S. CHRISTIAN ALBRIGHT | WAYNE L. WINSTON

SIXTH EDITION

Business Analytics

Data Analysis and Decision Making

Copyright ©2017 Cengage Learning. All Rights Reserved.

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.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Overview of Applications in the Book, by Discipline

Accounting

Accounts receivable	285, 297
Auditing for price errors	329
Developing a flexible budget	537
Estimating total tax refunds	325
Estimating total taxable income	325
Overhead cost analysis	423, 437, 471, 490, 520, 524

Economics/Government

Demand and cost for electricity	461
Demand for desktops and laptops	402
Demand for French bread	481
Demand for heating oil	536
Demand for microwaves	182
Electricity pricing	736
Home and condo prices	78
Housing price structure	480
Presidential elections	19
Sales of new houses	566, 572

Finance

Bond investment strategy	893
Capital budgeting	714
Cash management	852
DJIA index	58, 77
Investing for college	892
Investing for retirement	481, 857
Investment strategy	703
Investor's after-tax profit	181
Land purchase decision	274
Liquidity risk management	829
Market return scenarios	152, 157
Mutual fund returns	171, 195
New car development	847
Pension fund management	708
Portfolio analysis	743
Random walk of stock prices	562
Stock hedging	757

Human Resources

Employee empowerment	389
Employee retention	361
Gender discrimination	450, 457, 514
Jobs in statistics and mathematics	897
Personnel testing	178
Productivity due to exercise	384

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Marketing

Catalog marketing	503, 508
Churn in cellular phone market	136
Clustering shoe customers	934
Customer complaints	349, 378
Customer valuation	865
DVD movie renters	310
Electronics sales	108
Frozen lasagna dinner buyers	125, 915, 919, 923
Furniture pricing	480
Marketing and selling condos	873
New pizza style decisions	365, 373
New product decisions	233, 240, 243, 260
Olympics sponsors	363
Response to new sandwich	319, 346, 348
Running shoe style decisions	274
Sales presentation ratings	339
Sales response to coupons	343
Sales versus promotions	421, 433
Soft-drink can style decisions	380
Supermarket sales	197
Supermarket transactions	27
Value of free maintenance agreement	868

Miscellaneous Statistical

Crime in the U.S.	54
Cruise ship entertainment	16
Election returns	200
Family income sampling	283
Forecasting U.S. population	557
IQ and bell curve	166
Predictors of successful movies	79, 482
Questionnaire responses	23
Relationship between smoking and drinking	82
Removing Vioxx from market	412
Sample size determination in legal case	279
Saving, spending, social climbing	136
Simpson's paradox	165
University admissions	360

Operations Management

Aggregate planning	693
Airline boarding strategies	759
Airline hub location decisions	729
Arrivals at bank	135
Automobile production	755
Battery lifetimes	191
Bidding for contracts	831
Blending oil	670
Developing Army helicopter component	276

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Developing electronic timing system 275
Delivery times at restaurant 361
Distribution of metal strip widths 396
Employee scheduling 663
Expensive watch production 219
Forecasting sales 551, 554, 559, 566, 572, 576, 581, 586
Inventory management 208
Learning curve for production 466
Manufacturing plastics operations 599
Ordering decisions 781, 784, 796, 806, 812, 815
Out-of-spec products 350
Overbooking at airlines 198
Product mix decisions 603, 631, 721
Production quantity decisions 827, 828
Production scheduling 641, 840
Production, inventory, distribution decisions 661
Quality control at paper company 179
Reliability of motors 336
Site selection of motor inns 417
Timing uncertainty in construction 144
Transportation, logistics decisions 677, 686
Variability in machine parts 333
Warranty costs 835

Sports/Gaming

Baseball salaries 31, 40, 46, 49, 88
Games at McDonald's 139
Golf stats on PGA tour 95
NCAA basketball tournament simulation 882
Revenue management at casino 539
Streak shooting in basketball 201
Wheel of fortune simulation 300
Winning at craps 879
Winning the lottery 220

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

6TH EDITION

Business Analytics: Data Analysis and Decision Making

S. Christian Albright
Kelly School of Business, Indiana University, Emeritus

Wayne L. Winston
Kelly School of Business, Indiana University

CENGAGE Learning®
Australia • Brazil • Mexico • Singapore • United Kingdom • United States

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Business Analytics: Data Analysis & Decision Making, Sixth Edition

S. Christian Albright and Wayne L. Winston

Vice President, General Manager, Social Science & Qualitative Business: Balraj Kalsi

Product Director: Mike Schenk

Product Manager: Aaron Arnsparger

Associate Content Developer: Brad Sullender

Product Assistant: Audrey Jacobs

Sr. Marketing Manager: Nathan Anderson

Manufacturing Planner: Ron J. Montgomery

Art and Cover Direction, Production Management, and Composition: Lumina Datamatics, Inc.

Cover Image: iStockphoto.com/derrrek

Intellectual Property

Analyst: Brittani Morgan

Project Manager: Nick Barrows

© 2017, 2015 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, 1-800-354-9706**

For permission to use material from this text or product, submit all requests online at **www.cengage.com/permissions**
Further permissions questions can be emailed to **permissionrequest@cengage.com**

Library of Congress Control Number: 2016934323

ISBN: 978-1-305-94754-2

Loose Leaf Edition ISBN: 978-1-305-86647-8

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage 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.com**

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

Printed in the United States of America
Print Number: 01 Print Year: 2016

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

To my wonderful wife Mary—my best friend and travel mate; to Sam, Lindsay, Teddy, and Archie; and to Bryn, our ball-playing Welsh corgi! *S.C.A*

To my wonderful family *W.L.W.*

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

About the Authors

© Cengage Learning

S. Christian Albright got his B.S. degree in Mathematics from Stanford in 1968 and his PhD in Operations Research from Stanford in 1972. He taught in the Operations & Decision Technologies Department in the Kelley School of Business at Indiana University (IU) for close to 40 years, before retiring from teaching in 2011. While at IU, he taught courses in management science, computer simulation, statistics, and computer programming to all levels of business students, including undergraduates, MBAs, and doctoral students. In addition, he taught simulation modeling at General Motors and Whirlpool, and he taught database analysis for the Army. He published over 20 articles in leading operations research journals in the area of applied probability, and he has authored the books *Statistics for Business and Economics, Practical Management Science, Spreadsheet Modeling and Applications, Data Analysis for Managers*, and *VBA for Modelers*. He worked for several years after "retirement" with the Palisade Corporation developing training materials for its software products, he has developed a commercial version of his Excel® tutorial, called ExcelNow!, and he continues to revise his textbooks.

On the personal side, Chris has been married for 44 years to his wonderful wife, Mary, who retired several years ago after teaching 7th grade English for 30 years. They have one son, Sam, who lives in Philadelphia with his wife Lindsay and their two sons, Teddy and Archer. Chris has many interests outside the academic area. They include activities with his family (especially traveling with Mary), going to cultural events at IU, power walking while listening to books on his iPod, and reading. And although he earns his livelihood from statistics and management science, his *real* passion is for playing classical piano music.

© Cengage Learning

Wayne L. Winston taught in the Operations & Decision Technologies Department in the Kelley School of Business at Indiana University for close to 40 before retiring a few years ago. Wayne received his B.S. degree in Mathematics from MIT and his PhD in Operations Research from Yale. He has written the successful textbooks *Operations Research: Applications and Algorithms, Mathematical Programming: Applications and Algorithms, Simulation Modeling Using @RISK, Practical Management Science, Data Analysis and Decision Making, Financial Models Using Simulation and Optimization*, and *Mathletics*. Wayne has published more than 20 articles in leading journals and has won many teaching awards, including the school-wide MBA award four times. He has taught classes at Microsoft, GM, Ford, Eli Lilly, Bristol-Myers Squibb, Arthur Andersen, Roche, PricewaterhouseCoopers, and NCR, and in "retirement," he is currently teaching several courses at the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

University of Houston. His current interest is showing how spreadsheet models can be used to solve business problems in all disciplines, particularly in finance and marketing.

Wayne enjoys swimming and basketball, and his passion for trivia won him an appearance several years ago on the television game show *Jeopardy!*, where he won two games. He is married to the lovely and talented Vivian. They have two children, Gregory and Jennifer.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Brief Contents

Preface xviii
1 Introduction to Business Analytics 1

Part 1 Exploring Data 17
2 Describing the Distribution of a Single Variable 19
3 Finding Relationships among Variables 79

Part 2 Probability and Decision Making Under Uncertainty 137
4 Probability and Probability Distributions 139
5 Normal, Binomial, Poisson, and Exponential Distributions 166
6 Decision Making under Uncertainty 222

Part 3 Statistical Inference 277
7 Sampling and Sampling Distributions 279
8 Confidence Interval Estimation 311
9 Hypothesis Testing 363

Part 4 Regression Analysis and Time Series Forecasting 415
10 Regression Analysis: Estimating Relationships 417
11 Regression Analysis: Statistical Inference 482
12 Time Series Analysis and Forecasting 539

Part 5 Optimization and Simulation Modeling 597
13 Introduction to Optimization Modeling 599
14 Optimization Models 661
15 Introduction to Simulation Modeling 759
16 Simulation Models 829

Part 6 Advanced Data Analysis 895
17 Data Mining 897
Introduction to Spreadsheet Modeling (only in MindTap)

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Part 7 Bonus Online Material* 18-1

18 Importing Data into Excel 18-3

19 Analysis of Variance and Experimental Design 19-1

20 Statistical Process Control 20-1

Appendix A Statistical Reporting A-1

*Bonus Online Material for this text can be found on the text companion website at cengagebrain.com.

References 943

Index 945

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Contents

Preface xviii

1 Introduction to Business Analytics 1
- 1-1 Introduction 3
- 1-2 Overview of the Book 4
  - 1-2a The Methods 4
  - 1-2b The Software 7
- 1-3 Modeling and Models 10
  - 1-3a Graphical Models 10
  - 1-3b Algebraic Models 11
  - 1-3c Spreadsheet Models 12
  - 1-3d A Seven-Step Modeling Process 13
- 1-4 Conclusion 15

PART 1 EXPLORING DATA 17

2 Describing the Distribution of a Single Variable 19
- 2-1 Introduction 21
- 2-2 Basic Concepts 22
  - 2-2a Populations and Samples 22
  - 2-2b Data Sets, Variables, and Observations 23
  - 2-2c Types of Data 24
- 2-3 Descriptive Measures for Categorical Variables 26
- 2-4 Descriptive Measures for Numerical Variables 30
  - 2-4a Numerical Summary Measures 31
  - 2-4b Numerical Summary Measures with StatTools 40
  - 2-4c Analysis ToolPak Add-In 45
  - 2-4d Charts for Numerical Variables 45
- 2-5 Time Series Data 54
- 2-6 Outliers and Missing Values 61
  - 2-6a Outliers 61
  - 2-6b Missing Values 61
- 2-7 Excel Tables for Filtering, Sorting, and Summarizing 63
- 2-8 Conclusion 71

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

3 Finding Relationships among Variables 79
3-1 Introduction 80
3-2 Relationships among Categorical Variables 82
3-3 Relationships among Categorical Variables and a Numerical Variable 86
3-3a Stacked and Unstacked Formats 86
3-4 Relationships among Numerical Variables 95
3-4a Scatterplots 95
3-4b Correlation and Covariance 101
3-5 Pivot Tables 108
3-6 Conclusion 131

PART 2 PROBABILITY AND DECISION MAKING UNDER UNCERTAINTY 137

4 Probability and Probability Distributions 139
4-1 Introduction 140
4-2 Probability Essentials 142
4-2a Rule of Complements 142
4-2b Addition Rule 142
4-2c Conditional Probability and the Multiplication Rule 143
4-2d Probabilistic Independence 146
4-2e Equally Likely Events 147
4-2f Subjective Versus Objective Probabilities 147
4-3 Probability Distribution of a Single Random Variable 150
4-3a Summary Measures of a Probability Distribution 151
4-3b Conditional Mean and Variance 154
4-4 Introduction to Simulation 156
4-5 Conclusion 160

5 Normal, Binomial, Poisson, and Exponential Distributions 166
5-1 Introduction 167
5-2 The Normal Distribution 168
5-2a Continuous Distributions and Density Functions 168
5-2b The Normal Density 169
5-2c Standardizing: Z-Values 170
5-2d Normal Tables and Z-Values 172
5-2e Normal Calculations in Excel 174
5-2f Empirical Rules Revisited 177
5-2g Weighted Sums of Normal Random Variables 177
5-3 Applications of the Normal Distribution 178

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

5-4 The Binomial Distribution 190
5-4a Mean and Standard Deviation of the Binomial Distribution 193
5-4b The Binomial Distribution in the Context of Sampling 193
5-4c The Normal Approximation to the Binomial 194
5-5 Applications of the Binomial Distribution 195
5-6 The Poisson and Exponential Distributions 207
5-6a The Poisson Distribution 207
5-6b The Exponential Distribution 210
5-7 Conclusion 212

6 Decision Making under Uncertainty 222
6-1 Introduction 223
6-2 Elements of Decision Analysis 225
6-2a Identifying the Problem 225
6-2b Possible Decisions 226
6-2c Possible Outcomes 226
6-2d Probabilities of Outcomes 226
6-2e Payoffs and Costs 227
6-2f Decision Criterion 227
6-2g More about the EMV Criterion 228
6-2h Decision Trees 230
6-3 One-Stage Decision Problems 232
6-4 The PrecisionTree Add-In 236
6-5 Multistage Decision Problems 239
6-6 The Role of Risk Aversion 257
6-6a Utility Functions 258
6-6b Exponential Utility 259
6-6c Certainty Equivalents 262
6-6d Is Expected Utility Maximization Used? 263
6-7 Conclusion 264

PART 3 STATISTICAL INFERENCE 277

7 Sampling and Sampling Distributions 279
7-1 Introduction 280
7-2 Sampling Terminology 280
7-3 Methods for Selecting Random Samples 282
7-3a Simple Random Sampling 282
7-3b Systematic Sampling 287
7-3c Stratified Sampling 288
7-3d Cluster Sampling 289
7-3e Multistage Sampling Schemes 290

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

7-4 Introduction to Estimation 292
7-4a Sources of Estimation Error 292
7-4b Key Terms in Sampling 293
7-4c Sampling Distribution of the Sample Mean 295
7-4d The Central Limit Theorem 299
7-4e Sample Size Selection 304
7-4f Summary of Key Ideas for Simple Random Sampling 305
7-5 Conclusion 307

8 Confidence Interval Estimation 311
8-1 Introduction 312
8-2 Sampling Distributions 314
8-2a The *t* Distribution 314
8-2b Other Sampling Distributions 317
8-3 Confidence Interval for a Mean 317
8-4 Confidence Interval for a Total 324
8-5 Confidence Interval for a Proportion 326
8-6 Confidence Interval for a Standard Deviation 331
8-7 Confidence Interval for the Difference between Means 335
8-7a Independent Samples 335
8-7b Paired Samples 339
8-8 Confidence Interval for the Difference between Proportions 342
8-9 Sample Size Selection 344
8-9a Sample Size Selection for Estimation of the Mean 345
8-9b Sample Size Selection for Estimation of Other Parameters 347
8-10 Conclusion 352

9 Hypothesis Testing 363
9-1 Introduction 364
9-2 Concepts in Hypothesis Testing 365
9-2a Null and Alternative Hypotheses 366
9-2b One-Tailed Versus Two-Tailed Tests 366
9-2c Types of Errors 367
9-2d Significance Level and Rejection Region 368
9-2e Significance from *p*-values 368
9-2f Type II Errors and Power 370
9-2g Hypothesis Tests and Confidence Intervals 371
9-2h Practical versus Statistical Significance 371
9-3 Hypothesis Tests for a Population Mean 372
9-4 Hypothesis Tests for Other Parameters 377
9-4a Hypothesis Tests for a Population Proportion 377
9-4b Hypothesis Tests for Differences between Population Means 379

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

9-4c Hypothesis Test for Equal Population Variances 387
9-4d Hypothesis Tests for Differences between Population Proportions 388
9-5 Tests for Normality 395
9-6 Chi-Square Test for Independence 401
9-7 Conclusion 406

PART 4 REGRESSION ANALYSIS AND TIME SERIES FORECASTING 415

10 Regression Analysis: Estimating Relationships 417
10-1 Introduction 418
10-2 Scatterplots: Graphing Relationships 421
10-2a Linear versus Nonlinear Relationships 426
10-2b Outliers 426
10-2c Unequal Variance 427
10-2d No Relationship 427
10-3 Correlations: Indicators of Linear Relationships 428
10-4 Simple Linear Regression 430
10-4a Least Squares Estimation 430
10-4b Standard Error of Estimate 438
10-4c The Percentage of Variation Explained: R-Square 440
10-5 Multiple Regression 443
10-5a Interpretation of Regression Coefficients 443
10-5b Interpretation of Standard Error of Estimate and R-Square 446
10-6 Modeling Possibilities 449
10-6a Dummy Variables 450
10-6b Interaction Variables 456
10-6c Nonlinear Transformations 460
10-7 Validation of the Fit 470
10-8 Conclusion 472

11 Regression Analysis: Statistical Inference 482
11-1 Introduction 484
11-2 The Statistical Model 484
11-3 Inferences about the Regression Coefficients 488
11-3a Sampling Distribution of the Regression Coefficients 489
11-3b Hypothesis Tests for the Regression Coefficients and *p*-Values 491
11-3c A Test for the Overall Fit: The ANOVA Table 492
11-4 Multicollinearity 496
11-5 Include/Exclude Decisions 502
11-6 Stepwise Regression 507
11-7 Outliers 512

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

11-8 Violations of Regression Assumptions 517
11-8a Nonconstant Error Variance 517
11-8b Nonnormality of Residuals 518
11-8c Autocorrelated Residuals 519
11-9 Prediction 521
11-10 Conclusion 527

12 Time Series Analysis and Forecasting 539
12-1 Introduction 540
12-2 Forecasting Methods: An Overview 541
12-2a Extrapolation Models 541
12-2b Econometric Models 542
12-2c Combining Forecasts 543
12-2d Components of Time Series Data 543
12-2e Measures of Accuracy 546
12-3 Testing for Randomness 548
12-3a The Runs Test 550
12-3b Autocorrelation 552
12-4 Regression-Based Trend Models 556
12-4a Linear Trend 556
12-4b Exponential Trend 559
12-5 The Random Walk Model 562
12-6 Moving Averages Forecasts 565
12-7 Exponential Smoothing Forecasts 570
12-7a Simple Exponential Smoothing 571
12-7b Holt's Model for Trend 575
12-8 Seasonal Models 580
12-8a Winters' Exponential Smoothing Model 581
12-8b Deseasonalizing: The Ratio-to-Moving-Averages Method 584
12-8c Estimating Seasonality with Regression 585
12-9 Conclusion 590

PART 5 Optimization And Simulation Modeling 597

13 Introduction to Optimization Modeling 599
13-1 Introduction 600
13-2 Introduction to Optimization 601
13-3 A Two-Variable Product Mix Model 602
13-4 Sensitivity Analysis 615
13-4a Solver's Sensitivity Report 616
13-4b SolverTable Add-In 619
13-4c Comparison of Solver's Sensitivity Report and SolverTable 626

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

13-5 Properties of Linear Models 626
13-6 Infeasibility and Unboundedness 629
13-7 A Larger Product Mix Model 631
13-8 A Multiperiod Production Model 640
13-9 A Comparison of Algebraic and Spreadsheet Models 649
13-10 A Decision Support System 750
13-11 Conclusion 652

14 Optimization Models 661
14-1 Introduction 662
14-2 Employee Scheduling Models 663
14-3 Blending Models 670
14-4 Logistics Models 676
14-4a Transportation Models 677
14-4b Other Logistics Models 685
14-5 Aggregate Planning Models 693
14-6 Financial Models 703
14-7 Integer Optimization Models 714
14-7a Capital Budgeting Models 714
14-7b Fixed-Cost Models 720
14-7c Set-Covering Models 729
14-8 Nonlinear Optimization Models 735
14-8a Basic Ideas of Nonlinear Optimization 735
14-8b Managerial Economics Models 736
14-8c Portfolio Optimization Models 740
14-9 Conclusion 749

15 Introduction to Simulation Modeling 759
15-1 Introduction 760
15-2 Probability Distributions for Input Variables 762
15-2a Types of Probability Distributions 763
15-2b Common Probability Distributions 766
15-2c Using @RISK to Explore Probability Distributions 770
15-3 Simulation and the Flaw of Averages 780
15-4 Simulation with Built-in Excel Tools 783
15-5 Introduction to @RISK 794
15-5a @RISK Features 795
15-5b Loading @RISK 795
15-5c @RISK Models with a Single Random Input Variable 796
15-5d Some Limitations of @RISK 806
15-5e @RISK Models with Several Random Input Variables 806

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

15-6 The Effects of Input Distributions on Results 811
15-6a Effect of the Shape of the Input Distribution(s) 812
15-6b Effect of Correlated Input Variables 815
15-7 Conclusion 820

16 Simulation Models 829
16-1 Introduction 831
16-2 Operations Models 831
16-2a Bidding for Contracts 831
16-2b Warranty Costs 835
16-2c Drug Production with Uncertain Yield 840
16-3 Financial Models 847
16-3a Financial Planning Models 847
16-3b Cash Balance Models 852
16-3c Investment Models 857
16-4 Marketing Models 864
16-4a Models of Customer Loyalty 864
16-4b Marketing and Sales Models 872
16-5 Simulating Games of Chance 879
16-5a Simulating the Game of Craps 879
16-5b Simulating the NCAA Basketball Tournament 882
16-6 Conclusion 885

PART 6 ADVANCED DATA ANALYSIS 895

17 Data Mining 897
17-1 Introduction 898
17-2 Data Exploration and Visualization 900
17-2a Introduction to Relational Databases 900
17-2b Online Analytical Processing (OLAP) 901
17-2c Power Pivot and Self-Service BI Tools in Excel 904
17-2d Visualization Software 911
17-3 Classification Methods 912
17-3a Logistic Regression 913
17-3b Neural Networks 918
17-3c Naïve Bayes 923
17-3d Classification Trees 926
17-3e Measures of Classification Accuracy 927
17-3f Classification with Rare Events 930
17-4 Clustering 933
17-5 Conclusion 938

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

18 Importing Data into Excel 18-3
18-1 Introduction 18-4
18-2 Rearranging Excel Data 18-5
18-3 Importing Text Data 18-9
18-4 Importing Data into Excel 18-15
18-4a Importing from Access with Old Tools 18-15
18-4b Importing from Access with Power Query 18-16
18-4c Using Microsoft Query 18-18
18-4d SQL Statements and M 18-26
18-4e Web Queries 18-26
18-5 Cleansing Data 18-28
18-6 Conclusion 18-35

19 Analysis of Variance and Experimental Design 19-1
19-1 Introduction 19-2
19-2 One-Way ANOVA 19-5
19-2a The Equal-Means Test 19-5
19-2b Confidence Intervals for Differences between Means 19-8
19-2c Using a Logarithmic Transformation 19-11
19-3 Using Regression to Perform ANOVA 19-17
19-4 The Multiple Comparison Problem 19-20
19-5 Two-Way ANOVA 19-24
19-5a Confidence Intervals for Contrasts 19-31
19-5b Assumptions of Two-Way ANOVA 19-34
19-6 More about Experimental Design 19-35
19-6a Randomization 19-36
19-6b Blocking 19-38
19-6c Incomplete Designs 19-42
19-7 Conclusion 19-45

20 Statistical Process Control 20-1
20-1 Introduction 20-3
20-2 Deming's 14 Points 20-4
20-3 Introduction to Control Charts 20-7
20-4 Control Charts for Variables 20-9
20-4a Control Charts and Hypothesis Testing 20-15
20-4b Other Out-of-Control Indications 20-16
20-4c Rational Subsamples 20-17
20-4d Deming's Funnel Experiment and Tampering 20-20
20-4e Control Charts in the Service Industry 20-23

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

20-5 Control Charts for Attributes 20-27
20-5a The *p* Chart 20-27
20-5b The Red Bead Experiment 20-31
20-6 Process Capability 20-34
20-6a Process Capability Indexes 20-37
20-6b More on Motorola and 6-sigma 20-42
20-7 Conclusion 20-45

Appendix A: Statistical Reporting A-1
A-1 Introduction A-1
A-2 Suggestions for Good Statistical Reporting A-2
A-2a Planning A-2
A-2b Developing a Report A-3
A-2c Be Clear A-4
A-2d Be Concise A-5
A-2e Be Precise A-5
A-3 Examples of Statistical Reports A-7
A-4 Conclusion A-18

References 943

Index 945

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Preface

With today's technology, companies are able to collect tremendous amounts of data with relative ease. Indeed, many companies now have more data than they can handle. However, the data are usually meaningless until they are analyzed for trends, patterns, relationships, and other useful information. This book illustrates in a practical way a variety of methods, from simple to complex, to help you analyze data sets and uncover important information. In many business contexts, data analysis is only the first step in the solution of a problem. Acting on the solution and the information it provides to make good decisions is a critical next step. Therefore, there is a heavy emphasis throughout this book on analytical methods that are useful in decision making. Again, the methods vary considerably, but the objective is always the same—to equip you with decision-making tools that you can apply in your business careers.

We recognize that the majority of students in this type of course are *not* majoring in a quantitative area. They are typically business majors in finance, marketing, operations management, or some other business discipline who will need to analyze data and make quantitative-based decisions in their jobs. We offer a hands-on, example-based approach and introduce fundamental concepts as they are needed. Our vehicle is spreadsheet software—specifically, Microsoft Excel®. This is a package that most students already know and will almost surely use in their careers. Our MBA students at Indiana University have been so turned on by the required course that is based on this book that *almost all* of them (mostly finance and marketing majors) have taken at least one of our follow-up elective courses in spreadsheet modeling. We are convinced that students see value in quantitative analysis when the course is taught in a practical and example-based approach.

Rationale for Writing This Book

Business Analytics: Data Analysis and Decision Making is different from the many fine textbooks written for statistics and management science. Our rationale for writing this book is based on four fundamental objectives.

- **Integrated coverage and applications.** The book provides a unified approach to business-related problems by integrating methods and applications that have been traditionally taught in separate courses, specifically statistics and management science.
- **Practical in approach.** The book emphasizes realistic business examples and the processes managers actually use to analyze business problems. The emphasis is *not* on abstract theory or computational methods.
- **Spreadsheet-based teaching.** The book provides students with the skills to analyze business problems with tools they have access to and will use in their careers. To this end, we have adopted Excel and commercial spreadsheet add-ins.
- **Latest tools.** This is not a static field. The software keeps changing, and even the mathematical algorithms behind the software continue to evolve. Each edition of this book has presented the most recent tools in Excel and the accompanying Excel add-ins, and the current edition is no exception.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Integrated Coverage and Applications

In the past, many business schools, including ours at Indiana University, have offered a required statistics course, a required decision-making course, and a required management science course—or some subset of these. A current trend, however, is to have only one required course that covers the basics of statistics, some regression analysis, some decision making under uncertainty, some linear programming, some simulation, and some advanced data analysis methods. Essentially, faculty in the quantitative area get one opportunity to teach all business students, so we attempt to cover a variety of useful quantitative methods. We are not necessarily arguing that this trend is ideal, but rather that it is a reflection of the reality at our university and, we suspect, at many others. After several years of teaching this course, we have found it to be a great opportunity to attract students to the subject and to more advanced study.

The book is also integrative in another important aspect. It not only integrates a number of analytical methods, but it also applies them to a wide variety of business problems—that is, it analyzes realistic examples from many business disciplines. We include examples, problems, and cases that deal with portfolio optimization, workforce scheduling, market share analysis, capital budgeting, new product analysis, and many others.

Practical in Approach

This book has been designed to be very example-based and practical. We strongly believe that students learn best by working through examples, and they appreciate the material most when the examples are realistic and interesting. Therefore, our approach in the book differs in two important ways from many competitors. First, there is just enough conceptual development to give students an understanding and appreciation for the issues raised in the examples. We often introduce important concepts, such as standard deviation as a measure of variability, in the context of examples rather than discussing them in the abstract. Our experience is that students gain greater intuition and understanding of the concepts and applications through this approach.

Second, we place virtually no emphasis on hand calculations. We believe it is more important for students to understand why they are conducting an analysis and what it means than to emphasize the tedious calculations associated with many analytical techniques. Therefore, we illustrate how powerful software can be used to create graphical and numerical outputs in a matter of seconds, freeing the rest of the time for in-depth interpretation of the results, sensitivity analysis, and alternative modeling approaches. In our own courses, we move directly into a discussion of examples, where we focus almost exclusively on interpretation and modeling issues, and we let the software perform the number crunching.

Spreadsheet-based Teaching

We are strongly committed to teaching spreadsheet-based, example-driven courses, regardless of whether the basic area is data analysis or management science. We have found tremendous enthusiasm for this approach, both from students and from faculty around the world who have used our books. Students learn and remember more, and they appreciate the material more. In addition, instructors typically enjoy teaching more, and they usually receive immediate reinforcement through better teaching evaluations. We were among the first to move to spreadsheet-based teaching about two decades ago, and we have never regretted the move.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

What We Hope to Accomplish in This Book

Condensing the ideas in the previous paragraphs, we hope to:

- continue to make quantitative courses attractive to a wide audience by making these topics real, accessible, and interesting;
- give students plenty of hands-on experience with real problems and challenge them to develop their intuition, logic, and problem-solving skills;
- expose students to real problems in many business disciplines and show them how these problems can be analyzed with quantitative methods; and
- develop spreadsheet skills, including experience with powerful spreadsheet add-ins, that add immediate value to students' other courses and for their future careers.

New in the Sixth Edition

There are several important changes in this edition.

- ***MindTap:*** Offered for the first time with this text, MindTap is a customizable digital course solution that includes an interactive eBook, auto-graded exercises from the textbook, author-created videos, flashcards, and more. MindTap includes all videos in support of the text, using StatTools or using JMP, as well as Excel solutions files for students to use for checking selected problems from the text (odd-numbered questions). MindTap also includes a chapter on Spreadsheet Modeling, which is not found in the print text, as an additional resource for faculty and students. For more information on MindTap, as well as ordering options, please contact your Cengage Learning consultant.
- ***Focus on Excel 2016:*** The newest version of Excel was released just in time for this book's revision, so all of the explanations and screenshots are based on this newest version. Except for cosmetic changes in the user interface, you will see almost no changes, and if you are still using Excel 2013 or a previous version, you shouldn't have any problems following along with this book. However, Excel 2016 does have some nice features that are included here, including histograms, box plots, and the "power" tools discussed in Chapters 17 and 18.
- ***Revised Chapter 6:*** The chapter on decision making under uncertainty has been totally rewritten. Now, a single "new product decisions" example is developed and extended throughout the chapter to promote continuity.
- ***BigPicture diagrams:*** In the optimization and simulation chapters, it has always been difficult for students to go from a verbal description of a problem to an eventual spreadsheet model. In this edition, we include "big picture" diagrams of the models that will hopefully act as a bridge from the verbal descriptions to the spreadsheet models. These diagrams have been created from the latest add-in in the Palisade DecisionTools Suite, the BigPicture add-in. Users of the book have access to BigPicture, just like @RISK and the other Palisade add-ins.
- ***Somewhat less reliance on StatTools:*** Although we continue to rely on the StatTools add-in for much of the statistical number crunching, especially in the regression and time series chapters, we rely on Excel formulas for the "easier" material in the confidence interval and hypothesis testing chapters, where Excel's functions are perfectly adequate and might even be more insightful. Nevertheless, we include many brief videos that walk you through the StatTools procedures. These videos can be found within the MindTap product that accompanies this text.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

- ***Inclusion of JMP:*** The book continues to use Excel and Excel add-ins as the primary tools for data analysis. However, the student edition of the statistical software package JMP from SAS can be bundled with the text for minimal cost for users of our book. We do not try to replicate JMP's fine and extensive online help for learning the software, but for many of the statistical examples in the book, we provide short videos showing how JMP can generate the results from Excel or StatTools. These videos can be found within the MindTap product that is available for this text. For ordering information on how to include JMP student edition with the book, please contact your Cengage learning consultant.
- ***Updated videos:*** As in the fifth edition, the materials for the book include over 50 videos, particularly for the models in the optimization and simulation chapters. These videos have been redone (and shortened). These videos can be found within the MindTap product that accompanies this text.
- ***Updated Problems:*** As in previous editions, there are some new and some updated problems. Again, we have included a file, essentially a database of problems, which is available to instructors. This file, **DADM 6e Problem Database.xlsx,** indicates the context of each of the problems, and it also shows the correspondence between problems in this edition and those in the previous edition.
- ***More Data Mining Content:*** Chapter 17 on data mining has been expanded. First, there is more coverage on Excel's newest "power" tools, especially Power Pivot, which are now included with Excel 2016 (at least in versions Professional Plus and higher). Second, there are detailed sections on the Naïve Bayes method for classification and an Excel-only method for clustering.

Software

This book is based entirely on Microsoft Excel, the spreadsheet package that has become the standard analytical tool in business. Excel is an extremely powerful package, and one of our goals is to convert *casual* users into *power* users who can take full advantage of its features. If you learn no more than this, you will be acquiring a valuable skill for the business world. However, Excel has some limitations. Therefore, this book includes several Excel add-ins that greatly enhance Excel's capabilities. As a group, these add-ins comprise what is arguably the most impressive assortment of spreadsheet-based software accompanying any book on the market.

DecisionTools® Suite Add-in

The textbook website for *Business Analytics: Data Analysis and Decision Making* provides a link to the powerful DecisionTools® Suite by Palisade Corporation. (The version available is compatible with Excel 2016 and previous versions of Excel.) This suite includes seven separate add-ins:

- **@RISK**, an add-in for simulation
- **StatTools**, an add-in for statistical data analysis
- **PrecisionTree**, a graphical-based add-in for creating and analyzing decision trees
- **TopRank**, an add-in for performing what-if analyses
- **NeuralTools®**, an add-in for estimating complex, nonlinear relationships
- **Evolver™**, an add-in for performing optimization (an alternative to Excel's Solver)
- **BigPicture**, a very smart drawing add-in, useful for depicting model relationships

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

We use @RISK and PrecisionTree extensively in the chapters on simulation and decision making under uncertainty, and we use StatTools extensively in the data analysis chapters. We also use BigPicture in the optimization and simulation chapters to provide a "bridge" between a problem statement and an eventual spreadsheet model.

Online access to the DecisionTools Suite, available with new copies of the book, is an academic version, slightly scaled down from the professional version that sells for hundreds of dollars and is used by many leading companies. It functions for two years when properly installed, and it puts only modest limitations on the size of data sets or models that can be analyzed.[1]

SolverTable Add-in

We also include SolverTable, a supplement to Excel's built-in Solver for optimization.[2] If you have ever had difficulty understanding Solver's sensitivity reports, you will appreciate SolverTable. It works like Excel's data tables, except that for each input (or pair of inputs), the add-in runs Solver and reports the *optimal* output values. SolverTable is used extensively in the optimization chapters.

Windows versus Mac

In our own courses, we have seen an increasing number of students using Macintosh laptops rather than Windows laptops. Fortunately, this is *not* a problem, and our students have followed along fine with their Macs. However, these students should be advised to use a Windows emulation program (Bootcamp or Parallels are good candidates), along with Office for *Windows*. Be aware that a few Excel features discussed in the book, such as pivot charts and Power Pivot, are *not* yet supported by Excel 2016 for the Mac, and the SolverTable and Palisade add-ins will *not* work in Excel 2016 for the Mac.

Potential Course Structures

Although we have used the book for our own required one-semester course, there is admittedly much more material than can be covered adequately in one semester. We have tried to make the book as modular as possible, allowing an instructor to cover, say, simulation before optimization or vice-versa, or to omit either of these topics. The one exception is statistics. Due to the natural progression of statistical topics, the basic topics in the early chapters should be covered before the more advanced topics (regression and time series analysis) in the later chapters. With this in mind, there are several possible ways to cover the topics.

- **One-semester Required Course, with No Statistics Prerequisite** (or where MBA students need a refresher for whatever statistics they learned previously): If data analysis is the primary focus of the course, then Chapters 2–5, 7–11, and possibly Chapter 17 should be covered. Depending on the time remaining, any of the topics in Chapters 6 (decision making under uncertainty), 12 (time series analysis), 13–14 (optimization), or 15–16 (simulation) can be covered in practically any order.
- **One-semester Required Course, with a Statistics Prerequisite**: Assuming that students know the basic elements of statistics (up through hypothesis testing), the

[1] Visit www.kelley.iu.edu/albrightbooks for specific details on these limitations.
[2] Although SolverTable is available on this textbook's website, it is also available for free from Albright's website, www.kelley.iu.edu/albrightbooks.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

material in Chapters 2–5 and 7–9 can be reviewed quickly, primarily to illustrate how Excel and add-ins can be used to do the number crunching. The instructor can then choose among any of the topics in Chapters 6, 10–11, 12, 13–14, or 15–16 (in practically any order), or to fill the remainder of the course.

- **Two-semester Required Sequence**: Given the luxury of spreading the topics over two semesters, the entire book, or at least most of it, can be covered. The statistics topics in Chapters 2–5 and 7–9 should be covered in chronological order before other statistical topics (regression and time series analysis), but the remaining chapters can be covered in practically any order.

Custom Publishing

Cengage Learning is dedicated to making the educational experience unique for all learners by creating custom materials that best suit your course needs. With Cengage Learning you can create a custom solution where you have the ability to choose your book's content, length, sequence, even the cover design. You may combine content from multiple Cengage Learning titles and add other materials, including your own original work, to create your ideal customized text. If you would like to learn more about our custom publishing services, please contact your Cengage Learning representative[3] or visit us at www.cengage.com/custom.

Instructor Supplements

Textbook Website: cengage.com/login

The companion website provides immediate access to an array of teaching resources—including data and solutions files for all of the Examples, Problems, and Cases in the book, Chapters 18–20 and Appendix A, Test Bank files, PowerPoint slides, and access to the DecisionTools® Suite by Palisade Corporation and the SolverTable add-in. You can easily download the instructor resources you need from the password-protected, instructor-only section of the site.

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
- deliver tests from your LMS, your classroom, or wherever you want

[3] Find your Learning Consultant at sites.cengage.com/repfinder.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Student Supplements

Textbook Website: www.cengagebrain.com

Every new student edition of this book comes with access to the *Business Analytics: Data Analysis and Decision Making, 6e* textbook website that links to the following files and tools:

- Excel files for the examples in the chapters (usually two versions of each—a template, or data-only version, and a finished version)
- Data files required for the Problems and Cases
- **Excel Tutorial for Windows.xlsx**, which contains a useful tutorial for getting up to speed in Excel (**Excel Tutorial for the Mac.xlsx** is also available)
- Chapters 18–20 and Appendix A
- DecisionTools® Suite software by Palisade Corporation (described earlier)
- SolverTable add-in

The resources listed above can be accessed through our MindTap learning system. For additional free resources go to www.cengagebrain .com, search by ISBN 9781305947542, click on the "Free Materials" tab, and select "Access Now." The resources you need will be listed both per chapter (by selecting a chapter from the drop-down list) and for the entire book (under Book Resources).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Acknowledgements

We are also grateful to many of the professionals who worked behind the scenes to make this book a success: Mike Schenk, Product Director; Sr. Product Team Manager, Joe Sabatino; Product Manager, Aaron Arnsparger; Associate Content Developer, Brad Sullender; Product Project Manager, Kailash Rawat; Marketing Manager, Nate Anderson, Marketing Coordinator, Eileen Corcoran; and Product Assistant, Audrey Jacobs.

We also extend our sincere appreciation to the reviewers who provided feedback on the authors' proposed changes that resulted in this sixth edition:

John Aloysius, *Walton College of Business, University of Arkansas*

Henry F. Ander, *Arizona State University*

Dr. Baabak Ashuri, *School of Building Construction, Georgia Institute of Technology*

James Behel, *Harding University*

Robert H. Burgess, *Scheller College of Business, Georgia Institute of Technology*

Paul Damien, *McCombs School of Business, University of Texas in Austin*

Parviz Ghandforoush, *Virginia Tech*

Betsy Greenberg, *University of Texas*

Anissa Harris, *Harding University*

Tim James, *Arizona State University*

Norman Johnson, *C.T. Bauer College of Business, University of Houston*

Shivraj Kanungo, *The George Washington University*

Miguel Lejeune, *The George Washington University*

José Lobo, *Arizona State University*

Stuart Low, *Arizona State University*

Lance Matheson, *Virginia Tech*

Patrick R. McMullen, *Wake Forest University*

Barbara A. Price, PhD, *Georgia Southern University*

Laura Wilson-Gentry, *University of Baltimore*

Toshiyuki Yuasa, *University of Houston*

S. Christian Albright

Wayne L. Winston

December 2015

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

The DecisionTools Suite

Upgrade Your Textbook Version of DecisionTools Suite

The Textbook Edition of Palisade Corporation's industry-leading DecisionTools Suite Industrial for Microsoft Excel is included with this textbook. This Textbook Edition is a two-year, time-locked version of the software that has been carefully calibrated to handle all of the appropriate exercises, examples, and cases in this textbook.

You've used the DecisionTools Suite in your course work. You've seen how Monte Carlo simulation, decision trees, and statistics can lead to better, smarter decisions. Now, take the power into the business world with your own, unlimited version of the DecisionTools Suite!

You'll receive:

@RISK risk analysis using Monte Carlo simulation

PrecisionTree decision analysis with decision trees

TopRank automated "what if" sensitivity analysis

NeuralTools predictive analysis with intelligent neural networks

StatTools forecasting and advanced statistics

RISKOptimizer optimization under uncertainty with Monte Carlo simulation

Evolver optimization with sophisticated algorithms

BigPicture diagramming and mind mapping in Excel

Mention this ad and textbook title and save 25% off the retail price of one copy of the commercial version of DecisionTools Suite Industrial Edition. Limit one copy per textbook.

PALISADE

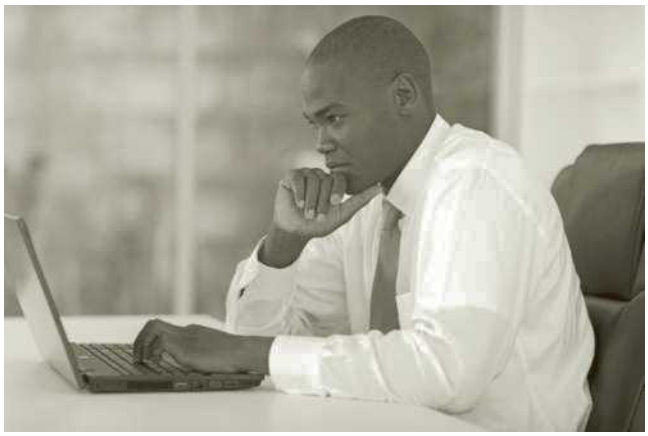
800-432-RISK
607-277-8000
sales@palisade.com
www.palisade.com

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

CHAPTER

1

Introduction to Business Analytics



michaeljung/Shutterstock.com

BUSINESS ANALYTICS PROVIDES INSIGHTS AND IMPROVES PERFORMANCE

This book is all about using quantitative modeling to help companies make better decisions and improve performance. We have been teaching management science for decades, and companies have been using the management science methods discussed in this book for decades to improve performance and save millions of dollars. Indeed, the applied journal *Interfaces*, discussed later in this chapter, has chronicled management science success stories for years. Therefore, we were a bit surprised when a brand new term, Business Analytics (BA), became hugely popular several years ago. All of a sudden, BA promised to be the road to success. By using quantitative BA methods—data analysis, optimization, simulation, prediction, and others—companies could drastically improve business performance. Haven't those of us in management science been doing this for years? What is different about BA that has made it so popular, both in the academic world and even more so in the business world?

The truth is that BA *does* use the same quantitative methods that have been the hallmark of management science for years, the same methods you will learn in this book. BA has not all of a sudden invented brand new quantitative methods to eclipse traditional management science methods. The main difference is that BA uses *big data* to solve business problems and provide insights. Companies now have access to huge sources of data, and the technology is now available to use huge data sets for statistical and quantitative analysis, predictive modeling, optimization, and simulation. In short, the same quantitative methods that have been available for years can now be even more effective by utilizing big data and the corresponding technology.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

For a quick introduction to BA, you should visit the BA Wikipedia site (search the Web for "business analytics"). Among other things, it lists areas where BA plays a prominent role, including the following: retail sales analytics; financial services analytics; risk and credit analytics; marketing analytics; pricing analytics; supply chain analytics; and transportation analytics. If you glance through the examples and problems in this book, you will see that most of them come from these same areas. Again, the difference is that we use relatively small data sets to get you started—we do not want to overwhelm you with gigabytes of data—whereas real applications of BA use huge data sets to advantage.

A more extensive discussion of BA can be found in the Fall 2011 research report, *Analytics: The Widening Divide*, published in the MIT Sloan Management Review in collaboration with IBM, a key developer of BA software (search the Web for the article's title). This 22-page article discusses what BA is and provides several case studies. In addition, it lists three key competencies people need to compete successfully in the BA world—and hopefully you will be one of these people.

- **Competency 1: Information management skills to manage the data.** This competency involves expertise in a variety of techniques for managing data. Given the key role of data in BA methods, data quality is extremely important. With data coming from a number of disparate sources, both internal and external to an organization, achieving data quality is no small feat.
- **Competency 2: Analytics skills and tools to understand the data.** We were not surprised, but rather very happy, to see this competency listed among the requirements because these skills are exactly the skills we cover throughout this book—optimization with advanced quantitative algorithms, simulation, and others.
- **Competency 3: Data-oriented culture to act on the data.** This refers to the culture within the organization. Everyone involved, especially top management, must believe strongly in fact-based decisions arrived at using analytical methods.

The article argues persuasively that the companies that have these competencies and have embraced BA have a distinct competitive advantage over companies that are just starting to use BA methods or are not using them at all. This explains the title of the article. The gap between companies that embrace BA and those that do not will only widen in the future.

One final note about the relationship between BA and management science is that, at the time this book was being revised (Winter 2014), a special issue of the journal *Management Science* was about to be published. The entire focus of this special issue is on BA. The following is an excerpt from the Call for Papers for this issue (search the Web for "management science business analytics special issue").

"We envision business analytics applied to many domains, including, but surely not limited to: digital market design and operation; network and social-graph analysis; pricing and revenue management; targeted marketing and customer relationship management; fraud and security; sports and entertainment; retailing to healthcare to financial services to many other industries. We seek novel modeling and empirical work which includes, among others, probability modeling, structural empirical models, and/or optimization methods."

This is even more confirmation of the tight relationship between BA and management science. As you study this book, you will see examples of most of the topics listed in this quote. ■

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

1-1 INTRODUCTION

We are living in the age of technology. This has two important implications for everyone entering the business world. First, technology has made it possible to collect huge amounts of data. Retailers collect point-of-sale data on products and customers every time a transaction occurs; credit agencies have all sorts of data on people who have or would like to obtain credit; investment companies have a limitless supply of data on the historical patterns of stocks, bonds, and other securities; and government agencies have data on economic trends, the environment, social welfare, consumer product safety, and virtually everything else imaginable. It has become relatively *easy* to collect the data. As a result, data are plentiful. However, as many organizations have discovered, it is quite a challenge to make sense of all the data they have collected.

A second important implication of technology is that it has given many more people the power and responsibility to analyze data and make decisions on the basis of quantitative analysis. People entering the business world can no longer pass all of the quantitative analysis to the "quant jocks," the technical specialists who have traditionally done the number crunching. The vast majority of employees now have a desktop or laptop computer at their disposal, access to relevant data, and training in easy-to-use software, particularly spreadsheet and database software. For these employees, statistics and other quantitative methods are no longer forgotten topics they once learned in college. Quantitative analysis is now an integral part of their daily jobs.

A large amount of data already exists, and it will only increase in the future. Many companies already complain of swimming in a sea of data. However, enlightened companies are seeing this expansion as a source of competitive advantage. In fact, one of the hottest topics in today's business world is **business analytics**, also called **data analytics**. These terms have been created to encompass all of the types of analysis discussed in this book, so they aren't really new; we have been teaching them for years. The new aspect of business analytics is that it typically implies the analysis of very *large* data sets, the kind that companies currently encounter. (For this reason, the term **big data** has also become popular.) By using quantitative methods to uncover the *information* in these data sets and then acting on this information—again guided by quantitative analysis—companies are able to gain advantages that their less enlightened competitors are not able to gain. Here are several pertinent examples.

- Direct marketers analyze enormous customer databases to see which customers are likely to respond to various products and types of promotions. Marketers can then target different classes of customers in different ways to maximize profits—and give their customers what they want.
- Hotels and airlines also analyze enormous customer databases to see what their customers want and are willing to pay for. By doing this, they have been able to devise very clever pricing strategies, where different customers pay different prices for the same accommodations. For example, a business traveler typically makes a plane reservation closer to the time of travel than a vacationer. The airlines know this. Therefore, they reserve seats for these business travelers and charge them a higher price for the same seats. The airlines profit from clever pricing strategies, and the customers are happy.
- Financial planning services have a virtually unlimited supply of data about security prices, and they have customers with widely differing preferences for various types of investments. Trying to find a match of investments to customers is a very challenging problem. However, customers can easily take their business elsewhere if good decisions are not made on their behalf. Therefore, financial planners are under extreme competitive pressure to analyze masses of data so that they can make informed decisions for their customers.[1]

[1]For a great overview of how quantitative techniques have been used in the financial world, read the book *The Quants*, by Scott Patterson. It describes how quantitative models made millions for a lot of bright young analysts, but it also describes the dangers of relying totally on quantitative models, at least in the complex world of global finance.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

- We all know about the pressures U.S. manufacturing companies have faced from foreign competition in the past couple of decades. The automobile companies, for example, have had to change the way they produce and market automobiles to stay in business. They have had to improve quality and cut costs by orders of magnitude. Although the struggle continues, much of the success they have had can be attributed to data analysis and wise decision making. Starting on the shop floor and moving up through the organization, these companies now measure almost everything, analyze these measurements, and then act on the results of their analysis.

We talk about companies analyzing data and making decisions. However, *companies* don't really do this; *people* do it. And who will these people be in the future? They will be *you*! We know from experience that students in all areas of business, at both the undergraduate and graduate level, will be *required* to describe large complex data sets, run regression analyses, make quantitative forecasts, create optimization models, and run simulations. You are the person who will be analyzing data and making important decisions to help your company gain a competitive advantage. And if you are *not* willing or able to do so, there will be plenty of other technically trained people who will be more than happy to replace you.

The goal of this book is to teach you how to use a variety of quantitative methods to analyze data and make decisions in a very hands-on way. We discuss a number of quantitative methods and illustrate their use in a large variety of realistic business situations. As you will see, this book includes many examples from finance, marketing, operations, accounting, and other areas of business. To analyze these examples, we take advantage of the Microsoft Excel® spreadsheet software, together with a number of powerful Excel add-ins. In each example we provide step-by-step details of the method and its implementation in Excel.

This is *not* a "theory" book. It is also not a book where you can lean comfortably back in your chair and read about how *other* people use quantitative methods. It is a "get your hands dirty" book, where you will learn best by actively following the examples throughout the book on your own computer. By the time you have finished, you will have acquired some very useful skills for today's business world.

1-2 OVERVIEW OF THE BOOK

This book is packed with quantitative methods and examples, probably more than can be covered in any single course. Therefore, we purposely intend to keep this introductory chapter brief so that you can get on with the analysis. Nevertheless, it is useful to introduce the methods you will be learning and the tools you will be using. This section provides an overview of the methods covered in this book and the software that is used to implement them. Then the next section presents a brief discussion of models and the modeling process. The primary goal at this point is to stimulate your interest in what follows.

1-2a The Methods

This book is rather unique in that it combines topics from two separate fields: statistics and management science. Statistics is the study of data analysis, whereas management science is the study of model building, optimization, and decision making. In the academic arena these two fields have traditionally been separated, sometimes widely. Indeed, they are often housed in separate academic departments. However, from a user's standpoint it makes little sense to separate them. Both are useful in accomplishing what the title of this book promises: data analysis and decision making.

Therefore, we do not distinguish between the statistics and the management science parts of this book. Instead, we view the entire book as a collection of useful quantitative methods that can be used to analyze data and help make business decisions. In addition, our

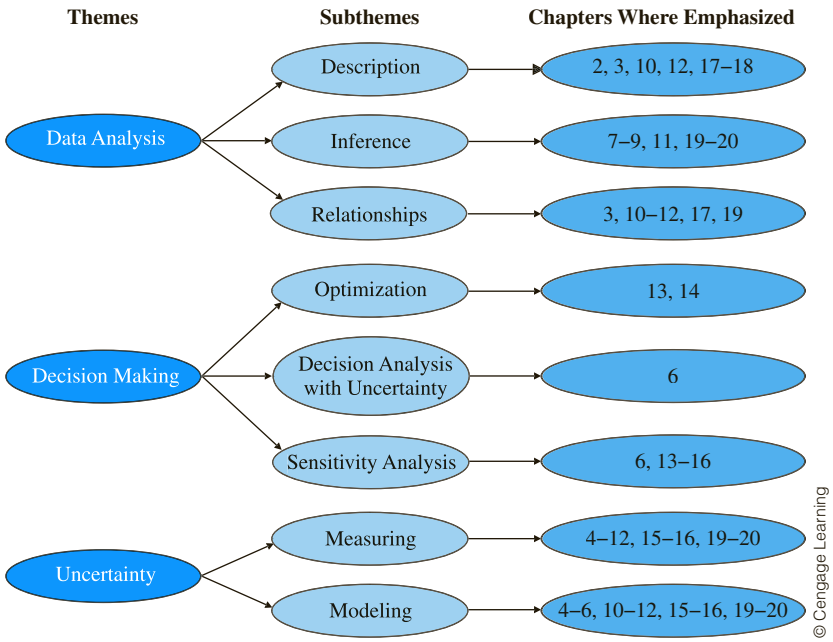
Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

choice of software helps to integrate the various topics. By using a single package, Excel, together with a number of add-ins, you will see that the methods of statistics and management science are similar in many important respects. Most importantly, their combination gives you the power and flexibility to solve a wide range of business problems.

Three important themes run through this book. Two of them are in the title: data analysis and decision making. The third is *dealing with uncertainty*.[2] Each of these themes has subthemes. Data analysis includes data description, data inference, and the search for relationships in data. Decision making includes *optimization* techniques for problems with no uncertainty, *decision analysis* for problems with uncertainty, and structured *sensitivity analysis*. Dealing with uncertainty includes measuring uncertainty and modeling uncertainty explicitly. There are obvious overlaps between these themes and subthemes. When you make inferences from data and search for relationships in data, you must deal with uncertainty. When you use *decision trees* to help make decisions, you must deal with uncertainty. When you use *simulation models* to help make decisions, you must deal with uncertainty, and then you often make inferences from the simulated data.

Figure 1.1 shows where these themes and subthemes are discussed in the book. The next few paragraphs discuss the book's contents in more detail.

Figure 1.1
Themes and Subthemes



We begin in Chapters 2 and 3 by illustrating a number of ways to summarize the information in data sets. These include graphical and tabular summaries, as well as numerical summary measures such as means, medians, and standard deviations. The material in these two chapters is elementary from a mathematical point of view, but it is extremely important. As stated earlier, organizations are now able to collect huge amounts of raw data, but what does it all mean? Although there are very sophisticated methods for analyzing data, some of which are covered in later chapters, the "simple" methods in Chapters 2 and 3 are crucial for obtaining an initial understanding of the data. Fortunately, Excel and available add-ins now make this quite easy. For example, Excel's pivot table tool for "slicing and dicing" data is an analyst's dream come true. You will be amazed by the insights you can gain from pivot tables—with very little effort.

[2]The fact that the uncertainty theme did not find its way into the title of this book does not detract from its importance. We just wanted to keep the title reasonably short.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

Uncertainty is a key aspect of most business problems. To deal with uncertainty, you need a basic understanding of probability. We discuss the key concepts in Chapters 4 and 5. Chapter 4 covers basic rules of probability and then discusses the extremely important concept of probability distributions. Chapter 5 follows up this discussion by focusing on two of the most important probability distributions, the normal and binomial distributions.

One of the best ways to make probabilistic concepts easier to understand is by using computer simulation. Therefore, simulation is a common theme that runs through this book, beginning in Chapter 4. Although Chapters 15 and 16 are devoted entirely to simulation, we do not hesitate to use simulation in earlier chapters to illustrate statistical concepts.

In Chapter 6 we apply probability to decision making under uncertainty. These types of problems—faced by all companies on a continual basis—are characterized by the need to make a decision *now,* even though important information, such as demand for a product or returns from investments, will not be known until later. The methods in Chapter 6 provide a rational basis for making such decisions. These methods do not guarantee perfect outcomes—the future could unluckily turn out differently than expected—but they do enable decision makers to proceed rationally and make the best of their current information. Additionally, the software used to implement these methods allows decision makers, with very little extra work, to see how sensitive the optimal decisions are to inputs. This is crucial, because the inputs to many business problems are, at best, educated guesses. Finally, we examine the role of risk aversion in these types of decision problems.

In Chapters 7, 8, and 9 we discuss sampling and statistical inference. Here the basic problem is to estimate one or more characteristics of a population. If it is too expensive or time-consuming to learn about the *entire* population—and it usually is—it is instead common to select a random sample from the population and then use the information in the sample to *infer* the characteristics of the population. You see this continually on news shows that describe the results of various polls. You also see it in many business contexts. For example, auditors typically sample only a fraction of a company's records. Then they infer the characteristics of the entire population of records from the results of the sample to conclude whether the company has been following acceptable accounting standards.

In Chapters 10 and 11 we discuss the extremely important topic of regression analysis, which is used to study relationships between variables. The power of regression analysis is its generality. Every part of a business has variables that are related to one another, and regression can often be used to estimate relationships between these variables. In managerial accounting, regression can be used to estimate how overhead costs depend on direct labor hours and production volume. In marketing, regression can be used to estimate how sales volume depends on advertising and other marketing variables. In finance, regression can be used to estimate how the return of a stock depends on the "market" return. In real estate studies, regression can be used to estimate how the selling price of a house depends on the assessed valuation of the house and characteristics such as the number of bedrooms and square footage.

From regression, we move to time series analysis and forecasting in Chapter 12. This topic is particularly important for providing inputs into business decision problems. For example, manufacturing companies must forecast demand for their products to make sensible decisions about order quantities from their suppliers. Similarly, fast-food restaurants must forecast customer arrivals, sometimes down to the level of 15-minute intervals, so that they can staff their restaurants appropriately. There are many approaches to forecasting, ranging from simple to complex. Some involve regression-based methods, whereas other methods are based on extrapolation. In an extrapolation method the historical patterns of a time series variable, such as product demand or customer arrivals, are studied carefully and are then *extrapolated* into the future to obtain forecasts. In Chapter 12 we discuss both regression and extrapolation methods for forecasting.

Chapters 13 and 14 are devoted to spreadsheet optimization. We assume a company must make several decisions, and there are constraints that limit the possible decisions. The

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

job of the decision maker is to choose the decisions such that all of the constraints are satisfied and an objective, such as total profit or total cost, is optimized. The solution process consists of two steps. The first step is to build a spreadsheet model that relates the decision variables to other relevant quantities by means of logical formulas. In this first step there is no attempt to find the *optimal* solution; its only purpose is to relate all relevant quantities in a logical way. The second step is then to find the optimal solution. Fortunately, Excel contains a Solver add-in that performs the optimization. All you need to do is specify the objective, the decision variables, and the constraints; Solver then uses powerful algorithms to find the optimal solution. As with regression, the power of this approach is its generality. An enormous variety of problems can be solved with spreadsheet optimization.

Chapters 15 and 16 illustrate a number of computer simulation models. This is not your first exposure to simulation—it is used in several previous chapters to illustrate statistical concepts—but here it is studied in its own right. As mentioned earlier, most business problems have some degree of uncertainty. The demand for a product is unknown, future interest rates are unknown, the delivery lead time from a supplier is unknown, and so on. Simulation allows you to build this uncertainty *explicitly* into spreadsheet models. Essentially, some cells in the model contain random values with given probability distributions. Every time the spreadsheet recalculates, these random values change, which causes "bottom-line" output cells to change as well. The trick then is to force the spreadsheet to recalculate many times and keep track of interesting outputs. In this way you can see an entire distribution of output values that might occur, not just a single best guess.

Spreadsheet simulations can be performed entirely with Excel's built-in tools, but this is quite tedious. Therefore, we use a spreadsheet add-in to streamline the process. In particular, you will learn how the @RISK add-in from Palisade Corporation can be used to run replications of a simulation, keep track of outputs, create useful charts, and perform sensitivity analyses. With the inherent power of spreadsheets and the ease of using add-ins such as @RISK, spreadsheet simulation is becoming one of the most popular quantitative tools in the business world.

Chapter 17 then returns to data analysis. It provides an introduction to data mining, a topic of increasing importance in today's data-driven world. Data mining is all about exploring data sets, especially large data sets, for relationships and patterns that can help companies gain a competitive advantage. It employs a number of relatively new technologies to implement various algorithms, several of which are discussed in this chapter.

Finally, there are three online chapters, 18–20, that complement topics included in the book itself. Chapter 18 discusses methods for importing data from a variety of data sources into Excel for analysis. Chapter 19 discusses analysis of variance (ANOVA) and experimental design. Chapter 20 discusses quality control and statistical process control. These three online chapters follow the same structure as the chapters in the book, complete with many examples and problems.

1-2b The Software

The quantitative methods in this book can be used to analyze a wide variety of business problems. However, they are not of much practical use unless you have the software to perform the number crunching. Very few business problems are small enough to be solved with pencil and paper. They require powerful software.

The software included in new copies of this book, together with Microsoft Excel, provides you with a powerful combination. This software is being used—and will continue to be used—by leading companies all over the world to analyze large, complex problems. We firmly believe that the experience you obtain with this software, through working the examples and problems in this book, will give you a key competitive advantage in the business world.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

It all begins with Excel. All of the quantitative methods that we discuss are implemented in Excel. Specifically, in this edition, we use Excel 2016.[3] We cannot forecast the state of computer software in the long-term future, but Excel is currently *the* most heavily used spreadsheet package on the market, and there is every reason to believe that this state will persist for many years. Most companies use Excel, most employees and most students have been trained in Excel, and Excel is a *very* powerful, flexible, and easy-to-use package.

Built-in Excel Features

Virtually everyone in the business world knows the basic features of Excel, but relatively few know some of its more powerful features. In short, relatively few people are the "power users" we expect you to become by working through this book. To get you started, the files **Excel Tutorial for Windows.xlsx** and **Excel Tutorial for the Mac.xlsx** explain some of the "intermediate" features of Excel—features that we expect you to be able to use. (See the Preface for instructions on how to access the resources that accompany this textbook.) These include the **SUMPRODUCT**, **VLOOKUP**, **IF**, **NPV**, and **COUNTIF**, functions. They also include range names, data tables, Paste Special, Goal Seek, and many others. Finally, although we assume you can perform routine spreadsheet tasks such as copying and pasting, the tutorial provides many tips to help you perform these tasks more efficiently.[4]

In the body of the book, we describe several of Excel's advanced features in more detail. For example, we provide a fairly extensive explanation of pivot tables in Chapter 3. This Excel tool enables you to summarize data sets in an almost endless variety of ways. As another example, we explain Excel's **RAND** and **RANDBETWEEN** functions for generating random numbers in Chapter 4. These functions are used in all spreadsheet simulations (at least those that do not take advantage of an add-in). In short, when an Excel tool is useful for a particular type of analysis, we usually provide step-by-step instructions on how to use it.

Solver Add-in

Chapters 13 and 14 make heavy use of Excel's Solver add-in. This add-in, developed by Frontline Systems®, not Microsoft, uses powerful algorithms to perform spreadsheet optimization. Before this type of spreadsheet optimization add-in was available, specialized (nonspreadsheet) software was required to solve optimization problems. Now you can do it all within the familiar Excel environment.

SolverTable Add-in

An important theme throughout this book is sensitivity analysis: How do outputs change when inputs change? Typically these changes are made in spreadsheets with a data table, a built-in Excel tool. However, data tables don't work in optimization models, where we want to see how the *optimal* solution changes when certain inputs change. Therefore, we include an Excel add-in called SolverTable, which works almost exactly like Excel's data tables. (This add-in was developed by Albright.) Chapters 13 and 14 illustrate the use of SolverTable.

DecisionTools Suite

In addition to SolverTable and built-in Excel add-ins, an educational version of Palisade Corporation's powerful DecisionTools® Suite is available. (See the Preface for instructions on how to access it.) All of the programs in this suite are Excel add-ins, so the learning curve isn't very steep. There are seven separate add-ins in this suite: @RISK, BigPicture,

[3]As you probably know, Microsoft made sweeping changes to Excel when it went from version 2003 to version 2007. The subsequent move to version 2010 introduced a few new changes, but nothing too dramatic, and the same is true of the two most recent versions, Excel 2013 and 2016. If you are using version 2007, 2010, or 2013 you will see a few differences in the screenshots (which are from version 2016), but you should be able to follow along without trouble.

[4]Albright and several colleagues have created a more robust commercial version of this tutorial called **ExcelNow!**. The Excel Tutorial files explain how you can upgrade to this commercial version at a very reasonable price.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203

StatTools, PrecisionTree, NeuralTools, TopRank, and Evolver.[5] We use only the first five in this book, but all are useful for certain tasks and are described briefly below.

@RISK

The simulation add-in @RISK enables you to run as many replications of a spreadsheet simulation as you like. As the simulation runs, @RISK automatically keeps track of the outputs you select, and it then displays the results in a number of tabular and graphical forms. @RISK also enables you to perform a sensitivity analysis, so that you can see which inputs have the most effect on the outputs. Finally, @RISK provides a number of spreadsheet functions that enable you to generate random numbers from a variety of probability distributions.

RISKOptimizer, part of @RISK, combines optimization with simulation. There are often times when you want to use simulation to model some business problem, but you also want to optimize a summary measure, such as a mean, of an output distribution. This optimization can be performed in a trial-and-error fashion, where you try a few values of the decision variable(s) and see which provides the best solution. However, RISKOptimizer provides a more automatic optimization procedure.

BigPicture

BigPicture is the newest addition to the Palisade suite. It is a smart drawing add-in that complements Excel's built-in drawing tools. The goal of BigPicture, at least for our purposes, is to represent the elements and relationships in a model so that you can better visualize the problem. We use it in most of the optimization and simulation examples as a "bridge" from the verbal description of a problem to the eventual spreadsheet model of the problem.

StatTools

Much of this book discusses basic statistical analysis. Fortunately, the Palisade suite includes a statistical add-in called StatTools. StatTools is powerful, easy to use, and capable of generating output quickly in an easily interpretable form.

PrecisionTree

The PrecisionTree add-in is used in Chapter 6 to analyze decision problems with uncertainty. The primary method for performing this type of analysis is to draw a decision tree. Decision trees are inherently graphical, and they have always been difficult to implement in spreadsheets, which are based on rows and columns. However, PrecisionTree does this in a very clever and intuitive way. Equally important, once the basic decision tree is built, you can use PrecisionTree to perform a sensitivity analysis on the model's inputs.

NeuralTools

Chapters 10 and 11 discuss how regression can be used to find a linear equation that quantifies the relationship between a dependent variable and one or more explanatory variables. Although linear regression is a powerful tool, it is not capable of quantifying all possible relationships. The NeuralTools add-in mimics the working of the human brain to find "neural networks" that quantify complex nonlinear relationships. It is used in Chapter 17 as one of several possible classification methods.

Evolver

Although we will not use it in this book, Evolver provides an alternative to Excel's built-in Solver add-in for optimization.

[5]The Palisade suite used to have another separate add-in, RISKOptimizer. Now RISKOptimizer is contained entirely in @RISK.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-203