BUSINESS ANALYTICS

THE ART OF MODELING WITH SPREADSHEETS

STEPHEN G. POWELL
KENNETH R. BAKER

FIFTH EDITION

WILEY

Fifth Edition

BUSINESS ANALYTICS
The Art of Modeling with Spreadsheets

STEPHEN G. POWELL
Dartmouth College

KENNETH R. BAKER
Dartmouth College

WILEY

EXECUTIVE EDITOR	Lise Johnson
SPONSORING EDITOR	Jennifer Manias
PROJECT MANAGER	Gladys Soto
PROJECT SPECIALIST	Nichole Urban
CONTENT MANAGEMENT DIRECTOR	Lisa Wojcik
SENIOR CONTENT SPECIALIST	Nicole Repasky
PRODUCTION EDITOR	Ezhilan Vikraman
COVER PHOTO CREDIT	© traffic_analyzer/iStockphoto

This book was set in 10/12 TimesTenLTStd by SPi Global. Printed and bound by Strategic Content Imaging.

This book is printed on acid free paper. ∞

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.

Copyright © 2017, 2014, 2011, 2009, 2007 John Wiley & Sons, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923 (Web site: www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201) 748-6011, fax (201) 748-6008, or online at: www.wiley.com/go/permissions.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at: www.wiley.com/go/returnlabel. If you have chosen to adopt this textbook for use in your course, please accept this book as your complimentary desk copy. Outside of the United States, please contact your local sales representative.

ISBN: 978-1-119-29842-7 (PBK)
ISBN: 978-1-119-29840-3 (EVALC)

Library of Congress Cataloging-in-Publication Data:

Names: Powell, Stephen G., author. | Baker, Kenneth R., 1943- author. | Powell, Stephen G. Art of modeling with spreadsheets.
Title: Business analytics : the art of modeling with spreadsheets / Stephen G. Powell, Kenneth R. Baker.
Other titles: Management science.
Description: Fifth Edition. | Hoboken : Wiley, 2016. | Revised edition of the authors' Management science, [2014] | Includes index.
Identifiers: LCCN 2016032593| ISBN 9781119298427 (pbk.) | ISBN 9781119298311 (Adobe PDF) | ISBN 9781119298335 (epub)
Subjects: LCSH: Business—Computer simulation. | Electronic spreadsheets.
Classification: LCC HF5548.2 .P654 2016 | DDC 650.0285/554—dc23 LC record available at https://lccn.loc.gov/2016032593

The inside back cover will contain printing identification and country of origin if omitted from this page. In addition, if the ISBN on the back cover differs from the ISBN on this page, the one on the back cover is correct.

Printed in the United States of America

To Becky and Judy
for all their encouragement and support

Brief Contents

CHAPTER 1 *INTRODUCTION* 1

CHAPTER 2 *MODELING IN A PROBLEM-SOLVING FRAMEWORK* 15

CHAPTER 3 *SPREADSHEET ENGINEERING* 49

CHAPTER 4 *ANALYSIS USING SPREADSHEETS* 71

CHAPTER 5 *DATA EXPLORATION AND PREPARATION* 89

CHAPTER 6 *CLASSIFICATION AND PREDICTION METHODS* 117

CHAPTER 7 *SHORT-TERM FORECASTING* 187

CHAPTER 8 *NONLINEAR OPTIMIZATION* 207

CHAPTER 9 *LINEAR OPTIMIZATION* 239

CHAPTER 10 *OPTIMIZATION OF NETWORK MODELS* 277

CHAPTER 11 *INTEGER OPTIMIZATION* 309

CHAPTER 12 *OPTIMIZATION OF NONSMOOTH MODELS* 335

CHAPTER 13 *DECISION ANALYSIS* 357

CHAPTER 14 *MONTE CARLO SIMULATION* 383

CHAPTER 15 *OPTIMIZATION IN SIMULATION* 435

MODELING CASES 463

APPENDICES

APPENDIX 1 479

APPENDIX 2 507

APPENDIX 3 515

Table of Contents

PREFACE **xi**

ABOUT THE AUTHORS **xv**

CHAPTER 1 *INTRODUCTION* **1**

1.1 Models and Modeling **1**
- 1.1.1 Why Study Modeling? **2**
- 1.1.2 Models in Business **2**
- 1.1.3 Models in Business Education **3**
- 1.1.4 Benefits of Business Models **3**

1.2 The Role of Spreadsheets **4**
- 1.2.1 Risks of Spreadsheet Use **5**
- 1.2.2 Challenges for Spreadsheet Users **6**
- 1.2.3 Background Knowledge for Spreadsheet Modeling **7**

1.3 The Real World and the Model World **7**

1.4 Lessons from Expert and Novice Modelers **9**
- 1.4.1 Expert Modelers **9**
- 1.4.2 Novice Modelers **11**

1.5 Organization of the Book **12**

1.6 Summary **13**

Suggested Readings **14**

CHAPTER 2 *MODELING IN A PROBLEM-SOLVING FRAMEWORK* **15**

2.1 Introduction **15**

2.2 The Problem-Solving Process **16**
- 2.2.1 Some Key Terms **16**
- 2.2.2 The Six-Stage Problem-Solving Process **18**
- 2.2.3 Mental Models and Formal Models **23**

2.3 Influence Charts **24**
- 2.3.1 A First Example **25**
- 2.3.2 An Income Statement as an Influence Chart **27**
- 2.3.3 Principles for Building Influence Charts **27**
- 2.3.4 Two Additional Examples **28**

2.4 Craft Skills for Modeling **31**
- 2.4.1 Simplify the Problem **33**
- 2.4.2 Break the Problem into Modules **34**
- 2.4.3 Build a Prototype and Refine It **35**
- 2.4.4 Sketch Graphs of Key Relationships **38**
- 2.4.5 Identify Parameters and Perform Sensitivity Analysis **39**
- 2.4.6 Separate the Creation of Ideas from Their Evaluation **41**
- 2.4.7 Work Backward from the Desired Answer **42**
- 2.4.8 Focus on Model Structure, not on Data Collection **43**

2.5 Summary **45**

Suggested Readings **46**

Exercises **46**

CHAPTER 3 *SPREADSHEET ENGINEERING* **49**

3.1 Introduction **49**

3.2 Designing a Spreadsheet **51**
- 3.2.1 Sketch the Spreadsheet **51**
- 3.2.2 Organize the Spreadsheet into Modules **52**
- 3.2.3 Start Small **53**
- 3.2.4 Isolate Input Parameters **54**
- 3.2.5 Design for Use **54**
- 3.2.6 Keep It Simple **54**
- 3.2.7 Design for Communication **55**
- 3.2.8 Document Important Data and Formulas **55**

3.3 Designing a Workbook **57**
- 3.3.1 Use Separate Worksheets to Group Similar Kinds of Information **58**
- 3.3.2 Design Workbooks for Ease of Navigation and Use **59**
- 3.3.3 Design a Workbook as a Decision-Support System **60**

3.4 Building a Workbook **62**
- 3.4.1 Follow a Plan **62**
- 3.4.2 Build One Worksheet or Module at a Time **62**
- 3.4.3 Predict the Outcome of Each Formula **62**
- 3.4.4 Copy and Paste Formulas Carefully **62**
- 3.4.5 Use Relative and Absolute Addressing to Simplify Copying **62**
- 3.4.6 Use the Function Wizard to Ensure Correct Syntax **63**
- 3.4.7 Use Range Names to Make Formulas Easy to Read **63**
- 3.4.8 Choose Input Data to Make Errors Stand Out **64**

3.5 Testing a Workbook **64**
- 3.5.1 Check That Numerical Results Look Plausible **64**
- 3.5.2 Check That Formulas Are Correct **65**
- 3.5.3 Test That Model Performance Is Plausible **68**

3.6 Summary **68**

Suggested Readings **69**

Exercises **69**

CHAPTER 4 *ANALYSIS USING SPREADSHEETS* **71**

4.1 Introduction **71**

4.2 Base-case Analysis **72**

4.3 What-if Analysis **72**
- 4.3.1 Benchmarking **73**
- 4.3.2 Scenarios **74**
- 4.3.3 Parametric Sensitivity **77**
- 4.3.4 Tornado Charts **79**

4.4 Breakeven Analysis **81**

4.5 Optimization Analysis **83**

4.6 Simulation and Risk Analysis 84
4.7 Summary 85
Exercises 85

CHAPTER 5 *DATA EXPLORATION AND PREPARATION* 89

5.1 Introduction 89
5.2 Dataset Structure 90
5.3 Types of Data 93
5.4 Data Exploration 93
5.4.1 Understand the Data 94
5.4.2 Organize and Subset the Data 94
5.4.3 Examine Individual Variables Graphically 98
5.4.4 Calculate Summary Measures for Individual Variables 99
5.4.5 Examine Relationships among Variables Graphically 101
5.4.6 Examine Relationships among Variables Numerically 105
5.5 Data Preparation 109
5.5.1 Handling Missing Data 109
5.5.2 Handling Errors and Outliers 111
5.5.3 Binning Continuous Data 111
5.5.4 Transforming Categorical Data 111
5.5.5 Functional Transformations 112
5.5.6 Normalizations 113
5.6 Summary 113
Suggested Readings 114
Exercises 114

CHAPTER 6 *CLASSIFICATION AND PREDICTION METHODS* 117

6.1 Introduction 117
6.2 Preliminaries 117
6.2.1 The Data-Mining Process 118
6.2.2 The Problem of Overfitting 118
6.2.3 Partitioning the Dataset 120
6.2.4 Measures of Model Quality 120
6.2.5 Variable Selection 125
6.2.6 Setting the Cutoff in Classification 126
6.3 Classification and Prediction Trees 127
6.3.1 Classification Trees 128
6.3.2 An Application of Classification Trees 130
6.3.3 Prediction Trees 137
6.3.4 An Application of Prediction Trees 138
6.3.5 Ensembles of Trees 141
6.4 Additional Algorithms for Classification 143
6.4.1 Logistic Regression 144
6.4.2 Naïve Bayes 150
6.4.3 *k*-Nearest Neighbors 158
6.4.4 Neural Networks 162
6.5 Additional Algorithms for Prediction 169
6.5.1 Multiple Linear Regression 169
6.5.2 *k*-Nearest Neighbors 177
6.5.3 Neural Networks 178
6.6 Strengths and Weaknesses of Algorithms 181
6.7 Practical Advice 182
6.8 Summary 183
Suggested Readings 184
Exercises 184

CHAPTER 7 *SHORT-TERM FORECASTING* 187

7.1 Introduction 187
7.2 Forecasting with Time-Series Models 187
7.2.1 The Moving-Average Model 188
7.2.2 Measures of Forecast Accuracy 191
7.3 The Exponential Smoothing Model 192
7.4 Exponential Smoothing with a Trend 196
7.5 Exponential Smoothing with Trend and Cyclical Factors 198
7.6 Using XLMiner for Short-Term Forecasting 202
7.7 Summary 202
Suggested Readings 203
Exercises 203

CHAPTER 8 *NONLINEAR OPTIMIZATION* 207

8.1 Introduction 207
8.2 An Optimization Example 208
8.2.1 Optimizing Q1 208
8.2.2 Optimization Over All Four Quarters 210
8.2.3 Incorporating the Budget Constraint 211
8.3 Building Models for Solver 213
8.3.1 Formulation 213
8.3.2 Layout 214
8.3.3 Interpreting Results 215
8.4 Model Classification and the Nonlinear Solver 215
8.5 Nonlinear Programming Examples 217
8.5.1 Facility Location 217
8.5.2 Revenue Maximization 219
8.5.3 Curve Fitting 221
8.5.4 Economic Order Quantity 225
8.6 Sensitivity Analysis for Nonlinear Programs 227
8.7 The Portfolio Optimization Model 231
8.8 Summary 234
Suggested Readings 234
Exercises 234

CHAPTER 9 *LINEAR OPTIMIZATION* 239

9.1 Introduction 239
9.1.1 Linearity 239
9.1.2 Simplex Algorithm 240
9.2 Allocation Models 241
9.2.1 Formulation 241
9.2.2 Spreadsheet Model 242
9.2.3 Optimization 244
9.3 Covering Models 246
9.3.1 Formulation 246
9.3.2 Spreadsheet Model 247
9.3.3 Optimization 247
9.4 Blending Models 248
9.4.1 Blending Constraints 249
9.4.2 Formulation 251
9.4.3 Spreadsheet Model 252
9.4.4 Optimization 252
9.5 Sensitivity Analysis for Linear Programs 253
9.5.1 Sensitivity to Objective Function Coefficients 254
9.5.2 Sensitivity to Constraint Constants 255
9.6 Patterns in Linear Programming Solutions 258
9.6.1 Identifying Patterns 258
9.6.2 Further Examples 260
9.6.3 Review 264

9.7 Data Envelopment Analysis 265
9.8 Summary 269
Suggested Readings 270
Exercises 270
Appendix 9.1 The Solver Sensitivity Report 274

CHAPTER 10 *OPTIMIZATION OF NETWORK MODELS* 277

10.1 Introduction 277
10.2 The Transportation Model 277
10.2.1 Flow Diagram 278
10.2.2 Model Formulation 278
10.2.3 Spreadsheet Model 279
10.2.4 Optimization 280
10.2.5 Modifications to the Model 281
10.2.6 Sensitivity Analysis 282
10.3 Assignment Model 286
10.3.1 Model Formulation 287
10.3.2 Spreadsheet Model 287
10.3.3 Optimization 288
10.3.4 Sensitivity Analysis 288
10.4 The Transshipment Model 289
10.4.1 Formulation 290
10.4.2 Spreadsheet Model 291
10.4.3 Optimization 292
10.4.4 Sensitivity Analysis 293
10.5 A Standard Form for Network Models 293
10.6 Network Models with Yields 295
10.6.1 Yields as Reductions in Flow 295
10.6.2 Yields as Expansions in Flow 297
10.6.3 Patterns in General Network Models 300
10.7 Network Models for Process Technologies 301
10.7.1 Formulation 301
10.7.2 Spreadsheet Model 303
10.7.3 Optimization 304
10.8 Summary 304
Exercises 305

CHAPTER 11 *INTEGER OPTIMIZATION* 309

11.1 Introduction 309
11.2 Integer Variables and the Integer Solver 310
11.3 Binary Variables and Binary Choice Models 312
11.3.1 The Capital Budgeting Problem 312
11.3.2 The Set Covering Problem 315
11.4 Binary Variables and Logical Relationships 316
11.4.1 Relationships Among Projects 317
11.4.2 Linking Constraints and Fixed Costs 319
11.4.3 Threshold Levels and Quantity Discounts 323
11.5 The Facility Location Model 324
11.5.1 The Capacitated Problem 325
11.5.2 The Uncapacitated Problem 327
11.5.3 The Assortment Model 329
11.6 Summary 330
Suggested Readings 331
Exercises 331

CHAPTER 12 *OPTIMIZATION OF NONSMOOTH MODELS* 335

12.1 Introduction 335
12.2 Features of the Evolutionary Solver 335
12.3 Curve Fitting (Revisited) 338
12.4 The Advertising Budget Problem (Revisited) 339
12.5 The Capital Budgeting Problem (Revisited) 342
12.6 The Fixed Cost Problem (Revisited) 344
12.7 The Machine-Sequencing Problem 345
12.8 The Traveling Salesperson Problem 347
12.9 Group Assignment 350
12.10 Summary 352
Exercises 352

CHAPTER 13 *DECISION ANALYSIS* 357

13.1 Introduction 357
13.2 Payoff Tables and Decision Criteria 358
13.2.1 Benchmark Criteria 358
13.2.2 Incorporating Probabilities 359
13.3 Using Trees to Model Decisions 361
13.3.1 Decision Trees 362
13.3.2 Decision Trees for a Series of Decisions 364
13.3.3 Principles for Building and Analyzing Decision Trees 367
13.3.4 The Cost of Uncertainty 368
13.4 Using Decision Tree Software 369
13.4.1 Solving a Simple Example with Decision Tree 370
13.4.2 Sensitivity Analysis with Decision Tree 371
13.4.3 Minimizing Expected Cost with Decision Tree 373
13.5 Maximizing Expected Utility With Decision Tree 375
13.6 Summary 378
Suggested Readings 378
Exercises 378

CHAPTER 14 *MONTE CARLO SIMULATION* 383

14.1 Introduction 383
14.2 A Simple Illustration 384
14.3 The Simulation Process 386
14.3.1 Base-Case Model 387
14.3.2 Sensitivity Analysis 388
14.3.3 Specifying Probability Distributions 390
14.3.4 Specifying Outputs 391
14.3.5 Setting Simulation Parameters 391
14.3.6 Analyzing Simulation Outputs 391
14.4 Corporate Valuation Using Simulation 395
14.4.1 Base-Case Model 396
14.4.2 Sensitivity Analysis 398
14.4.3 Selecting Probability Distributions 399
14.4.4 Simulation Analysis 399
14.4.5 Simulation Sensitivity 402
14.5 Option Pricing Using Simulation 404
14.5.1 The Logic of Options 405
14.5.2 Modeling Stock Prices 405
14.5.3 Pricing an Option 408
14.5.4 Sensitivity to Volatility 410
14.5.5 Simulation Precision 410
14.6 Selecting Uncertain Parameters 411
14.7 Selecting Probability Distributions 413
14.7.1 Empirical Data and Judgmental Data 413
14.7.2 Six Essential Distributions 414
14.7.3 Fitting Distributions to Data 418
14.8 Ensuring Precision in Outputs 420
14.8.1 Illustrations of Simulation Error 420
14.8.2 Precision versus Accuracy 421
14.8.3 An Experimental Method 422

14.8.4 Precision Using the MSE 423
14.8.5 Simulation Error in a Decision Context 423
14.9 Interpreting Simulation Outcomes 424
14.9.1 Simulation Results 424
14.9.2 Displaying Results on the Spreadsheet 426
14.10 When to Simulate and When Not To Simulate 426
14.11 Summary 428
Suggested Readings 428
Exercises 429

CHAPTER 15 *OPTIMIZATION IN SIMULATION* 435

15.1 Introduction 435
15.2 Optimization with One or Two Decision Variables 435
15.2.1 Base-case Model 436
15.2.2 Grid Search 438
15.2.3 Optimizing using Simulation Sensitivity 439
15.2.4 Optimizing using Solver 442
15.3 Stochastic Optimization 442
15.3.1 Optimization of the Base-Case Model 442
15.3.2 A Portfolio Optimization Problem 445
15.4 Chance Constraints 448
15.5 Two-Stage Problems with Recourse 453
15.6 Summary 457
Suggested Readings 458
Exercises 458

MODELING CASES 463

APPENDIX 1 *BASIC EXCEL SKILLS* 479

Introduction 479
Excel Prerequisites 479
The Excel Window 480
Configuring Excel 482
Manipulating Windows and Sheets 483
Navigation 484
Selecting Cells 485
Entering Text and Data 485
Editing Cells 486
Formatting 487
Basic Formulas 488
Basic Functions 489
Charting 493
Printing 495
Help Options 496
Keyboard Shortcuts 497
Cell Comments 497
Naming Cells and Ranges 499
Some Advanced Tools 502
R1C1 Reference Style 502
Mixed Addresses 503
Advanced Functions 503

APPENDIX 2 *MACROS AND VBA* 507

Introduction 507
Recording a Macro 507
Editing a Macro 510
Creating a User-Defined Function 512
Suggested Readings 514

APPENDIX 3 *BASIC PROBABILITY CONCEPTS* 515

Introduction 515
Probability Distributions 515
Examples of Discrete Distributions 518
Examples of Continuous Distributions 519
Expected Values 521
Cumulative Distribution Functions 522
Tail Probabilities 523
Variability 524
Sampling 525

Preface

This is a book for business analysts about *modeling*. A *model* is a simplified representation of a situation or problem, and *modeling* is the process of building, refining, and analyzing that representation for greater insight and improved decision making. Some models are so common that they are thought of as routine instruments rather than models. A budget, a cash flow projection, or a business plan may have many uses, but each one is a model. In addition, many sophisticated models are embedded in software. Option pricing models, credit scoring models, or inventory models are key components of important decision-support systems. Beyond these types, we encounter many customized models built by the millions of people who routinely use spreadsheet software to analyze business situations. This group includes consultants, venture capitalists, marketing analysts, and operations specialists. Almost anyone who uses spreadsheets in business has been involved with models and can benefit from formal training in the use of models.

Models also play a central role in management education. A short list of models that nearly every business student encounters would include cash flow models, stock price models, option pricing models, product life cycle models, market diffusion models, order quantity models, and project scheduling models. For the management student, a basic ability to model in spreadsheets can be a powerful tool for acquiring a deeper understanding of the various functional areas of business. But to fully understand the implications of these models, a student needs to appreciate what a model is and how to learn from it. Our book provides that knowledge.

For many years, modeling was performed primarily by highly trained specialists using mainframe computers. Consequently, even a simple model was costly and frequently required a long development time. The assumptions and results often seemed impenetrable to business managers because they were removed from the modeling process. This situation has changed radically with the advent of personal computers and electronic spreadsheets. Now, managers and analysts can build their own models and produce their own analyses. This newer kind of modeling is known as *end-user modeling*. Now that virtually every analyst has access to a powerful computer, the out-of-pocket costs of modeling have become negligible. The major cost now is the analyst's *time*: time to define the problem, gather the data, build and debug a model, and use the model to support the decision process. For this time to be well spent, the analyst must be efficient and effective in the modeling process. This book is designed to improve modeling *efficiency* by focusing on the most important tasks and tools and by suggesting how to avoid unproductive steps in the modeling effort. This book is also designed to improve modeling *effectiveness* by introducing the most relevant analytic methods and emphasizing procedures that lead to the deepest business insights.

One of our reasons for writing this book was the conviction that many analysts were not being appropriately educated as modelers. Business students typically take courses in statistics and management science, and these quantitative fields are often covered by the umbrella term Business Analytics. But these courses generally offer little training in practical modeling, and students, also, often receive inadequate training in the use of spreadsheets for modeling. In most educational programs, the emphasis is on *models,* rather than on *modeling*. That is, the curriculum covers a number of classical models that have proven useful in management education or business. Although studying the classics may be valuable for a number of reasons (and our book covers a number of the classics), studying models does not provide the full range of skills needed to build models for new

situations. We have also met many analysts who view modeling, essentially, as a matter of having strong spreadsheet skills. But spreadsheet skills alone are not sufficient. The spreadsheet is only one tool in the creative, open-ended problem-solving process we call modeling. Modeling is both a technical discipline and a *craft*. The craft aspects of the process have largely been overlooked in the education of business analysts. Our purpose is to provide both the technical knowledge and the craft skills needed to develop real expertise in business modeling. In this book, therefore, we cover the three skill areas that a business analyst needs to become an effective modeler:

- spreadsheet engineering
- management science
- modeling craft

NEW IN THE FIFTH EDITION

We have changed the title of this edition from *Management Science* to the more widely recognized term *Business Analytics*. This term has recently become popular to describe a wide range of quantitative fields, including classical statistics, data exploration and data mining, management science, and modeling.

The major change in this edition is to the organization of Chapter 6 on data mining. Many data mining textbooks stress the mathematical details of the algorithms. In contrast, our presentation emphasizes the broad set of skills necessary to carry out a data mining project. A basic understanding of the algorithms is necessary, but equally important are skills such as recognizing and dealing with overfitting, and tailoring the output measures to the problem at hand. In order to better communicate these skills we have reorganized the chapter by starting with a single algorithm and several applications. Then more algorithms for both classification and prediction are introduced in separate sections and applied to a single data set. As in the previous edition, we use the software XLMiner to support both data exploration and data mining. XLMiner is one component of the integrated software suite Analytic Solver Platform, which we use throughout the book.

Beyond the capabilities of XLMiner, several features of Analytic Solver Platform have been corrected or changed, and we have updated our coverage accordingly. Excel itself has been updated with the advent of Office 2016, and we have made corresponding changes in our exhibits and narratives. Further changes in Excel or in Analytic Solver Platform will be reflected promptly in the software that accompanies the text and in electronic versions of the *Fifth Edition*.

TO THE READER

Modeling, like painting or singing, cannot be learned entirely from a book. However, a book can establish principles, provide examples, and offer additional practice. We suggest that the reader take an *active learning* attitude toward this book. This means working to internalize the skills taught here by tackling as many new problems as possible. It also means applying these skills to everyday situations in other classes or on the job. Modeling expertise (as opposed to modeling appreciation) can be acquired only by *doing* modeling. There is no substitute for experience.

The book is organized into four parts:

- Spreadsheet modeling in the context of problem solving (Chapters 1–4)
- Data analysis (Chapters 5–7)
- Optimization (Chapters 8–12)
- Decision analysis and simulation (Chapters 13–15)

Our table of contents provides details on the topic coverage in the various chapters, and in Chapter we provide a diagram of the prerequisite logic among the chapters. Several chapters contain advanced material in sections marked with (∗). Students can find spreadsheet files for all models presented in the text on the book's website at http://faculty.tuck.dartmouth.edu/management-science/.

TO THE TEACHER

It is far easier to teach technical skills in Excel or in business analytics than it is to teach modeling. Nonetheless, modeling skills *can* be taught successfully, and a variety of effective approaches are available. Feedback from users of our book and reviewers of previous editions suggests that almost as many course designs exist as there are instructors for this subject. Our book does not represent an idealized version of our own course; rather, it is intended to be a versatile resource that can support a selection of topics in management science, spreadsheet engineering, and modeling craft.

On the book's website, http://faculty.tuck.dartmouth.edu/management-science/, we provide some teaching tips and describe our views on some of the ways this material can be delivered successfully in a graduate or undergraduate course. All spreadsheet files for the models in the text, as well as PowerPoint slides, can be found on the site. In addition, we provide some sample syllabi to illustrate the course designs that other instructors have delivered with the help of this book. For access to the Analytic Solver Platform for Education software, contact Frontline Systems at academic@solver.com or call 775-831-0300.

SOFTWARE ACCOMPANYING THE FIFTH EDITION

Users of the *Fifth Edition* have access to spreadsheet files for all the models presented in the text. Users also have access to Analytic Solver Platform for Education, an integrated software platform for sensitivity analysis, optimization, decision trees, data exploration and data mining, and simulation.

Purchasers of a new text (in either hard copy or electronic format) have access to Analytic Solver Platform for Education through their course instructor—see www.solver.com/student. Instructors, as well as purchasers not enrolled in a course, may contact Frontline Systems Inc. at academic@solver.com or 775-831-0300.

ACKNOWLEDGMENTS

A book such as this evolves over many years of teaching and research. Our ideas have been influenced by our students and by other teachers, not all of whom we can acknowledge here. Our students at Dartmouth's Tuck School of Business have participated in many of our teaching experiments and improved our courses through their invaluable feedback. Without the collaborative spirit our students bring to their education, we could not have developed our ideas as we have.

As in the first edition, we wish to mention the many excellent teachers and writers whose ideas we have adapted. We acknowledge Don Plane, Cliff Ragsdale, and Wayne Winston for their pioneering work in teaching management science with spreadsheets, and the later influence of Tom Grossman, Peter Bell, Zeger Degraeve, and Erhan Erkut on our work.

The first edition benefited from careful reviews from the following reviewers: Jerry Allison (University of Central Oklahoma), Jonathan Caulkins (Carnegie-Mellon University), Jean-Louis Goffin (McGill University), Roger Grinde (University of New Hampshire), Tom Grossman (University of Calgary), Raymond Hill (Air Force Institute of Technology), Alan Johnson (United States Military Academy), Prafulla Joglekar (LaSalle University), Tarja Joro (University of Alberta), Ron Klimberg (Saint Joseph's University), Larry Leblanc (Vanderbilt University), Jerry May (University of Pittsburgh), Jim Morris (University of Wisconsin), Jim Mote (RPI), Chuck Noon (University of Tennessee), Tava Olsen (Washington University), Fred Raafat (San Diego State University), Gary Reeves (University of South Carolina), Moshe Rosenwein (Columbia University), David Schilling (Ohio State University), Linus Schrage (University of Chicago), Donald Simmons (Ithaca College), George Steiner (McMaster University), and Stephen Thorpe (Drexel University).

Additional feedback has come from the following: R. Kim Craft (Southern Utah University), Joan Donohue (University of South Carolina), Steve Ford (University of the South), Phillip Fry (Boise State University), Li Guodong (Maryville University), LeRoy

Honeycutt (Gardner-Webb University), Rich Kilgore (St. Louis University), Frank Krzystofiak (University at Buffalo SUNY), Shailesh Kulkarni (University of North Texas), Dale Lehman (Alaska Pacific University), Vedran Lelas (Plymouth State University), David Walter Little (High Point University), Leo Lopes (University of Arizona), Alvin J. Martinez (University of Puerto Rico, Rio Piedras), Jacquelynne McLellan (Frostburg State University), Ajay Mishra (Binghamton University SUNY), Shimon Y. Nof (Purdue University), Manuel Nunez (University of Connecticut), Alan Olinsky (Bryant University), Tava Olsen (Washington University), Susan Palocsay (James Madison University), Ganga P. Ramdas (Lincoln University), B. Madhu Rao (Bowling Green State University), Jim Robison (Sonoma State University), Christopher M. Rump (Bowling Green State University), Thomas Sandman (California State University, Sacramento), Sergei Savin (Columbia University), Daniel Shimshak (University of Massachusetts Boston), Minghe Sun (University of Texas at San Antonio), and David Tufte (Southern Utah University).

Beth Golub of John Wiley & Sons encouraged us to write this book for years and supported us in writing a new kind of textbook. She also tapped an extensive network of contacts for useful feedback and helped us improve successive editions. With the *Third Edition*, the responsibility passed from Beth to Lisé Johnson, and with this edition to Darren Lalonde, whose team has continued providing the editorial support we have come to appreciate.

SGP
KRB

About The Authors

Steve Powell is a Professor at the Tuck School of Business at Dartmouth College. His primary research interest lies in modeling production and service processes, but he has also been active in research in energy economics, marketing, and operations. At Tuck, he has developed a variety of courses in management science, including the core Decision Science course and electives in the Art of Modeling, Business Analytics, and Simulation. He originated the Teacher's Forum column in *Interfaces*, and he has written a number of articles on teaching modeling to practitioners. He was the Academic Director of the annual INFORMS Teaching of Management Science Workshops. In 2001, he was awarded the INFORMS Prize for the Teaching of Operations Research/Management Science Practice. Along with Ken Baker, he has directed the Spreadsheet Engineering Research Project. In 2008, he co-authored *Modeling for Insight: A Master Class for Business Analysts* with Robert J. Batt.

Ken Baker is a faculty member at Dartmouth College. He is currently the Nathaniel Leverone Professor of Management at the Tuck School of Business and Adjunct Professor at the Thayer School of Engineering. At Dartmouth, he has taught courses related to Management Science, Decision Support Systems, Manufacturing Management, and Environmental Management. Along with Steve Powell, he has directed the Spreadsheet Engineering Research Project. He is the author of two other textbooks, *Optimization Modeling with Spreadsheets* and *Principles of Sequencing and Scheduling* (with Dan Trietsch), in addition to a variety of technical articles. He has served as the Tuck School's Associate Dean and as the Co-Director of the Master's Program in Engineering Management. He is an INFORMS Fellow as well as a Fellow of the Manufacturing and Service Operations Management (MSOM) Society.

1 Introduction

1.1 MODELS AND MODELING

Modeling is the process of creating a simplified representation of reality and working with this representation in order to understand or control some aspect of the world. While this book is devoted to *mathematical* models, modeling itself is a ubiquitous human activity. In fact, it seems to be one of just a few fundamental ways in which we humans understand our environment.

As an example, a map is one of the most common models we encounter. Maps are models because they simplify reality by leaving out most geographic details in order to highlight the important features we need. A state road map, for example, shows major roads but not minor ones, gives rough locations of cities but not individual addresses, and so on. The map we choose must be appropriate for the need we have: a long trip across several states requires a regional map, while a trip across town requires a detailed street map. In the same way, a good model must be appropriate for the specific uses intended for it. A complex model of the economy is probably not appropriate for pricing an individual product. Similarly, a back-of-the-envelope calculation is likely to be inappropriate for acquiring a multibillion-dollar company.

Models take many different forms: mental, visual, physical, mathematical, and spreadsheet, to name a few. We use mental models constantly to understand the world and to predict the outcomes of our actions. Mental models are informal, but they do allow us to make a quick judgment about the desirability of a particular proposal. For example, mental models come into play in a hiring decision. One manager has a mental model that suggests that hiring older workers is not a good idea because they are slow to adopt new ways; another manager has a mental model that suggests hiring older workers is a good idea because they bring valuable experience to the job. We are often unaware of our own mental models, yet they can have a strong influence on the actions we take, especially when they are the primary basis for decision making.

While everyone uses mental models, some people routinely use other kinds of models in their professional lives. Visual models include maps, as we mentioned earlier. Organization charts are also visual models. They may represent reporting relationships, reveal the locus of authority, suggest major channels of communication, and identify responsibility for personnel decisions. Visual models are used in various sports, for instance, as when a coach sketches the playing area and represents team members and opponents as X's and O's. Most players probably don't realize that they are using a model for the purposes of understanding and communication.

Physical models are used extensively in engineering to assist in the design of airplanes, ships, and buildings. They are also used in science, as, for example, in depicting the spatial arrangement of amino acids in the DNA helix or the makeup of a chemical compound. Architects use physical models to show how a proposed building fits within its surroundings.

Mathematical models take many forms and are used throughout science, engineering, and public policy. For instance, a groundwater model helps determine where flooding is most likely to occur, population models predict the spread of infectious disease, and exposure-assessment models forecast the impact of toxic spills. In other settings, traffic-flow models predict the buildup of highway congestion, fault-tree models help reveal the causes of an accident, and reliability models suggest when equipment may need

replacement. Mathematical models can be extremely powerful, especially when they give clear insights into the forces driving a particular outcome.

1.1.1 Why Study Modeling?

What are the benefits of building and using formal models, as opposed to relying on mental models or just "gut feel?" The primary purpose of modeling is to generate *insight*, by which we mean an improved understanding of the situation or problem at hand. While mathematical models consist of numbers and symbols, the real benefit of using them is to make better *decisions*. Better decisions are most often the result of improved understanding, not just the numbers themselves.

Thus, we study modeling primarily because it improves our thinking skills. Modeling is a discipline that provides a structure for problem solving. The fundamental elements of a model—such as parameters, decisions, and outcomes—are useful concepts in all problem solving. Modeling provides examples of clear and logical analysis and helps raise the level of our thinking.

Modeling also helps improve our quantitative reasoning skills. Building a model demands care with units and with orders of magnitude, and it teaches the importance of numeracy. Many people are cautious about quantitative analysis because they do not trust their own quantitative skills. In the best cases, a well-structured modeling experience can help such people overcome their fears, build solid quantitative skills, and improve their performance in a business world that demands (and rewards) these skills.

Any model is a laboratory in which we can experiment and learn. An effective modeler needs to develop an open, inquiring frame of mind to go along with the necessary technical skills. Just as a scientist uses the laboratory to test ideas, hypotheses, and theories, a business analyst can use a model to test the implications of alternative courses of action and develop not only a recommended decision but, equally important, the rationale for why that decision is preferred. The easy-to-understand rationale behind the recommendation often comes from insights the analyst has discovered while testing a model.

1.1.2 Models in Business

Given the widespread use of mathematical models in science and engineering, it is not surprising to find that they are also widely used in the business world. We refer to people who routinely build and analyze formal models in their professional lives as **business analysts**. In our years of training managers and management students, we have found that strong modeling skills are particularly important for consultants, as well as for financial analysts, marketing researchers, entrepreneurs, and others who face challenging business decisions of real economic consequence. Practicing business analysts and students intending to become business analysts are the intended audience for this book.

Just as there are many types of models in science, engineering, public policy, and other domains outside of business, many different types of models are used in business. We distinguish here four model types that exemplify different levels of interaction with, and participation by, the people who use the models:

- One-time decision models
- Decision-support models
- Models embedded in computer systems
- Models used in business education

Many of the models business analysts create are used in **one-time decision problems**. A corporate valuation model, for example, might be used intensively during merger negotiations but never thereafter. In other situations, a one-time model might be created to evaluate the profit impact of a promotion campaign, or to help select a health insurance provider, or to structure the terms of a supply contract. One-time models are usually built by decision makers themselves, frequently under time pressure. Managerial judgment is often used as a substitute for empirical data in such models, owing to time constraints and data limitations. Most importantly, this type of model involves the user intensively because the model is usually tailored to a particular decision-making need. One major benefit of studying modeling is to gain skills in building and using one-time models effectively.

Decision-support systems are computer systems that tie together models, data, analysis tools, and presentation tools into a single integrated package. These systems are intended for repeated use, either by executives themselves or by their analytic staff. Decision-support systems are used in research and development planning at pharmaceutical firms, pricing decisions at oil companies, and product-line profitability analysis at manufacturing firms, to cite just a few examples. Decision-support systems are usually built and maintained by information systems personnel, but they represent the routine use of what were once one-time decision models. After a one-time model becomes established, it can be adapted for broader and more frequent use in the organization. Thus, the models within decision-support systems may initially be developed by managers and business analysts, but later streamlined by information systems staff for a less intensive level of human interaction. An additional benefit of studying modeling is to recognize possible improvements in the design and operation of decision-support systems.

Embedded models are those contained within computer systems that perform routine, repeated tasks with little or no human involvement. Many inventory replenishment decisions are made by automated computer systems. Loan payments on auto leases or prices for stock options are also determined by automated systems. Routine real estate appraisals may also be largely automated. In these cases, the models themselves are somewhat hidden in the software. Many users of embedded models are not aware of the underlying models; they simply assume that the "system" knows how to make the right calculations. An ancillary benefit of studying modeling is to become more aware, and perhaps more questioning, of these embedded models.

1.1.3 Models in Business Education

Models are useful not only in the business world, but also in the academic world where business analysts are educated. The modern business curriculum is heavily dependent on models for delivering basic concepts as well as for providing numerical results. An introductory course in Finance might include an option-pricing model, a cash-management model, and the classic portfolio model. A basic Marketing course might include demand curves for pricing analysis, a diffusion model for new-product penetration, and clustering models for market segmentation. In Operations Management, we might encounter inventory models for stock control, allocation models for scheduling production, and newsvendor models for trading off shortage and surplus outcomes. Both micro- and macroeconomics are taught almost exclusively through models. Aggregate supply-and-demand curves are models, as are production functions.

Most of the models used in education are highly simplified, or *stylized*, in order to preserve clarity. Stylized models are frequently used to provide insight into qualitative phenomena, not necessarily to calculate precise numerical results. In this book, we frequently use models from business education as examples, so that we can combine learning about business with learning about models. In fact, the tools presented in this book can be used throughout the curriculum to better understand the various functional areas of business.

1.1.4 Benefits of Business Models

Modeling can benefit business decision making in a variety of ways.

- Modeling allows us to make inexpensive errors. Wind-tunnel tests are used in airplane design partly because if every potential wing design had to be built into a full-scale aircraft and flown by a pilot, we would lose far too many pilots. In a similar way, we can propose ideas and test them in a model, without having to suffer the consequences of bad ideas in the real world.
- Modeling allows us to explore the impossible. Many companies have policies, procedures, or habits that prevent them from making certain choices. Sometimes these habits prevent them from discovering better ways of doing business. Modeling can be used to explore these "impossible" alternatives and to help convince the skeptics to try a different approach.
- Modeling can improve business intuition. As we have said, a model is a laboratory in which we perform experiments. We can usually learn faster from laboratory experiments than from experience in the real world. With a model, we can try thousands of

combinations that would take many years to test in the real world. We can also try extreme ideas that would be too risky to test in the real world. And we can learn about how the world works by simulating a hundred years of experience in just a few seconds.

- Modeling provides information in a timely manner. For example, while a survey could be used to determine the potential demand for a product, effective modeling can often give useful bounds on the likely range of demand in far less time.
- Finally, modeling can reduce costs. Data collection is often expensive and time-consuming. An effective modeler may be able to provide the same level of information at a much lower cost.

Even among those who do not build models, skill in working with models is very important. Most business students eventually find themselves on a team charged with recommending a course of action. If these teams do not build models themselves, they often work with internal or external consultants who do. Experience in building and analyzing models is, in our minds, the best training for working effectively on problem-solving teams. People who have not actually built a few models themselves often accept model results blindly or become intimidated by the modeling process. A well-trained analyst not only appreciates the power of modeling but also remains skeptical of models as panaceas.

We believe that modeling skills are useful to a very broad range of businesspeople, from junior analysts without a business degree to senior vice presidents who do their own analysis. Many recent graduates have only a superficial knowledge of these tools because their education emphasized passive consumption of other people's models rather than active model building. Thus, there is considerable potential even among master's-level graduates to improve their modeling skills so that they can become more capable of carrying out independent analyses of important decisions. The only absolute prerequisite for using this book and enhancing that skill is a desire to use logical, analytic methods to reach a higher level of understanding in the decision-making world.

1.2 THE ROLE OF SPREADSHEETS

Because spreadsheets are the principal vehicle for modeling in business, spreadsheet models are the major type we deal with in this book. Spreadsheet models are also mathematical models, but, for many people, spreadsheet mathematics is more accessible than algebra or calculus. Spreadsheet models do have limitations, of course, but they allow us to build more detailed and more complex models than traditional mathematics allows. They also have the advantage of being pervasive in business analysis. Finally, the spreadsheet format corresponds nicely to the form of accounting statements that are used for business communication; in fact, the word "spreadsheet" originates in accounting and only recently has come to mean the electronic spreadsheet.

It has been said that the spreadsheet is the *second best* way to do many kinds of analysis and is therefore the *best* way to do most modeling. In other words, for any one modeling task, a more powerful, flexible, and sophisticated software tool is almost certainly available. In this sense, the spreadsheet is the Swiss Army knife of business analysis. Most business analysts lack the time, money, and knowledge to learn and use a different software tool for each problem that arises, just as most of us cannot afford to carry around a complete toolbox to handle the occasional screw we need to tighten. The practical alternative is to use the spreadsheet (and occasionally one of its sophisticated add-ins) to perform most modeling tasks. An effective modeler will, of course, have a sense for the limitations of a spreadsheet and will know when to use a more powerful tool.

Despite its limitations, the electronic spreadsheet represents a breakthrough technology for practical modeling. Prior to the 1980s, modeling was performed only by specialists using demanding software on expensive hardware. This meant that only the most critical business problems could be analyzed using models because only these problems justified the large budgets and long time commitments required to build, debug, and apply the models of the day. This situation has changed dramatically in the past 30 years or so. First the personal computer, then the spreadsheet, and recently the arrival of add-ins for specialized analyses have put tremendous analytical power at the hands of

anyone who can afford a laptop and some training. In fact, we believe the 1990s will come to be seen as the dawn of the "end-user modeling" era. End-user modelers are analysts who are not specialists in modeling, but who can create an effective spreadsheet and manipulate it for insight. The problems that end-user modelers can solve are typically not the multibillion-dollar, multiyear variety; those are still the preserve of functional-area specialists and sophisticated computer scientists. Rather, the end user can apply modeling effectively to hundreds of important but smaller-scale situations that in the past would not have benefited from this approach. We provide many illustrations throughout this book.

Spreadsheet skills themselves are now in high demand in many jobs, although experts in Excel may not be skilled modelers. In our recent survey of MBAs from the Tuck School of Business (available at http://mba.tuck.dartmouth.edu/spreadsheet/), we found that 77 percent said that spreadsheets were either "very important" or "critical" in their work. Good training in spreadsheet modeling, in what we call **spreadsheet engineering**, is valuable because it can dramatically improve both the efficiency and effectiveness with which the analyst uses spreadsheets.

1.2.1 Risks of Spreadsheet Use

Countless companies and individuals rely on spreadsheets every day. Most users assume their spreadsheet models are error free. However, the available evidence suggests just the opposite: many, perhaps most, spreadsheets contain internal errors, and more errors are introduced as these spreadsheets are used and modified. Given this evidence, and the tremendous risks of relying on flawed spreadsheet models, it is critically important to learn how to create spreadsheets that are as close to error free as possible and to use spreadsheets in a disciplined way to avoid mistakes.

It is rare to read press reports on problems arising from erroneous spreadsheets. Most companies do not readily admit to these kinds of mistakes. However, the few reports that have surfaced are instructive. For many years the European Spreadsheet Risks Interest Group (EUSPRIG) has maintained a website (http://www.eusprig.org/horror-stories.htm) that documents dozens of verified stories about spreadsheet errors that have had a quantifiable impact on the organization. Here is just a small sample:

- Some candidates for police officer jobs are told that they have passed the test when in fact they have failed. Reason: improper sorting of the spreadsheet.
- An energy company overcharges consumers between $200 million and $1 billion. Reason: careless naming of spreadsheet files.
- A think-tank reports that only 11 percent of a local population has at least a bachelor's degree when in fact the figure is 20 percent. Reason: a copy-and-paste error in a spreadsheet.
- Misstated earnings lead the stock price of an online retailer to fall 25 percent in a day and the CEO to resign. Reason: a single erroneous numerical input in a spreadsheet.
- A school loses £30,000 because its budget is underestimated. Reason: numbers entered as text in a spreadsheet.
- The Business Council reports that its members forecast slow growth for the coming year when their outlook is actually quite optimistic. Reason: the spreadsheet shifted, so the wrong numbers appeared in the wrong columns.
- Benefits of unbundling telecommunication services are understated by $50 million. Reason: incorrect references in a spreadsheet formula.

These cases suggest that spreadsheets can lead to costly errors in a variety of ways. But are spreadsheets themselves properly *built* in the first place? Apparently they are not, at least according to several research studies. In our own investigation of 50 spreadsheets that were being used by various organizations, fewer than 10 percent were free of errors.[1] This evidence serves notice that errors in spreadsheets may be rampant and insidious.

Despite the research evidence, very few corporations employ even the most basic design methodologies and error-detection procedures. These procedures take time and effort, whereas one of the great appeals of spreadsheet modeling is that it can be done

[1] S. Powell, K. Baker and B. Lawson, "Errors in Operational Spreadsheets," *Journal of End User Computing* 21, (July–September, 2009): 24–36.

quickly and easily, even by business analysts who are not professional programmers. But ease of use is a delusion if the results contain significant errors.

Briefly stated, the business world is still at an early stage of understanding how to develop error-free spreadsheets. Organizations are in need of better methods for detecting errors and more reliable procedures for preventing errors in the first place. However, the research literature on these topics has not advanced very far, and the state of the art remains somewhat primitive.

1.2.2 Challenges for Spreadsheet Users

Spreadsheets represent the ubiquitous software platform of business. Millions of spreadsheet models are used each day to make decisions involving billions of dollars, and thousands of new spreadsheets come into being each day. Given this usage pattern, we might think that spreadsheet engineering is a well-developed discipline and that expertise in spreadsheet modeling can be found in just about any company. Amazingly, the opposite is true.

What is the current state of spreadsheet use by end-user modelers? The evidence available from audits of existing spreadsheets, laboratory experiments, surveys of end users, and field visits suggests that, despite widespread use, the quality with which spreadsheets are engineered generally remains poor. There are four major problem areas:

- End-user spreadsheets frequently have bugs.
- End users are overconfident about the quality of their own spreadsheets.
- The process that end users employ to create their spreadsheets is inefficient at best and chaotic at worst.
- End users fail to employ the most productive methods for generating insights from their spreadsheets.

Our own research, conducted as part of the Spreadsheet Engineering Research Project (http://mba.tuck.dartmouth.edu/spreadsheet/), found that a substantial majority of spreadsheets in use contain at least one error. A follow-up study found that most of these errors had a substantial impact on the quantitative analysis in the spreadsheets. However, our investigation also suggested that errors in individual cells may be only a symptom. The underlying cause often seems to be a high degree of complexity in the model, even when the corresponding problem is relatively simple. Complexity arises in many ways:

- Individual cell formulas that are excessively long and involved
- Poorly designed worksheets that are difficult to navigate and understand
- Poorly organized workbooks whose underlying structure is concealed

Spreadsheets that are overly complex and difficult for anyone other than the designer to use, even if they are technically correct, may be the cause of some of the costly mistakes attributed to spreadsheets.

Laboratory experiments have uncovered another disturbing fact about spreadsheet modeling: end users appear to be overconfident about the likelihood of errors in their own spreadsheets. In these experiments, undergraduate volunteers were asked to build a spreadsheet for a well-defined problem. After they were done, the volunteers were given time to review and audit their models. Finally, they were asked to evaluate the likelihood that their model contained one or more bugs. While 18 percent of the subjects thought their models had one or more bugs, the actual proportion proved to be 80 percent. That is, 80 percent of these spreadsheets actually had bugs, but only about 18 percent of those who built them suspected they had bugs. This finding of overconfidence is consistent with the findings of other studies: people tend to underestimate the possibility that they might make mistakes. Unfortunately, this overconfidence translates directly into a casual attitude toward spreadsheet design and ultimately into a disturbingly high error rate among spreadsheets in actual use.

Our observations and research into how end users actually construct spreadsheets suggest that the process is often inefficient:

- End users typically do not plan their spreadsheets. Instead, they build them live at the keyboard. The result in many cases is extensive rework. (In our survey of MBA graduates, we found that about 20 percent sketched a spreadsheet on paper first,

whereas about 50 percent started by entering data and formulas directly into the computer.)

- End users do not use a conscious prototyping approach, which involves building a series of models starting with the simplest and gradually adding complexity.
- End users rarely spend time debugging their models, unless the model performs in such a counterintuitive manner that it demands intervention.
- End users almost never subject their spreadsheets to review by another person. In general, end users appear to trust that the model they *thought* they had built is actually the model they *see* on their screens, despite the fact that spreadsheets show only numbers, not the relationships behind the numbers.
- Finally, many end users, even some who are experts in Excel, do not consistently use tools that can help generate the insights that make modeling worthwhile. Excel's Data Table and Goal Seek tools, to cite just two examples, are overlooked by the majority of end users. Without these tools, the end user either fails to ask questions that can provide telling insights, or else wastes time generating results that could be found more easily.

The evidence is strong that the existing state of spreadsheet design and use is generally inadequate. This is one reason we devote a significant portion of this book to spreadsheet engineering. Only with a solid foundation in spreadsheet engineering can the business analyst effectively generate real insights from spreadsheet models.

1.2.3 Background Knowledge for Spreadsheet Modeling

Many people new to modeling fear it because modeling reminds them of painful experiences with mathematics. We do not wish to downplay the essentially mathematical nature of modeling, even modeling using spreadsheets. However, an effective modeler does not need to know any really advanced math. Knowledge of basic algebra (including functions such as the quadratic, exponential, and logarithmic), simple logic (as expressed in an IF statement or the MAX function), and basic probability (distributions and sampling, for example) will usually suffice. When we find it necessary to use any higher math in this book, we provide explanations. But our focus here is less on the mathematical details of models than on the creative process of constructing and using models.

We assume throughout this book that the reader has a basic familiarity with Excel. This includes the ability to build a simple spreadsheet, enter and format text and data, use formulas and simple functions such as SUM, construct graphs, and so on. We do not assume the reader is an expert in Excel, nor do we assume knowledge of the advanced tools we cover, such as optimization and simulation. We have found that, in many situations, advanced Excel skills are not required for building effective models. And we believe that the main purpose of modeling is to improve the insight of the modeler. Thus, it is appropriate for a modeler with only basic Excel skills to build a model using only basic tools, and it is appropriate for a modeler with advanced skills to draw on advanced tools when needed. We have also found that too much skill in Excel can sometimes distract from the essential modeling tasks, which are almost always more about finding a simple and effective representation of the problem at hand than about finding some Excel trick.

For easy reference, we have included Appendix 1 to give an overview of Excel, from the basics of entering text and data to advanced formulas and functions. In addition, Appendix 2 covers the use of macros and an introduction to Visual Basic for Applications (VBA). We expect most readers to already know Excel to some degree, and to use these appendices as needed to hone specific skills. We believe that, by working through the examples in the book, the reader's Excel skills will improve naturally and painlessly, just as ours have improved over years of building models and teaching modeling to students whose Excel skills often exceeded ours.

1.3 THE REAL WORLD AND THE MODEL WORLD

We stated at the outset that modeling provides a structure for problem solving. It does this through a process of abstraction, in which the essence of the problem is captured in a simplified form. Because of this abstraction process, modeling does not come naturally to

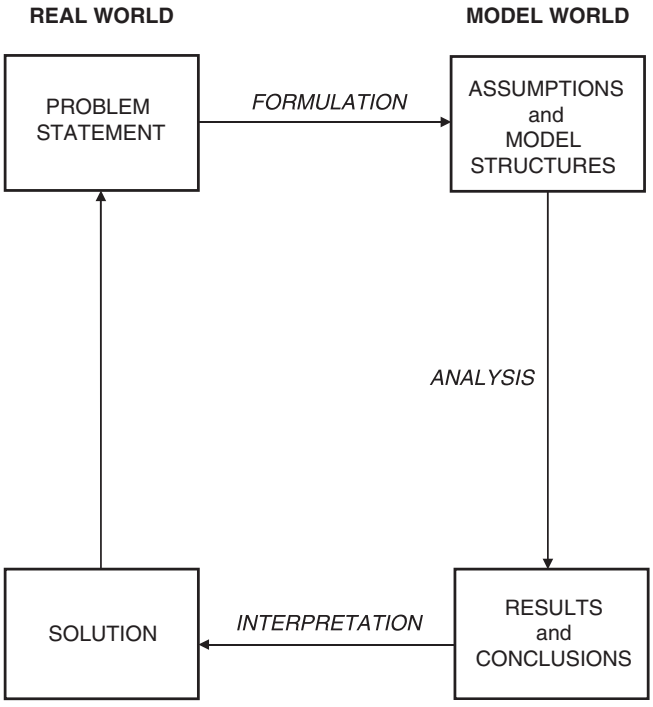

FIGURE 1.1 The Real-World and the Model World

most people but must be learned. Because it does not come naturally, it can appear to be artificial and counterintuitive, causing many students of modeling to become uncomfortable with the process. This section attempts to reduce that discomfort by placing modeling in the context of problem solving in the real world.

A model is an abstraction, or simplification, of the real world. It is a laboratory—an artificial environment—in which we can experiment and test ideas without the costs and risks of experimenting with real systems and organizations. Figure 1.1 is a schematic showing how modeling creates an artificial world. We begin in the real world, usually with a messy problem to solve. If we determine that modeling is an appropriate tool, we then move across an invisible boundary into the model world.

In order to move into the model world, we abstract the essential features of the real world, leaving behind all the nonessential detail and complexity. We then construct our laboratory by combining our abstractions with specific assumptions and building a model of the essential aspects of the real world. This is the process of **model formulation**. It is an exercise in simplifying the actual situation and capturing its essence, with a specific purpose in mind. The model formulation process typically forces us to confront four features of a model:

- Decisions
- Outcomes
- Structure
- Data

Decisions refers to possible choices, or courses of action, that we might take. These would be controllable variables, such as quantities to buy, manufacture, spend, or sell. (By contrast, uncontrollable variables such as tax rates or the cost of materials are not decision variables.) **Outcomes** refers to the consequences of the decisions—the performance measures we use to evaluate the results of taking action. Examples might include profit, cost, or efficiency. **Structure** refers to the logic and the mathematics that link the elements of our model together. A simple example might be the equation $P = R - C$, in which profit is calculated as the difference between revenue and cost. Another example might be the relationship $F = I + P - S$, in which final inventory is calculated from initial inventory, production, and shipments. Finally, **data** refers to specific numerical assumptions. That may mean actual observations of the real world (often called "raw" or "empirical" data), or it may mean estimates of uncontrollable variables in the problem's environment. Examples might include the interest rate on borrowed funds, the production capacity of a manufacturing facility, or the first-quarter sales for a new product.

Once it is built, we can use the model to test ideas and evaluate solutions. This is a process of **analysis**, in which we apply logic, often with the support of software, to take us from our assumptions and abstractions to a set of derived conclusions. Unlike model formulation, which tends to be mostly an art, analysis is much more of a science. It relies on mathematics and reason in order to explore the implications of our assumptions. This exploration process leads, hopefully, to insights about the problem confronting us. Sometimes, these insights involve an understanding of why one solution is beneficial and another is not; at other times, the insights involve understanding the sources of risk in a particular solution. In another situation, the insights involve identifying the decisions that are most critical to a good result, or identifying the inputs that have the strongest influence on a particular outcome. In each instance, it is crucial to understand that these insights are derived from the *model* world and not from the *real* world. Whether they apply to the real world is another matter entirely and requires managerial judgment.

To make the model insights useful, we must first translate them into the terms of the real world and then communicate them to the actual decision makers involved. Only then do *model* insights turn into useful *managerial* insights. And only then can we begin the process of evaluating solutions in terms of their impact on the real world. This is a process of **interpretation**, and here again, the process is an art. Good modelers can move smoothly back and forth between the model world and the real world, deriving crisp insights from the model, and translating the insights, modifying them as needed, to account for real-world complexities not captured in the model world.

This schematic description of the modeling process highlights some of the reasons it can be a challenge to incorporate modeling into problem solving. Powerful in competent hands, modeling is also somewhat esoteric. It involves deliberate abstraction and simplification of a situation, which appears to many people as a counterproductive exercise. Modeling requires a willingness to temporarily set aside much of the richness of the real world and to operate in the refined and artificial world of models and model insights. It also requires confidence that whatever insights arise in the model world can be translated into useful ideas in the real world. In addition, it requires an ability to mix art with science in order to exploit the modeling process to its full potential. Until we have some experience with this process, we may be resistant and skeptical. And it is always easy to criticize a model as being too simple. Good models are as simple as they can possibly be. But this very simplicity can appear to be a fatal flaw to skeptics. Nevertheless, modeling is one of the most powerful tools in the problem solver's tool kit, simply because there is no more practical way to arrive at the insights modeling can provide.

1.4 LESSONS FROM EXPERT AND NOVICE MODELERS

Perhaps the best way to become a good modeler is to serve an apprenticeship under an expert. Unfortunately, such opportunities are rare. Moreover, experts in all fields find it difficult to express their expertise or to teach it. While narrow, **technical skills** are relatively easy to teach (e.g., how to use the NPV function in Excel), expertise consists largely of **craft skills** that are more difficult to teach (e.g., what to include and exclude from the model). In the arts, there is a tradition of studio training, where a teacher poses artistic challenges to students and then coaches them as they work through the problems on their own. This is one way for students to acquire some of the difficult-to-articulate craft skills of the master. There is no comparable tradition in the mathematical fields; in fact, there is a long-standing belief that modeling cannot be taught but must simply be acquired by experience.

One way to improve modeling skills is to understand what expert and novice modelers actually do when they build and use models. From closely observing experts, we can attempt to articulate a set of modeling best practices. From observing novices, we can understand the reasons for their relatively lower level of modeling accomplishment: the blind alleys, counterproductive behaviors, misperceptions, and cognitive limitations that keep them from attaining expert performance. In this section, we summarize research studies on both expert and novice modelers.

1.4.1 Expert Modelers

An alternative to an apprenticeship under an expert is to study experts in a laboratory setting. Tom Willemain did this in a series of experiments with 12 expert modelers.