Fort Lewis College

Faizul Huq
Ohio University

Marco Lam
York College of Pennsylvania

Thomas Lee
University of California, Berkeley

Roger Myerson
Northwestern University

Ram Pakath
University of Kentucky

Susan Palocsay
James Madison University

Andy Shogan
University of California, Berkeley

Dothan Truong
Embry-Riddle Aeronautical University

Kai Wang
Wake Technical Community College

Ed Wasil
American University

We are indebted to our Senior Product Team Manager, Joe Sabatino and our Senior Product Manager, Aaron Arnsparger; our Senior Digital Content Designer, Brandon Foltz, our Senior Marketing Manager, Nathan Anderson, and our Content Developer, Anne Merrill; our Content Project Manager, Jean Buttrom; and others at Cengage Learning for their counsel and support during the preparation of this text.

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

Chapter 1

Introduction

CONTENTS

- 1.1 **DECISION MAKING**
- 1.2 **BUSINESS ANALYTICS DEFINED**
- 1.3 **A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS**
 - Descriptive Analytics
 - Predictive Analytics
 - Prescriptive Analytics
- 1.4 **BIG DATA**
 - Volume
 - Velocity
 - Variety
 - Veracity
- 1.5 **BUSINESS ANALYTICS IN PRACTICE**
 - Financial Analytics
 - Human Resource (HR) Analytics
 - Marketing Analytics
 - Health Care Analytics
 - Supply-Chain Analytics
 - Analytics for Government and Nonprofits
 - Sports Analytics
 - Web Analytics

You apply for a loan for the first time. How does the bank assess the riskiness of the loan it might make to you? How does Amazon.com know which books and other products to recommend to you when you log in to their web site? How do airlines determine what price to quote to you when you are shopping for a plane ticket? How can doctors better diagnose and treat you when you are ill or injured?

You may be applying for a loan for the first time, but millions of people around the world have applied for loans before. Many of these loan recipients have paid back their loans in full and on time, but some have not. The bank wants to know whether you are more like those who have paid back their loans or more like those who defaulted. By comparing your credit history, financial situation, and other factors to the vast database of previous loan recipients, the bank can effectively assess how likely you are to default on a loan.

Similarly, Amazon.com has access to data on millions of purchases made by customers on its web site. Amazon.com examines your previous purchases, the products you have viewed, and any product recommendations you have provided. Amazon.com then searches through its huge database for customers who are similar to you in terms of product purchases, recommendations, and interests. Once similar customers have been identified, their purchases form the basis of the recommendations given to you.

Prices for airline tickets are frequently updated. The price quoted to you for a flight between New York and San Francisco today could be very different from the price that will be quoted tomorrow. These changes happen because airlines use a pricing strategy known as revenue management. Revenue management works by examining vast amounts of data on past airline customer purchases and using these data to forecast future purchases. These forecasts are then fed into sophisticated optimization algorithms that determine the optimal price to charge for a particular flight and when to change that price. Revenue management has resulted in substantial increases in airline revenues.

Finally, consider the case of being evaluated by a doctor for a potentially serious medical issue. Hundreds of medical papers may describe research studies done on patients facing similar diagnoses, and thousands of data points exist on their outcomes. However, it is extremely unlikely that your doctor has read every one of these research papers or is aware of all previous patient outcomes. Instead of relying only on her medical training and knowledge gained from her limited set of previous patients, wouldn't it be better for your doctor to have access to the expertise and patient histories of thousands of doctors around the world?

A group of IBM computer scientists initiated a project to develop a new decision technology to help in answering these types of questions. That technology is called Watson, named after the founder of IBM, Thomas J. Watson. The team at IBM focused on one aim: How the vast amounts of data now available on the Internet can be used to make more data-driven, smarter decisions.

Watson became a household name in 2011, when it famously won the television game show, *Jeopardy!* Since that proof of concept in 2011, IBM has reached agreements with the health insurance provider WellPoint (now part of Anthem), the financial services company Citibank, Memorial Sloan-Kettering Cancer Center, and automobile manufacturer General Motors to apply Watson to the decision problems that they face.

Watson is a system of computing hardware, high-speed data processing, and analytical algorithms that are combined to make data-based recommendations. As more and more data are collected, Watson has the capability to learn over time. In simple terms, according to IBM, Watson gathers hundreds of thousands of possible solutions from a huge data bank, evaluates them using analytical techniques, and proposes only the best solutions for consideration. Watson provides not just a single solution, but rather a range of good solutions with a confidence level for each.

For example, at a data center in Virginia, to the delight of doctors and patients, Watson is already being used to speed up the approval of medical procedures. Citibank is beginning to explore how to use Watson to better serve its customers, and cancer specialists at

more than a dozen hospitals in North America are using Watson to assist with the diagnosis and treatment of patients.¹

This book is concerned with data-driven decision making and the use of analytical approaches in the decision-making process. Three developments spurred recent explosive growth in the use of analytical methods in business applications. First, technological advances—such as improved point-of-sale scanner technology and the collection of data through e-commerce and social networks, data obtained by sensors on all kinds of mechanical devices such as aircraft engines, automobiles, and farm machinery through the so-called Internet of Things and data generated from personal electronic devices—produce incredible amounts of data for businesses. Naturally, businesses want to use these data to improve the efficiency and profitability of their operations, better understand their customers, price their products more effectively, and gain a competitive advantage. Second, ongoing research has resulted in numerous methodological developments, including advances in computational approaches to effectively handle and explore massive amounts of data, faster algorithms for optimization and simulation, and more effective approaches for visualizing data. Third, these methodological developments were paired with an explosion in computing power and storage capability. Better computing hardware, parallel computing, and, more recently, cloud computing (the remote use of hardware and software over the Internet) have enabled businesses to solve big problems more quickly and more accurately than ever before.

In summary, the availability of massive amounts of data, improvements in analytic methodologies, and substantial increases in computing power have all come together to result in a dramatic upsurge in the use of analytical methods in business and a reliance on the discipline that is the focus of this text: business analytics. As stated in the Preface, the purpose of this text is to provide students with a sound conceptual understanding of the role that business analytics plays in the decision-making process. To reinforce the applications orientation of the text and to provide a better understanding of the variety of applications in which analytical methods have been used successfully, Analytics in Action articles are presented throughout the book. Each Analytics in Action article summarizes an application of analytical methods in practice.

1.1 Decision Making

It is the responsibility of managers to plan, coordinate, organize, and lead their organizations to better performance. Ultimately, managers' responsibilities require that they make strategic, tactical, or operational decisions. **Strategic decisions** involve higher-level issues concerned with the overall direction of the organization; these decisions define the organization's overall goals and aspirations for the future. Strategic decisions are usually the domain of higher-level executives and have a time horizon of three to five years. **Tactical decisions** concern how the organization should achieve the goals and objectives set by its strategy, and they are usually the responsibility of midlevel management. Tactical decisions usually span a year and thus are revisited annually or even every six months. **Operational decisions** affect how the firm is run from day to day; they are the domain of operations managers, who are the closest to the customer.

Consider the case of the Thoroughbred Running Company (TRC). Historically, TRC had been a catalog-based retail seller of running shoes and apparel. TRC sales revenues grew quickly as it changed its emphasis from catalog-based sales to Internet-based sales. Recently, TRC decided that it should also establish retail stores in the malls and downtown areas of major cities. This strategic decision will take the firm in a new direction that it hopes will complement its Internet-based strategy. TRC middle managers will therefore have to make a variety of tactical decisions in support of this strategic decision, including

¹"IBM's Watson Is Learning Its Way to Saving Lives," Fastcompany web site, December 8, 2012; "IBM's Watson Targets Cancer and Enlists Prominent Providers in the Fight," ModernHealthcare web site, May 5, 2015.

how many new stores to open this year, where to open these new stores, how many distribution centers will be needed to support the new stores, and where to locate these distribution centers. Operations managers in the stores will need to make day-to-day decisions regarding, for instance, how many pairs of each model and size of shoes to order from the distribution centers and how to schedule their sales personnel's work time.

Regardless of the level within the firm, *decision making* can be defined as the following process:

1. Identify and define the problem.
2. Determine the criteria that will be used to evaluate alternative solutions.
3. Determine the set of alternative solutions.
4. Evaluate the alternatives.
5. Choose an alternative.

If I were given one hour to save the planet, I would spend 59 minutes defining the problem and one minute resolving it.

—Albert Einstein

Step 1 of decision making, identifying and defining the problem, is the most critical. Only if the problem is well-defined, with clear metrics of success or failure (step 2), can a proper approach for solving the problem (steps 3 and 4) be devised. Decision making concludes with the choice of one of the alternatives (step 5).

There are a number of approaches to making decisions: tradition (“We’ve always done it this way”), intuition (“gut feeling”), and rules of thumb (“As the restaurant owner, I schedule twice the number of waiters and cooks on holidays”). The power of each of these approaches should not be underestimated. Managerial experience and intuition are valuable inputs to making decisions, but what if relevant data were available to help us make more informed decisions? With the vast amounts of data now generated and stored electronically, it is estimated that the amount of data stored by businesses more than doubles every two years. How can managers convert these data into knowledge that they can use to be more efficient and effective in managing their businesses?

1.2 Business Analytics Defined

What makes decision making difficult and challenging? Uncertainty is probably the number one challenge. If we knew how much the demand will be for our product, we could do a much better job of planning and scheduling production. If we knew exactly how long each step in a project will take to be completed, we could better predict the project's cost and completion date. If we knew how stocks will perform, investing would be a lot easier.

Another factor that makes decision making difficult is that we often face such an enormous number of alternatives that we cannot evaluate them all. What is the best combination of stocks to help me meet my financial objectives? What is the best product line for a company that wants to maximize its market share? How should an airline price its tickets so as to maximize revenue?

Business analytics is the scientific process of transforming data into insight for making better decisions.² Business analytics is used for data-driven or fact-based decision making, which is often seen as more objective than other alternatives for decision making.

As we shall see, the tools of business analytics can aid decision making by creating insights from data, by improving our ability to more accurately forecast for planning, by helping us quantify risk, and by yielding better alternatives through analysis and optimization. A study based on a large sample of firms that was conducted by researchers at MIT's Sloan School of Management and the University of Pennsylvania, concluded that firms guided by data-driven decision making have higher productivity and market value and increased output and profitability.³

*Some firms and industries use the simpler term, **analytics**. Analytics is often thought of as a broader category than **business analytics**, encompassing the use of analytical techniques in the sciences and engineering as well. In this text, we use **business analytics** and **analytics** synonymously.*

²We adopt the definition of analytics developed by the Institute for Operations Research and the Management Sciences (INFORMS).

³E. Brynjolfsson, L. M. Hitt, and H. H. Kim, “Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?” (April 18, 2013). Available at SSRN, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486.

1.3 A Categorization of Analytical Methods and Models

Business analytics can involve anything from simple reports to the most advanced optimization techniques (methods for finding the best course of action). Analytics is generally thought to comprise three broad categories of techniques: descriptive analytics, predictive analytics, and prescriptive analytics.

Descriptive Analytics

Descriptive analytics encompasses the set of techniques that describes what has happened in the past. Examples are data queries, reports, descriptive statistics, data visualization including data dashboards, some data-mining techniques, and basic what-if spreadsheet models.

A **data query** is a request for information with certain characteristics from a database. For example, a query to a manufacturing plant's database might be for all records of shipments to a particular distribution center during the month of March. This query provides descriptive information about these shipments: the number of shipments, how much was included in each shipment, the date each shipment was sent, and so on. A report summarizing relevant historical information for management might be conveyed by the use of descriptive statistics (means, measures of variation, etc.) and data-visualization tools (tables, charts, and maps). Simple descriptive statistics and data-visualization techniques can be used to find patterns or relationships in a large database.

Data dashboards are collections of tables, charts, maps, and summary statistics that are updated as new data become available. Dashboards are used to help management monitor specific aspects of the company's performance related to their decision-making responsibilities. For corporate-level managers, daily data dashboards might summarize sales by region, current inventory levels, and other company-wide metrics; front-line managers may view dashboards that contain metrics related to staffing levels, local inventory levels, and short-term sales forecasts.

Data mining is the use of analytical techniques for better understanding patterns and relationships that exist in large data sets. For example, by analyzing text on social network platforms like Twitter, data-mining techniques (including cluster analysis and sentiment analysis) are used by companies to better understand their customers. By categorizing certain words as positive or negative and keeping track of how often those words appear in tweets, a company like Apple can better understand how its customers are feeling about a product like the Apple Watch.

Predictive Analytics

Predictive analytics consists of techniques that use models constructed from past data to predict the future or ascertain the impact of one variable on another. For example, past data on product sales may be used to construct a mathematical model to predict future sales. This mode can factor in the product's growth trajectory and seasonality based on past patterns. A packaged-food manufacturer may use point-of-sale scanner data from retail outlets to help in estimating the lift in unit sales due to coupons or sales events. Survey data and past purchase behavior may be used to help predict the market share of a new product. All of these are applications of predictive analytics.

Linear regression, time series analysis, some data-mining techniques, and simulation, often referred to as risk analysis, all fall under the banner of predictive analytics. We discuss all of these techniques in greater detail later in this text.

Data mining, previously discussed as a descriptive analytics tool, is also often used in predictive analytics. For example, a large grocery store chain might be interested in developing a targeted marketing campaign that offers a discount coupon on potato chips. By studying historical point-of-sale data, the store may be able to use data mining to predict which customers are the most likely to respond to an offer on discounted chips by purchasing higher-margin items such as beer or soft drinks in addition to the chips, thus increasing the store's overall revenue.

Appendix B at the end of this book describes how to use Microsoft Access to conduct data queries.

Simulation involves the use of probability and statistics to construct a computer model to study the impact of uncertainty on a decision. For example, banks often use simulation to model investment and default risk in order to stress-test financial models. Simulation is also often used in the pharmaceutical industry to assess the risk of introducing a new drug.

Prescriptive Analytics

Prescriptive analytics differs from descriptive and predictive analytics in that **prescriptive analytics** indicates a course of action to take; that is, the output of a prescriptive model is a decision. Predictive models provide a forecast or prediction, but do not provide a decision. However, a forecast or prediction, when combined with a rule, becomes a prescriptive model. For example, we may develop a model to predict the probability that a person will default on a loan. If we create a rule that says if the estimated probability of default is more than 0.6, we should not award a loan, now the predictive model, coupled with the rule is prescriptive analytics. These types of prescriptive models that rely on a rule or set of rules are often referred to as **rule-based models**.

Other examples of prescriptive analytics are portfolio models in finance, supply network design models in operations, and price-markdown models in retailing. Portfolio models use historical investment return data to determine which mix of investments will yield the highest expected return while controlling or limiting exposure to risk. Supply-network design models provide plant and distribution center locations that will minimize costs while still meeting customer service requirements. Given historical data, retail price markdown models yield revenue-maximizing discount levels and the timing of discount offers when goods have not sold as planned. All of these models are known as **optimization models**, that is, models that give the best decision subject to the constraints of the situation.

Another type of modeling in the prescriptive analytics category is **simulation optimization** which combines the use of probability and statistics to model uncertainty with optimization techniques to find good decisions in highly complex and highly uncertain settings. Finally, the techniques of **decision analysis** can be used to develop an optimal strategy when a decision maker is faced with several decision alternatives and an uncertain set of future events. Decision analysis also employs **utility theory**, which assigns values to outcomes based on the decision maker's attitude toward risk, loss, and other factors.

In this text we cover all three areas of business analytics: descriptive, predictive, and prescriptive. Table 1.1 shows how the chapters cover the three categories.

1.4 Big Data

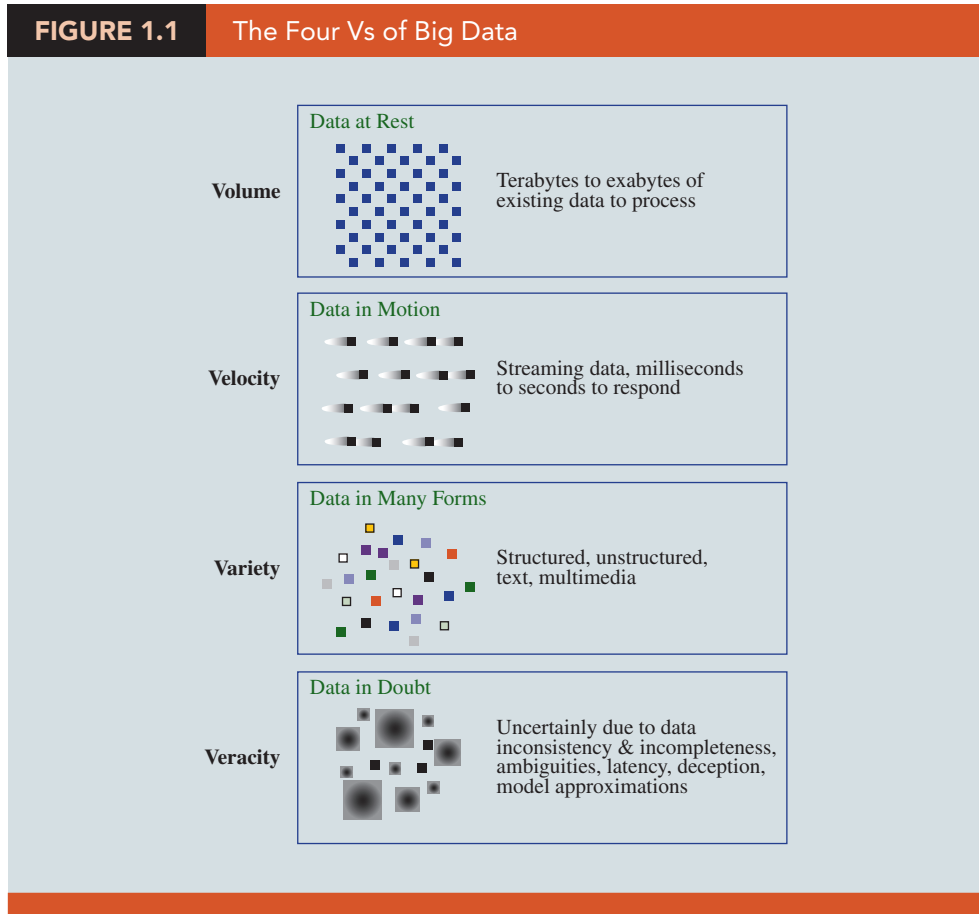
Walmart handles over 1 million purchase transactions per hour. Facebook processes more than 250 million picture uploads per day. Six billion cell phone owners around the world generate vast amounts of data by calling, texting, tweeting, and browsing the web on a daily basis.⁴ As Google CEO Eric Schmidt has noted, the amount of data currently created every 48 hours is equivalent to the entire amount of data created from the dawn of civilization until the year 2003. It is through technology that we have truly been thrust into the data age. Because data can now be collected electronically, the available amounts of it are staggering. The Internet, cell phones, retail checkout scanners, surveillance video, and sensors on everything from aircraft to cars to bridges allow us to collect and store vast amounts of data in real time.

In the midst of all of this data collection, the new term *big data* has been created. There is no universally accepted definition of big data. However, probably the most accepted and most general definition is that **big data** is any set of data that is too large or too complex to be handled by standard data-processing techniques and typical desktop software. IBM describes the phenomenon of big data through the four Vs: volume, velocity, variety, and veracity, as shown in Figure 1.1.⁵

⁴SAS White Paper, "Big Data Meets Big Data Analytics," SAS Institute, 2012.

⁵IBM web site: http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg.

TABLE 1.1		Coverage of Business Analytics Topics in This Text		
Chapter	Title	Descriptive	Predictive	Prescriptive
1	Introduction	●	●	●
2	Descriptive Statistics	●		
3	Data Visualization	●		
4	Descriptive Data Mining	●		
5	Probability: An Introduction to Modeling Uncertainty	●		
6	Statistical Inference	●		
7	Linear Regression		●	
8	Time Series and Forecasting		●	
9	Predictive Data Mining		●	
10	Spreadsheet Models	●	●	●
11	Monte Carlo Simulation		●	●
12	Linear Optimization Models			●
13	Integer Linear Optimization Models			●
14	Nonlinear Optimization Models			●
15	Decision Analysis			●



Source: IBM.

Volume

Because data are collected electronically, we are able to collect more of it. To be useful, these data must be stored, and this storage has led to vast quantities of data. Many companies now store in excess of 100 terabytes of data (a terabyte is 1,024 gigabytes).

Velocity

Real-time capture and analysis of data present unique challenges both in how data are stored and the speed with which those data can be analyzed for decision making. For example, the New York Stock Exchange collects 1 terabyte of data in a single trading session, and having current data and real-time rules for trades and predictive modeling are important for managing stock portfolios.

Variety

In addition to the sheer volume and speed with which companies now collect data, more complicated types of data are now available and are proving to be of great value to businesses. Text data are collected by monitoring what is being said about a company's products or services on social media platforms such as Twitter. Audio data are collected from service calls (on a service call, you will often hear "this call may be monitored for quality control"). Video data collected by in-store video cameras are used to analyze shopping behavior. Analyzing information generated by these nontraditional sources is more complicated in part because of the processing required to transform the data into a numerical form that can be analyzed.

Veracity

Veracity has to do with how much uncertainty is in the data. For example, the data could have many missing values, which makes reliable analysis a challenge. Inconsistencies in units of measure and the lack of reliability of responses in terms of bias also increase the complexity of the data.

Businesses have realized that understanding big data can lead to a competitive advantage. Although big data represents opportunities, it also presents challenges in terms of data storage and processing, security, and available analytical talent.

The four Vs indicate that big data creates challenges in terms of how these complex data can be captured, stored, and processed; secured; and then analyzed. Traditional databases more or less assume that data fit into nice rows and columns, but that is not always the case with big data. Also, the sheer volume (the first V) often means that it is not possible to store all of the data on a single computer. This has led to new technologies like **Hadoop**—an open-source programming environment that supports big data processing through distributed storage and distributed processing on clusters of computers. Essentially, Hadoop provides a divide-and-conquer approach to handling massive amounts of data, dividing the storage and processing over multiple computers. **MapReduce** is a programming model used within Hadoop that performs the two major steps for which it is named: the map step and the reduce step. The map step divides the data into manageable subsets and distributes it to the computers in the cluster (often termed nodes) for storing and processing. The reduce step collects answers from the nodes and combines them into an answer to the original problem. Without technologies like Hadoop and MapReduce, and relatively inexpensive computer power, processing big data would not be cost-effective; in some cases, processing might not even be possible.

While some sources of big data are publicly available (Twitter, weather data, etc.), much of it is private information. Medical records, bank account information, and credit card transactions, for example, are all highly confidential and must be protected from computer hackers. **Data security**, the protection of stored data from destructive forces or unauthorized users, is of critical importance to companies. For example, credit card transactions are potentially very useful for understanding consumer behavior, but compromise of these data could lead to unauthorized use of the credit card or identity theft. A 2016 study of 383 companies