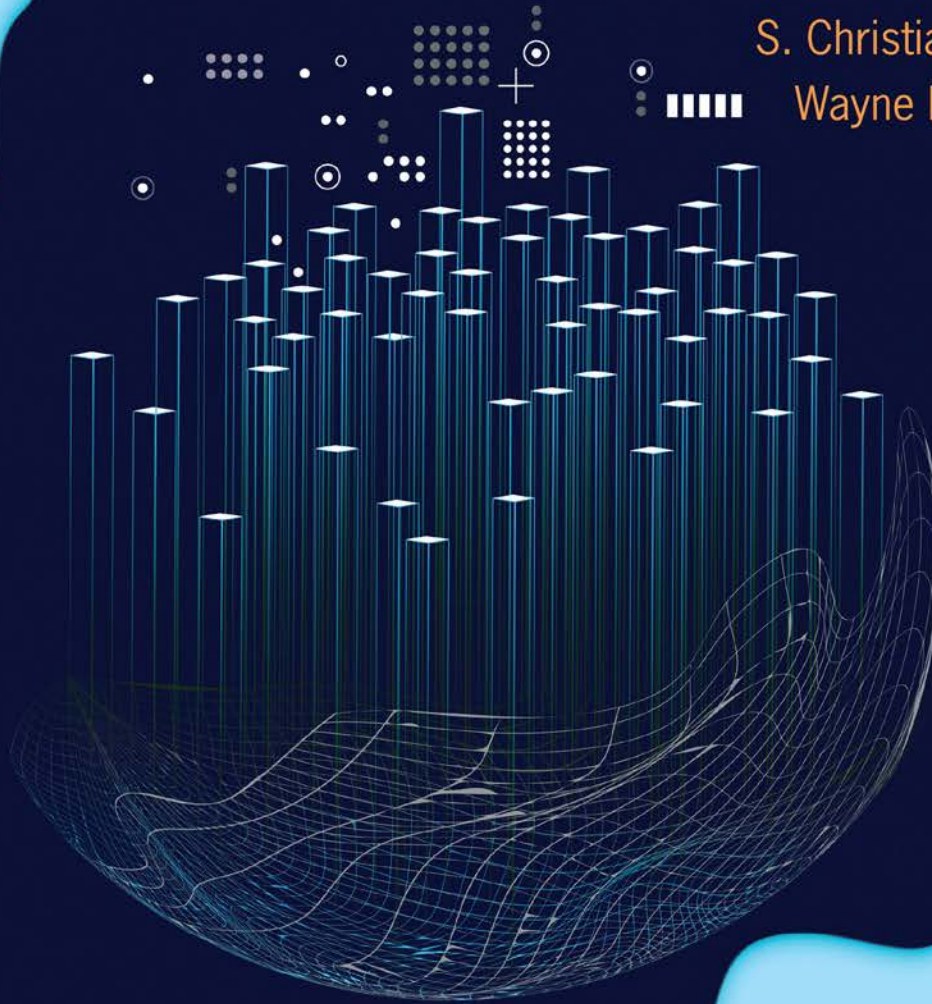


BUSINESS ANALYTICS

Data Analysis and Decision Making

SEVENTH EDITION

S. Christian Albright
Wayne L. Winston



BUSINESS ANALYTICS

Data Analysis and Decision Making

SEVENTH EDITION

S. Christian Albright

*Kelly School of Business,
Indiana University, Emeritus*

Wayne L. Winston

*Kelly School of Business,
Indiana University, Emeritus*



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

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: Data Analysis and
Decision Making, 7e***

S. Christian Albright and Wayne L. Winston

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

Product Director: Jason Fremder

Senior Product Manager: Aaron Arnsparger

Senior Learning Designer: Brandon Foltz

Content Manager: Conor Allen

Digital Delivery Lead: Mark Hopkinson

Product Assistant: Christian Wood

Marketing Manager: Chris Walz

Production Service: Lumina Datamatics Ltd.

Designer, Creative Studio: Erin Griffin

Text Designer: Erin Griffin

Cover Designer: Erin Griffin

Cover Image(s): Ron Dale/Shutterstock.com

Intellectual Property Analyst: Reba Frederics

Intellectual Property Project Manager: Nick
Barrows

© 2020, 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**.

Library of Congress Control Number: 2019935644

Student Edition:

ISBN: 978-0-357-10995-3

Loose-leaf Edition:

ISBN: 978-0-357-10996-0

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

*To my wonderful wife Mary—my best friend and companion; and to Sam, Lindsay,
Teddy, and Archie* S.C.A

To my wonderful family W.L.W.

ABOUT THE AUTHORS



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 47 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 Archie. Chris has many interests outside the academic area. They include activities with his family, traveling with Mary, going to cultural events, power walking while listening to books on his iPod, and reading. And although he earns his livelihood from quantitative methods, his *real* passion is for playing classical piano music.



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 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.

BRIEF CONTENTS

Preface xvi

1 Introduction to Business Analytics 1

PART 1 Data Analysis 37

- 2 Describing the Distribution of a Variable 38
- 3 Finding Relationships among Variables 84
- 4 Business Intelligence (BI) Tools for Data Analysis 132

PART 2 Probability and Decision Making under Uncertainty 183

- 5 Probability and Probability Distributions 184
- 6 Decision Making under Uncertainty 242

PART 3 Statistical Inference 293

- 7 Sampling and Sampling Distributions 294
- 8 Confidence Interval Estimation 323
- 9 Hypothesis Testing 368

PART 4 Regression Analysis and Time Series Forecasting 411

- 10 Regression Analysis: Estimating Relationships 412
- 11 Regression Analysis: Statistical Inference 472
- 12 Time Series Analysis and Forecasting 523

PART 5 Optimization and Simulation Modeling 575

- 13 Introduction to Optimization Modeling 576
- 14 Optimization Models 630
- 15 Introduction to Simulation Modeling 717
- 16 Simulation Models 779

PART 6 Advanced Data Analysis 837

- 17 Data Mining 838
- 18 Analysis of Variance and Experimental Design (MindTap Reader only)
- 19 Statistical Process Control (MindTap Reader only)
- APPENDIX A: Quantitative Reporting (MindTap Reader only)

References 873

Index 875



CONTENTS

Preface xvi

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 6

1-3 Introduction to Spreadsheet Modeling 8

1-3a Basic Spreadsheet Modeling: Concepts and Best Practices 9

1-3b Cost Projections 12

1-3c Breakeven Analysis 15

1-3d Ordering with Quantity Discounts and Demand Uncertainty 20

1-3e Estimating the Relationship between Price and Demand 24

1-3f Decisions Involving the Time Value of Money 29

1-4 Conclusion 33

PART 1 Data Analysis

37

2 Describing the Distribution of a Variable 38

2-1 Introduction 39

2-2 Basic Concepts 41

2-2a Populations and Samples 41

2-2b Data Sets, Variables, and Observations 41

2-2c Data Types 42

2-3 Summarizing Categorical Variables 45

2-4 Summarizing Numeric Variables 49

2-4a Numeric Summary Measures 49

2-4b Charts for Numeric Variables 57

2-5 Time Series Data 62

2-6 Outliers and Missing Values 69

2-7 Excel Tables for Filtering, Sorting, and Summarizing 71

2-8 Conclusion 77

Appendix: Introduction to StatTools 83

3 Finding Relationships among Variables 84

3-1 Introduction 85

3-2 Relationships among Categorical Variables 86

3-3 Relationships among Categorical Variables and a Numeric Variable	89
3-4 Relationships among Numeric Variables	96
3-4a Scatterplots	96
3-4b Correlation and Covariance	101
3-5 Pivot Tables	106
3-6 Conclusion	126
Appendix: Using StatTools to Find Relationships	131
4 Business Intelligence (BI) Tools for Data Analysis	132
4-1 Introduction	133
4-2 Importing Data into Excel with Power Query	134
4-2a Introduction to Relational Databases	134
4-2b Excel's Data Model	139
4-2c Creating and Editing Queries	146
4-3 Data Analysis with Power Pivot	152
4-3a Basing Pivot Tables on a Data Model	154
4-3b Calculated Columns, Measures, and the DAX Language	154
4-4 Data Visualization with Tableau Public	162
4-5 Data Cleansing	172
4-6 Conclusion	178
PART 2 Probability and Decision Making under Uncertainty	183
5 Probability and Probability Distributions	184
5-1 Introduction	185
5-2 Probability Essentials	186
5-2a Rule of Complements	187
5-2b Addition Rule	187
5-2c Conditional Probability and the Multiplication Rule	188
5-2d Probabilistic Independence	190
5-2e Equally Likely Events	191
5-2f Subjective Versus Objective Probabilities	192
5-3 Probability Distribution of a Random Variable	194
5-3a Summary Measures of a Probability Distribution	195
5-3b Conditional Mean and Variance	198
5-4 The Normal Distribution	200
5-4a Continuous Distributions and Density Functions	200
5-4b The Normal Density Function	201
5-4c Standardizing: Z-Values	202
5-4d Normal Tables and Z-Values	204

- 5-4e Normal Calculations in Excel 205
- 5-4f Empirical Rules Revisited 208
- 5-4g Weighted Sums of Normal Random Variables 208
- 5-4h Normal Distribution Examples 209

5-5 The Binomial Distribution 214

- 5-5a Mean and Standard Deviation of the Binomial Distribution 217
- 5-5b The Binomial Distribution in the Context of Sampling 217
- 5-5c The Normal Approximation to the Binomial 218
- 5-5d Binomial Distribution Examples 219

5-6 The Poisson and Exponential Distributions 226

- 5-6a The Poisson Distribution 227
- 5-6b The Exponential Distribution 229

5-7 Conclusion 231

6 Decision Making under Uncertainty 242

6-1 Introduction 243

6-2 Elements of Decision Analysis 244

6-3 EMV and Decision Trees 247

6-4 One-Stage Decision Problems 251

6-5 The PrecisionTree Add-In 254

6-6 Multistage Decision Problems 257

- 6.6a Bayes' Rule 262
- 6.6b The Value of Information 267
- 6.6c Sensitivity Analysis 270

6-7 The Role of Risk Aversion 274

- 6-7a Utility Functions 275
- 6-7b Exponential Utility 275
- 6-7c Certainty Equivalents 278
- 6-7d Is Expected Utility Maximization Used? 279

6-8 Conclusion 280

PART 3 Statistical Inference

293

7 Sampling and Sampling Distributions 294

7-1 Introduction 295

7-2 Sampling Terminology 295

7-3 Methods for Selecting Random Samples 297

- 7-3a Simple Random Sampling 297
- 7-3b Systematic Sampling 301
- 7-3c Stratified Sampling 301

7-3d Cluster Sampling	303
7-3e Multistage Sampling	303
7-4 Introduction to Estimation	305
7-4a Sources of Estimation Error	305
7-4b Key Terms in Sampling	306
7-4c Sampling Distribution of the Sample Mean	307
7-4d The Central Limit Theorem	312
7-4e Sample Size Selection	317
7-4f Summary of Key Ideas in Simple Random Sampling	318
7-5 Conclusion	320
8 Confidence Interval Estimation	323
8-1 Introduction	323
8-2 Sampling Distributions	325
8-2a The t Distribution	326
8-2b Other Sampling Distributions	327
8-3 Confidence Interval for a Mean	328
8-4 Confidence Interval for a Total	333
8-5 Confidence Interval for a Proportion	336
8-6 Confidence Interval for a Standard Deviation	340
8-7 Confidence Interval for the Difference between Means	343
8-7a Independent Samples	344
8-7b Paired Samples	346
8-8 Confidence Interval for the Difference between Proportions	348
8-9 Sample Size Selection	351
8-10 Conclusion	358
9 Hypothesis Testing	368
9-1 Introduction	369
9-2 Concepts in Hypothesis Testing	370
9-2a Null and Alternative Hypotheses	370
9-2b One-Tailed Versus Two-Tailed Tests	371
9-2c Types of Errors	372
9-2d Significance Level and Rejection Region	372
9-2e Significance from p -values	373
9-2f Type II Errors and Power	375
9-2g Hypothesis Tests and Confidence Intervals	375
9-2h Practical Versus Statistical Significance	375
9-3 Hypothesis Tests for a Population Mean	376
9-4 Hypothesis Tests for Other Parameters	380

- 9-4a Hypothesis Test for a Population Proportion 380
- 9-4b Hypothesis Tests for Difference between Population Means 382
- 9-4c Hypothesis Test for Equal Population Variances 388
- 9-4d Hypothesis Test for Difference between Population Proportions 388

9-5 Tests for Normality 395

9-6 Chi-Square Test for Independence 401

9-7 Conclusion 404

PART 4 Regression Analysis and Time Series Forecasting 411

10 Regression Analysis: Estimating Relationships 412

10-1 Introduction 413

10-2 Scatterplots: Graphing Relationships 415

10-3 Correlations: Indicators of Linear Relationships 422

10-4 Simple Linear Regression 424

10-4a Least Squares Estimation 424

10-4b Standard Error of Estimate 431

10-4c R-Square 432

10-5 Multiple Regression 435

10-5a Interpretation of Regression Coefficients 436

10-5b Interpretation of Standard Error of Estimate and R-Square 439

10-6 Modeling Possibilities 442

10-6a Dummy Variables 442

10-6b Interaction Variables 448

10-6c Nonlinear Transformations 452

10-7 Validation of the Fit 461

10-8 Conclusion 463

11 Regression Analysis: Statistical Inference 472

11-1 Introduction 473

11-2 The Statistical Model 474

11-3 Inferences About the Regression Coefficients 477

11-3a Sampling Distribution of the Regression Coefficients 478

11-3b Hypothesis Tests for the Regression Coefficients and p -Values 480

11-3c A Test for the Overall Fit: The ANOVA Table 481

11-4 Multicollinearity 485

11-5 Include/Exclude Decisions 489

11-6 Stepwise Regression 494

11-7 Outliers 499

11-8	Violations of Regression Assumptions	504
11-8a	Nonconstant Error Variance	504
11-8b	Nonnormality of Residuals	504
11-8c	Autocorrelated Residuals	505
11-9	Prediction	507
11-10	Conclusion	512
12	Time Series Analysis and Forecasting	523
12-1	Introduction	524
12-2	Forecasting Methods: An Overview	525
12-2a	Extrapolation Models	525
12-2b	Econometric Models	526
12-2c	Combining Forecasts	526
12-2d	Components of Time Series Data	527
12-2e	Measures of Accuracy	529
12-3	Testing for Randomness	531
12-3a	The Runs Test	534
12-3b	Autocorrelation	535
12-4	Regression-Based Trend Models	539
12-4a	Linear Trend	539
12-4b	Exponential Trend	541
12-5	The Random Walk Model	544
12-6	Moving Averages Forecasts	547
12-7	Exponential Smoothing Forecasts	551
12-7a	Simple Exponential Smoothing	552
12-7b	Holt's Model for Trend	556
12-8	Seasonal Models	560
12-8a	Winters' Exponential Smoothing Model	561
12-8b	Deseasonalizing: The Ratio-to-Moving-Averages Method	564
12-8c	Estimating Seasonality with Regression	565
12-9	Conclusion	569
PART 5	Optimization and Simulation Modeling	575
13	Introduction to Optimization Modeling	576
13-1	Introduction	577
13-2	Introduction to Optimization	577
13-3	A Two-Variable Product Mix Model	579

13-4 Sensitivity Analysis	590
13-4a Solver's Sensitivity Report	590
13-4b SolverTable Add-In	593
13-4c A Comparison of Solver's Sensitivity Report and SolverTable	599
13-5 Properties of Linear Models	600
13-6 Infeasibility and Unboundedness	602
13-7 A Larger Product Mix Model	604
13-8 A Multiperiod Production Model	612
13-9 A Comparison of Algebraic and Spreadsheet Models	619
13-10 A Decision Support System	620
13-11 Conclusion	622

14 Optimization Models 630

14-1 Introduction	631
14-2 Employee Scheduling Models	632
14-3 Blending Models	638
14-4 Logistics Models	644
14-4a Transportation Models	644
14-4b More General Logistics Models	651
14-5 Aggregate Planning Models	659
14-6 Financial Models	667
14-7 Integer Optimization Models	677
14-7a Capital Budgeting Models	678
14-7b Fixed-Cost Models	682
14-7c Set-Covering Models	689
14-8 Nonlinear Optimization Models	695
14-8a Difficult Issues in Nonlinear Optimization	695
14-8b Managerial Economics Models	696
14-8c Portfolio Optimization Models	700
14-9 Conclusion	708

15 Introduction to Simulation Modeling 717

15-1 Introduction	718
15-2 Probability Distributions for Input Variables	720
15-2a Types of Probability Distributions	721
15-2b Common Probability Distributions	724
15-2c Using @RISK to Explore Probability Distributions	728
15-3 Simulation and the Flaw of Averages	736
15-4 Simulation with Built-in Excel Tools	738

- 15-5 Simulation with @RISK 747**
 - 15-5a @RISK Features 748
 - 15-5b Loading @RISK 748
 - 15-5c @RISK Models with a Single Random Input 749
 - 15-5d Some Limitations of @RISK 758
 - 15-5e @RISK Models with Several Random Inputs 758
- 15-6 The Effects of Input Distributions on Results 763**
 - 15-6a Effect of the Shape of the Input Distribution(s) 763
 - 15-6b Effect of Correlated Inputs 766
- 15-7 Conclusion 771**

16 Simulation Models 779

- 16-1 Introduction 780**
- 16-2 Operations Models 780**
 - 16-2a Bidding for Contracts 780
 - 16-2b Warranty Costs 784
 - 16-2c Drug Production with Uncertain Yield 789
- 16-3 Financial Models 794**
 - 16-3a Financial Planning Models 795
 - 16-3b Cash Balance Models 799
 - 16-3c Investment Models 803
- 16-4 Marketing Models 810**
 - 16-4a Customer Loyalty Models 810
 - 16-4b Marketing and Sales Models 817
- 16-5 Simulating Games of Chance 823**
 - 16-5a Simulating the Game of Craps 823
 - 16-5b Simulating the NCAA Basketball Tournament 825
- 16-6 Conclusion 828**

PART 6 Advanced Data Analysis

837

17 Data Mining 838

- 17-1 Introduction 839**
- 17-2 Classification Methods 840**
 - 17-2a Logistic Regression 841
 - 17-2b Neural Networks 846
 - 17-2c Naïve Bayes 851
 - 17-2d Classification Trees 854
 - 17-2e Measures of Classification Accuracy 855
 - 17-2f Classification with Rare Events 857

17-3 Clustering Methods 860

17-4 Conclusion 870

18 Analysis of Variance and Experimental Design (MindTap Reader only)

18-1 Introduction 18-2

18-2 One-Way ANOVA 18-5

18-2a The Equal-Means Test 18-5

18-2b Confidence Intervals for Differences Between Means 18-7

18-2c Using a Logarithmic Transformation 18-11

18-3 Using Regression to Perform ANOVA 18-15

18-4 The Multiple Comparison Problem 18-18

18-5 Two-Way ANOVA 18-22

18-5a Confidence Intervals for Contrasts 18-28

18-5b Assumptions of Two-Way ANOVA 18-30

18-6 More About Experimental Design 18-32

18-6a Randomization 18-32

18-6b Blocking 18-35

18-6c Incomplete Designs 18-38

18-7 Conclusion 18-40

19 Statistical Process Control (MindTap Reader only)

19-1 Introduction 19-2

19-2 Deming's 14 Points 19-3

19-3 Introduction to Control Charts 19-6

19-4 Control Charts for Variables 19-8

19-4a Control Charts and Hypothesis Testing 19-13

19-4b Other Out-of-Control Indications 19-15

19-4c Rational Subsamples 19-16

19-4d Deming's Funnel Experiment and Tampering 19-18

19-4e Control Charts in the Service Industry 19-22

19-5 Control Charts for Attributes 19-26

19-5a *P* Charts 19-26

19-5b Deming's Red Bead Experiment 19-29

19-6 Process Capability 19-33

19-6a Process Capability Indexes 19-35

19-6b More on Motorola and 6-Sigma 19-40

19-7 Conclusion 19-43

APPENDIX A: Quantitative Reporting (MindTap Reader only)

A-1 Introduction A-1

A-2 Suggestions for Good Quantitative 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-4

A-2e Be Precise A-5

A-3 Examples of Quantitative Reports A-6

A-4 Conclusion A-16

References 873

Index 875



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, before the data can be useful, they must be analyzed for trends, patterns, and relationships. 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. 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 other 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.

Integrated Coverage and Applications

In the past, many business schools have offered a required statistics course, a required decision-making course, and a required management science course—or some subset of these. The 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 to interpret the results 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.

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.

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 Seventh Edition

There are several important changes in this edition.

- ***New introductory material on Excel:*** Chapter 1 now includes an introductory section on spreadsheet modeling. This provides business examples for getting students up to speed in Excel and covers such Excel tools as IF and VLOOKUP functions, data tables, goal seek, range names, and more.
- ***Reorganization of probability chapters:*** Chapter 4, Probability and Probability Distributions, and Chapter 5, Normal, Binomial, Poisson, and Exponential Distributions, have been shortened slightly and combined into a single Chapter 5, Probability and Probability Distributions. This created space for the new Chapter 4 discussed next.
- ***New material on “Power BI” tools and data visualization:*** The previous chapters on Data Mining and Importing Data into Excel have been reorganized and rewritten to include an increased focus on the tools commonly included under the Business Analytics umbrella. There is now a new Chapter 4, Business Intelligence Tools for Data Analysis, which includes Excel's Power Query tools for importing data into Excel, Excel's Power Pivot add-in (and the DAX language) for even more powerful data analysis with pivot tables, and Tableau Public for data visualization. The old online Chapter 18, Importing Data into Excel, has been eliminated, and its material has been moved to this new Chapter 4.

- **Updated for Office 365, Windows or Mac:** The 7th Edition is completely compatible with the latest version of Excel, and all screenshots in the book are from the latest version. However, because the changes from previous versions are not that extensive for Business Analytics purposes, the 7th Edition also works well even if you are still using Microsoft Office 2013, 2010, or 2007. Also, recognizing that many students are now using Macs, we have attempted to make the material compatible with Excel for Mac whenever possible.
- **Updated Problems:** Numerous problems have been modified to include the most updated data available. In addition, the **DADM 7e Problem Database.xlsx** file provides instructors with an entire database of problems. This file indicates the context of each of the problems and shows the correspondence between problems in this edition and problems in the previous edition.
- **Less emphasis on add-ins (when possible):** There is more emphasis in this edition on implementing spreadsheet calculations, especially statistical calculations, with built-in Excel tools rather than with add-ins. For example, there is no reliance on Palisade’s StatTools add-in in the descriptive statistics chapters 2 and 3 or in the confidence interval and hypothesis testing chapters 8 and 9. Nevertheless, Palisade’s add-ins are still relied on in chapters where they are really needed: PrecisionTree for decision trees in Chapter 6; StatTools for regression and time series analysis in Chapters 10, 11, and 12; @RISK for simulation in Chapters 15 and 16; and StatTools and NeuralTools for logistic regression and neural networks in Chapter 17.
- **New optional add-in:** Although it is not an “official” part of the book, Albright wrote a DADM_Tools add-in for Excel (Windows or Mac), with tools for creating summary stats, histograms, correlations and scatterplots, regression, time series analysis, decision trees, and simulation. This add-in provides a “lighter” alternative to the Palisade add-ins and is freely available at https://kelley.iu.edu/albrightbooks/free_downloads.htm.

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 relies on several Excel add-ins to 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. 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 smart drawing add-in, useful for depicting model relationships

We use @RISK and PrecisionTree extensively in the chapters on simulation and decision making under uncertainty, and we use StatTools as necessary 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 and for MindTap adopters, 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 one year when properly installed, and it puts only modest limitations on the size of data sets or models that can be analyzed.

SolverTable Add-in

We also include SolverTable, a supplement to Excel's built-in Solver for optimization.¹ 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

We have seen an increasing number of students using Macintosh laptops rather than Windows laptops. These students have two basic options when using our book. The first option is to use the latest version of Excel for Mac. Except for a few advanced tools such as Power Pivot (discussed in Chapter 4), the Mac version of Excel is very similar to the Windows version. However, the Palisade and SolverTable add-ins will not work with Excel for Mac. Therefore, the second option, the preferable option, is to use a Windows emulation program (Bootcamp and Parallels are good candidates), along with Office for Windows. Students at Indiana have used this second option for years and have had no problems.

Software Calculations by Chapter

This section indicates how the various calculations are implemented in the book. Specifically, it indicates which calculations are performed with built-in Excel tools and which require Excel add-ins.

Important note: The Palisade add-ins used in several chapters do *not* work in Excel for Mac. This is the primary reason Albright developed his own DADM_Tools add-in, which works in Excel for Windows *and* Excel for Mac. This add-in is freely available at the author's website (https://kelley.iu.edu/albrightbooks/free_downloads.htm), together with a Word document on how to use it. However, it is optional and is not used in the book.

Chapter 1 – Introduction to Business Analytics

- The section on basic spreadsheet modeling is implemented with built-in Excel functions.

Chapter 2 – Describing the Distribution of a Variable

- Everything is implemented with built-in Excel functions and charts.
 - Summary measures are calculated with built-in functions AVERAGE, STDEV.S, etc.
 - Histograms and box plots are created with the Excel chart types introduced in 2016.
 - Time series graphs are created with Excel line charts.
- Palisade's StatTools add-in can do all of this. It isn't used in the chapter, but it is mentioned in a short appendix, and an Intro to StatTools video is available.
- Albright's DADM_Tools add-in can do all of this except for time series graphs.

Chapter 3 – Finding Relationships among Variables

- Everything is implemented with built-in Excel functions and charts.
 - Summary measures of numeric variables, broken down by categories of a categorical variable, are calculated with built-in functions AVERAGE, STDEV.S, etc. (They are embedded in array formulas with IF functions.)
 - Side-by-side box plots are created with the Excel box plot chart type introduced in 2016.
 - Scatterplots are created with Excel scatter charts.
 - Correlations are calculated with Excel's CORREL function. A combination of the CORREL and INDIRECT functions is used to create tables of correlations.
- StatTools can do all of this. It isn't used in the chapter, but it is mentioned in a short appendix.
- DADM_Tools can do all of this.

¹ SolverTable is available on this textbook's website and on Albright's website, www.kelley.iu.edu/albrightbooks.

Chapter 4 - Business Intelligence (BI) Tools for Data Analysis

- Queries are implemented with Excel's Power Query tools (available only in Excel for Windows).
- Pivot table analysis is implemented with a combination of Excel's Data Model and the Power Pivot add-in (available only in Excel for Windows).
- Tableau Public (a free program for Windows or Mac) is used for data visualization.

Chapter 5 – Probability and Probability Distributions

- All calculations are performed with built-in Excel functions.

Chapter 6 – Decision Making Under Uncertainty

- Decision trees are implemented with Palisade's PrecisionTree add-in.
- DADM_Tools implements decision trees.

Chapter 7 – Sampling and Sampling Distributions

- All calculations are performed with built-in Excel functions.

Chapter 8 – Confidence Interval Estimation

- Everything is implemented with Excel functions, which are embedded in a **Confidence Interval Template.xlsx** file developed by Albright. This isn't an add-in; it is a regular Excel file where the confidence interval formulas have already been created, and users only need to enter their data.
- StatTools can do all of this, but it isn't used in the chapter.
- DADM_Tools doesn't perform confidence interval calculations.

Chapter 9 – Hypothesis Testing

- Everything is implemented with Excel functions, which are embedded in **Hypothesis Test Template.xlsx** and **Normality Tests Template.xlsx** files developed by Albright. These aren't add-ins; they are regular Excel files where the hypothesis test formulas have already been created, and users only need to enter their data.
- StatTools can do all of this, but it isn't used in the chapter.
- DADM_Tools doesn't perform hypothesis test calculations.

Chapters 10, 11 – Regression Analysis

- StatTools is used throughout to perform the regression calculations.
- Excel's built-in regression functions (SLOPE, INTERCEPT, etc.) are illustrated for simple linear regression, but they are used sparingly.
- Excel's Analysis ToolPak add-in is mentioned, but it isn't used.
- DADM_Tools implements regression calculations.

Chapter 12 – Time Series Analysis and Forecasting

- Autocorrelations and the runs test for randomness are implemented with Excel functions, which are embedded in **Autocorrelation Template.xlsx** and **Runs Test Template.xlsx** files developed by Albright. These aren't add-ins; they are regular Excel files where the relevant formulas have already been created, and users only need to enter their data.
- Moving averages and exponential smoothing are implemented in StatTools.
- DADM_Tools implements moving averages and exponential smoothing calculations.

Chapters 13, 14 – Optimization

- The optimization is performed with Excel’s Solver add-in.
- Albright’s SolverTable add-in performs sensitivity analysis on optimal solutions. It works in Excel for Windows but *not* in Excel for Mac.

Chapters 15, 16 – Simulation

- Built-in Excel functions for generating random numbers from various distributions are illustrated, and the Excel-only way of running simulations with data tables is shown in an introductory example.
- Palisade’s @RISK add-in is used in the rest of the examples.
- DADM_Tools implements simulation.

Chapter 17 – Data Mining

- Logistic regression is implemented with StatTools.
- Neural networks are implemented with Palisade’s NeuralTools add-in.
- Other calculations are implemented with Excel functions.

Chapter 18 – Analysis of Variance and Experimental Design (online)

- Most of the chapter is implemented with StatTools.
- DADM_Tools doesn’t perform ANOVA calculations.

Chapter 19 – Statistical Process Control

- All control charts are implemented with StatTools.
- DADM_Tools doesn’t implement control charts.

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 2–3, 5, 7–11 should be covered. Depending on the time remaining, any of the topics in Chapters 4 and 17 (advanced data analysis tools), 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 material in Chapters 2–3, 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 4, 6, 10–11, 12, 13–14, 15–16 (in practically any order), or 17 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 2–3 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.

Instructor Supplements

Textbook Website: www.cengage.com/decisionciences/albright/ba/7e

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, Test Bank files, PowerPoint slides, and access to the DecisionTools[®] Suite by Palisade Corporation and the SolverTable add-in. Instructors who want to compare the problems in the previous edition of this text to the problems in this edition can also download the file **DADM 7e Problem Database.xlsx** which details that correlation. 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 import, edit, and manipulate content from the text's *Test Bank* or elsewhere, including your own favorite test questions; create multiple test versions in an instant; and deliver tests from your LMS, your classroom, or wherever you want.

Student Supplements

Textbook Website: www.cengage.com/decisionciences/albright/ba/7e

Every new student edition of this book comes with access to the *Business Analytics: Data Analysis and Decision Making, 7e* 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)
- DecisionTools[®] Suite software by Palisade Corporation
- SolverTable add-in

Acknowledgements

We are also grateful to many of the professionals who worked behind the scenes to make this book a success: Aaron Arnsparger, Senior Product Manager; Brandon Foltz, Senior Learning Designer; Conor Allen, Content Manager; Project Manager, Anubhav Kaushal; and Marketing Manager, Chris Walz.

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

April 2019

CHAPTER 1

Introduction to Business Analytics



StockLite/Shutterstock.com

BUSINESS ANALYTICS PROVIDES INSIGHTS AND IMPROVES PERFORMANCE

This book is about analyzing data and using quantitative modeling to help companies understand their business, make better decisions, and improve performance. We have been teaching the methods discussed in this book for decades, and companies have been using these methods for decades to improve performance and save millions of dollars. 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 our field 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 used 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 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 quantitative analysis, predictive modeling, optimization, and simulation. In short, the same quantitative methods that have been used for years can now be even more effective by utilizing big data and the corresponding technology.

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.

This field of BA and, more specifically, data analysis is progressing very quickly, not only in academic areas but also in many applied areas. By analyzing big data sets from various sources, people are learning more and more about the way the world works. One recent book, *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, by Stephens-Davidowitz (2017), is especially interesting. The title of the book suggests that people lie to themselves, friends, Facebook, and surveys, but they don't lie on Google searches. The author, a social scientist, has focused on the data from Google searches to study wide-ranging topics, including economics, finance, education, sports, gambling, racism, and sex. Besides his insightful and often unexpected findings, the author presents several general conclusions about analysis with big data:

- Not surprisingly, insights are more likely to be obtained when analyzing data in novel ways, such as analyzing the organ sizes of race horses instead of their pedigrees. (A data analyst used data on race horse heart size to predict that American Pharoah would be great. The horse went on to win the Triple Crown in 2015, the first horse to do so since 1978.) On the other hand, if all available data have been analyzed in every conceivable way, as has been done with stock market data, even analysis with big data is unlikely to produce breakthroughs.
- Huge data sets allow analysts to zoom in on smaller geographical areas or smaller subsets of a population for keener insights. For example, when analyzing why some areas of the U.S. produce more people who warrant an entry on Wikipedia, you might find that richer and more educated states have more success. However, with a data set of more than 150,000 Americans of Wikipedia fame, the author was able to pinpoint that a disproportionate number were born in college towns. Traditional small data sets don't allow drilling down to this level of detail.
- Companies are increasingly able to run controlled experiments, usually on the Web, to learn the causes of behavior. The new term for these experiments is *A/B testing*. The purpose of a test is to determine whether a person prefers A or B. For example, Google routinely runs A/B tests, showing one randomly selected group of customers one thing and a second randomly selected group another. The difference could be as slight as the presence or absence of an arrow link on a Web page. Google gets immediate results from such tests by studying the customers' clicking behavior. The company can then use the results of one test to lead to another follow-up test. A/B testing is evidently happening all the time, whether or not we're aware of it.
- A potential drawback of analyzing big data is dimensionality, where there are so many potential explanatory variables that one will almost surely look like a winner—even though it isn't. The author illustrates this by imagining daily coin flips of 1000 coins and tracking which of them comes up heads on days when the stock market is up. One of these coins, just by chance, is likely to correlate highly with the stock market, but this coin is hardly a useful predictor of *future* stock market movement.

The analysis of big data will not solve all the world's problems, and there are ethical issues about studying people's behavior—and acting on the findings—from Web-based data such as Google searches. However, the potential for new and important insights into human behavior is enormous.

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. Technology companies like Google and Amazon capture click data from websites, retailers collect point-of-sale data on products and customers every time a transaction occurs; credit agencies collect 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 a challenge to make sense of 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 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 computer at their disposal, access to relevant data, and training in easy-to-use software, particularly spreadsheet and database software. For these employees, quantitative methods are no longer forgotten topics they once learned in college. Quantitative analysis is now an essential 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 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* from the data and then act on this information—again guided by quantitative analysis—companies are able to gain advantages over their less enlightened competitors. 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

- We all know about the pressures U.S. manufacturing companies have faced from foreign competition in the past few 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 analyze 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 do it.

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 several 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 section provides an overview of the methods covered in this book and the software that is used to implement them. Then the rest of the chapter illustrates how some of Excel's basic tools can be used to solve quantitative problems.

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 area of data analysis, whereas management science is the area 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, 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 for analyzing data and helping to make business decisions. In addition, our

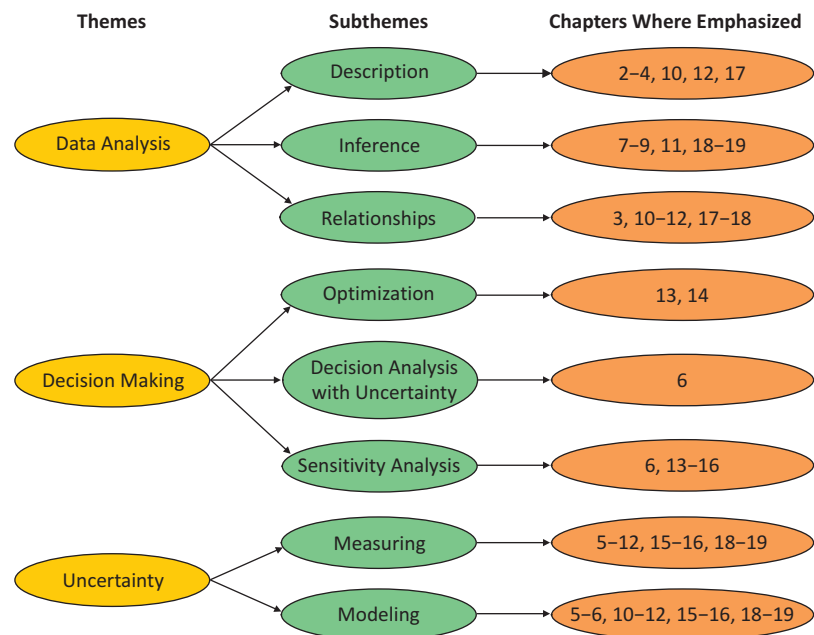
¹ For a great overview of how quantitative techniques have been used in the financial world, read the book *The Quants*, by Scott Patterson. It describes how quantitative models made millions for a lot of bright young analysts, but it also describes the dangers of relying totally on quantitative models,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

choice of software helps to integrate the various topics. By using a single package, Excel, together with several Excel add-ins, you will see that the methods of statistics and management science are similar in many important respects.

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 visualization, 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 simulation results.

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, 3, and 4 by illustrating a number of ways to summarize the information in data sets. These include graphical and tabular summaries, as well as numeric summary measures such as means, medians, and standard deviations. 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, 3, and 4 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 at the insights you can gain from pivot tables—with very little effort.

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 Chapter 5. This chapter covers basic rules of probability and then discusses the extremely important concept of probability distributions, with emphasis on two of the most important probability distributions, the normal and binomial distributions.

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.

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.

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.

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.

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. Chapter 12 illustrates some of most frequently used forecasting methods.

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 job of the decision maker is to choose the decisions such that all 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. 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.

Chapters 15 and 16 illustrate a number of computer simulation models. 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. 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.

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 two online chapters, 18 and 19, that complement topics included in the book itself. Chapter 18 discusses analysis of variance (ANOVA) and experimental design. Chapter 19 discusses quality control and statistical process control. These two online chapters follow the same structure as the chapters in the book, complete with many examples and problems.

1-2b The Software

The 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 implement them. Very few business problems are small enough to be solved with pencil and paper. They require powerful software.

The software available with 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 the book, will give you a key competitive advantage in the business world.

It all begins with Excel. Almost all the quantitative methods discussed in the book are implemented in Excel. 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.

It helps to understand Microsoft's versions of Excel. Until recently, we referred to Excel 2007, Excel 2010, and so on. Every few years, Microsoft released a new version of Office, and hence Excel. The latest was Excel 2016, and the next is going to be Excel 2019, which might be out by the time you read this. Each of these versions is basically fixed. For example, if you had Excel 2013, you had to wait for Excel 2016 to get the newest features. However, within the past few years, Microsoft has offered a subscription service, Office 365, which allows you to install free updates about once per month. Will Microsoft eventually discontinue the fixed versions and push everyone to the subscription service? We don't know, but it would be nice. Then we could assume that everyone had the same version of Excel, with all the latest features. Actually, this isn't quite true. There are different levels of Microsoft Office 365, some less expensive but with fewer features. You can find details on Microsoft's website. This book assumes you have the features in Excel 2016 or the ProPlus version of Office 365. You can see your version by opening Excel and selecting Account from the File menu.

We also realize that many of you are using Macs. There is indeed a version of Office 365 for the Mac, but its version of Excel doesn't yet have all the features of Excel for Windows. For example, we discuss two of Microsoft's "power" tools, Power Query and Power Pivot, in Chapter 4. At least as of this writing, these tools are not available in Excel for Mac. On the other hand, some features, including pivot charts, histograms, and box plots, were not in Office 365 for Mac until recently, when they suddenly appeared. So Microsoft is evidently catching up. In addition, some third-party add-ins for Excel, notably the Palisade DecisionTools Suite used in some chapters, are not available for Macs and probably never will be. If you really need these features and want to continue using your Mac, your best option is to install Windows emulation software such as Boot Camp or Parallels. Our students at Indiana have been doing this successfully for years.

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.³

In addition to this tutorial, the last half of this chapter presents several examples of modeling quantitative problems in Excel. These examples provide a head start in using Excel tools you will use in the rest of the book. Then later chapters will introduce other useful Excel tools as they are needed.

³ 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.

Analysis ToolPak

All versions of Excel, extending back at least two decades, have included an add-in called Analysis ToolPak. It has several tools for data analysis, including correlation, regression, and inference. Unfortunately, Microsoft hasn't updated Analysis ToolPak for years, so it is not as good as it could be. We will mention Analysis ToolPak a few times throughout the book, but we will use other Excel tools whenever they are available.

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 the goal is 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 for Excel for Windows by Albright. Unfortunately, it doesn't work for Excel for Mac.) Chapters 13 and 14 illustrate the use of SolverTable.

Palisade DecisionTools Suite

In addition to SolverTable and built-in Excel add-ins, an educational version of Palisade Corporation's powerful DecisionTools® Suite is available. All 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, StatTools, PrecisionTree, NeuralTools, TopRank, and Evolver. The add-ins we will use most often are StatTools (for data analysis), PrecisionTree (for decision trees), and @RISK (for simulation). These add-ins will be discussed in some detail in the chapters where they are used.

DADM_Tools Add-In

We realize that some of you prefer not to use the Palisade software because it might not be available in companies where your students are eventually employed. Nevertheless, some of the methods discussed in the book, particularly decision trees and simulation, are difficult to implement with Excel tools only. Therefore, Albright recently developed an add-in called DADM_Tools that implements decision trees and simulation, as well as forecasting and several basic data analysis tools. This add-in is freely available from the author's website at https://kelley.iu.edu/albrightbooks/free_downloads.htm, and students can continue to use it after the course is over, even in their eventual jobs, for free. You can decide whether you want to use the Palisade software, the DADM_Tools add-in, or neither.

1-3 Introduction to Spreadsheet Modeling⁴

A common theme in this book is spreadsheet modeling, where the essential elements of a business problem are entered and related in an Excel spreadsheet for further analysis. This section provides an introduction to spreadsheet modeling with some relatively simple models. Together with the Excel tutorial mentioned previously, the goal of this section is to get you "up to speed" in using Excel effectively for the rest of the book.

⁴ If you (or your students) are already proficient in basic Excel tools, you can skip this section, which is new to this edition of the book.

1-3a Basic Spreadsheet Modeling: Concepts and Best Practices

Most spreadsheet models involve *inputs*, *decision variables*, and *outputs*. The **inputs** have given fixed values, at least for the purposes of the model. The **decision variables** are those a decision maker controls. The **outputs** are the ultimate values of interest; they are determined by the inputs and the decision variables.

Spreadsheet modeling is the process of entering the inputs and decision variables into a spreadsheet and then relating them appropriately, by means of formulas, to obtain the outputs. After you have done this, you can then proceed in several directions. You might want to perform a sensitivity analysis to see how one or more outputs change as selected inputs or decision variables change. You might want to find the values of the decision variable(s) that minimize or maximize a particular output, possibly subject to certain constraints. You might also want to create charts that show graphically how certain parameters of the model are related.

Getting all the spreadsheet logic correct and producing useful results is a big part of the battle. However, it is also important to use good spreadsheet modeling *practices*. You probably won't be developing spreadsheet models for your sole use; instead, you will be sharing them with colleagues or even a boss (or an instructor). The point is that other people will be reading and trying to make sense out of your spreadsheet models. Therefore, you should construct your spreadsheet models with *readability* in mind. Features that can improve readability include the following:⁵

- A clear, logical layout to the overall model
- Separation of different parts of a model, possibly across multiple worksheets
- Clear headings for different sections of the model and for all inputs, decision variables, and outputs
- Use of range names
- Use of boldface, italics, larger font size, coloring, indentation, and other formatting features
- Use of cell comments
- Use of text boxes for assumptions and explanations

The following example illustrates the process of building a spreadsheet model according to these guidelines. We build this model in stages. In the first stage, we build a model that is correct, but not very readable. At each subsequent stage, we modify the model to enhance its readability.

EXAMPLE

1.1 ORDERING NCAA T-SHIRTS

It is March, and the annual NCAA Basketball Tournament is down to the final four teams. Randy Kitchell is a T-shirt vendor who plans to order T-shirts with the names of the final four teams from a manufacturer and then sell them to the fans. The fixed cost of any order is \$750, the variable cost per T-shirt to Randy is \$8, and Randy's selling price is \$18. However, this price will be charged only until a week after the tournament. After that time, Randy figures that interest in the T-shirts will be low, so he plans to sell all remaining T-shirts, if any, at \$6 each. His best guess is that demand for the T-shirts during the full-price period will be 1500. He is thinking about ordering 1450 T-shirts, but he wants to build a spreadsheet model that will let him experiment with the uncertain demand and his order quantity. How should he proceed?

Objective

To build a spreadsheet model in a series of stages, with all stages being correct but each stage being more readable and flexible than the previous stages.

⁵ For further guidelines that attempt to make spreadsheet models more flexible and less prone to errors, see the article by LeBlanc et al. (2018).