GLOBAL EDITION

Business Analytics

THIRD EDITION

James R. Evans

Business Analytics

This page intentionally left blank

Business Analytics

Methods, Models, and Decisions

James R. Evans | University of Cincinnati

THIRD EDITION

GLOBAL EDITION

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • São Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Please contact https://support.pearson.com/getsupport/s/contactsupport with any queries on this content

Cover image by Allies Interactive / Shutterstock.

Pearson Education Limited
KAO Two
KAO Park
Hockham Way
Harlow
Essex
CM17 9SR
United Kingdom

and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsonglobaleditions.com

Authorized adaptation from the United States edition, entitled *Business Analytics*, 3rd Edition, ISBN 978-0-13-523167-8 by James R. Evans, published by Pearson Education © 2020.

Print ISBN 10: 1-292-33906-3
Print ISBN 13: 978-1-292-33906-1
eBook ISBN 13: 978-1-292-33904-7

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

1 20

Typeset in Times NR MT Pro by SPi Global

Brief Contents

Preface *17*
About the Author *25*
Credits *27*

Part 1 Foundations of Business Analytics

Chapter 1 Introduction to Business Analytics 29
Chapter 2 Database Analytics 75

Part 2 Descriptive Analytics

Chapter 3 Data Visualization 113
Chapter 4 Descriptive Statistics 143
Chapter 5 Probability Distributions and Data Modeling 201
Chapter 6 Sampling and Estimation 247
Chapter 7 Statistical Inference 275

Part 3 Predictive Analytics

Chapter 8 Trendlines and Regression Analysis 311
Chapter 9 Forecasting Techniques 353
Chapter 10 Introduction to Data Mining 383
Chapter 11 Spreadsheet Modeling and Analysis 405
Chapter 12 Simulation and Risk Analysis 451

Part 4 Prescriptive Analytics

Chapter 13 Linear Optimization 493
Chapter 14 Integer and Nonlinear Optimization 551
Chapter 15 Optimization Analytics 593

Part 5 Making Decisions

Chapter 16 Decision Analysis 631

Appendix A *661*
Glossary *685*
Index *693*

This page intentionally left blank

Contents

Preface 17
About the Author 25
Credits 27

Part 1: Foundations of Business Analytics

Chapter 1: Introduction to Business Analytics 29

Learning Objectives 29
What Is Business Analytics? 31
Using Business Analytics 32 • Impacts and Challenges 33
Evolution of Business Analytics 34
Analytic Foundations 34 • Modern Business Analytics 35 • Software Support and Spreadsheet Technology 37
Analytics in Practice: Social Media Analytics 38
Descriptive, Predictive, and Prescriptive Analytics 39
Analytics in Practice: Analytics in the Home Lending and Mortgage Industry 41
Data for Business Analytics 42
Big Data 44 • Data Reliability and Validity 44
Models in Business Analytics 45
Descriptive Models 47 • Predictive Models 49 • Prescriptive Models 50 • Model Assumptions 51 • Uncertainty and Risk 53
Problem Solving with Analytics 54
Recognizing a Problem 54 • Defining the Problem 54 • Structuring the Problem 55 • Analyzing the Problem 55 • Interpreting Results and Making a Decision 55 • Implementing the Solution 55
Analytics in Practice: Developing Effective Analytical Tools at Hewlett-Packard 56
Key Terms 57 • Chapter 1 Technology Help 57 • Problems and Exercises 57 • Case: Performance Lawn Equipment 59

Appendix A1: Basic Excel Skills 61

Excel Formulas and Addressing 62
Copying Formulas 63
Useful Excel Tips 63
Excel Functions 64
Basic Excel Functions 64 • Functions for Specific Applications 65 • Insert Function 66 • Date and Time Functions 67
Miscellaneous Excel Functions and Tools 68
Range Names 68 • VALUE Function 71 • Paste Special 71 • Concatenation 72 • Error Values 72
Problems and Exercises 73

Chapter 2: Database Analytics 75

Learning Objectives 75
Data Sets and Databases 77
Using Range Names in Databases 78
Analytics in Practice: Using Big Data to Monitor Water Usage in Cary, North Carolina 79
Data Queries: Tables, Sorting, and Filtering 79
Sorting Data in Excel 80 • Pareto Analysis 81 • Filtering Data 82 • Database Functions 84
Analytics in Practice: Discovering the Value of Database Analytics at Allders International 86
Logical Functions 87
Lookup Functions for Database Queries 89
Excel Template Design 92
Data Validation Tools 93 • Form Controls 95
PivotTables 98
PivotTable Customization 100 • Slicers 103
Key Terms 104 • Chapter 2 Technology Help 104 • Problems and Exercises 105 • Case: People's Choice Bank 109 • Case: Drout Advertising Research Project 110

Part 2: Descriptive Analytics

Chapter 3: Data Visualization 113

Learning Objectives 113
The Value of Data Visualization 114
Tools and Software for Data Visualization 116
Analytics in Practice: Data Visualization for the New York City Police Department's Domain Awareness System 116
Creating Charts in Microsoft Excel 116
Column and Bar Charts 117 • Data Label and Data Table Chart Options 118 • Line Charts 119 • Pie Charts 120 • Area Charts 121 • Scatter Charts and Orbit Charts 122 • Bubble Charts 123 • Combination Charts 124 • Radar Charts 125 • Stock Charts 125 • Charts from PivotTables 125 • Geographic Data 126
Other Excel Data Visualization Tools 126
Data Bars 126 • Color Scales 127 • Icon Sets 128 • Sparklines 129
Dashboards 131
Analytics in Practice: Driving Business Transformation with IBM Business Analytics 132
Key Terms 133 • Chapter 3 Technology Help 133 • Problems and Exercises 134 • Case: Performance Lawn Equipment 135

Appendix A3: Additional Tools for Data Visualization 136

Hierarchy Charts 136
Waterfall Charts 136
PivotCharts 138
Tableau 139
Problems and Exercises 141

Chapter 4: Descriptive Statistics 143

Learning Objectives 143
Analytics in Practice: Applications of Statistics in Health Care 145
Metrics and Data Classification 146
Frequency Distributions and Histograms 148
Frequency Distributions for Categorical Data 148 • Relative Frequency Distributions 149 • Frequency Distributions for Numerical Data 150 • Grouped Frequency Distributions 151 • Cumulative Relative Frequency Distributions 154 • Constructing Frequency Distributions Using PivotTables 155
Percentiles and Quartiles 157
Cross-Tabulations 158
Descriptive Statistical Measures 160
Populations and Samples 160 • Statistical Notation 161 • Measures of Location: Mean, Median, Mode, and Midrange 161 • Using Measures of Location in Business Decisions 163 • Measures of Dispersion: Range, Interquartile Range, Variance, and Standard Deviation 165 • Chebyshev's Theorem and the Empirical Rules 168 • Standardized Values (*Z*-Scores) 170 • Coefficient of Variation 171 • Measures of Shape 172 • Excel *Descriptive Statistics* Tool 174
Computing Descriptive Statistics for Frequency Distributions 175
Descriptive Statistics for Categorical Data: The Proportion 177
Statistics in PivotTables 178
Measures of Association 179
Covariance 180 • Correlation 181 • Excel Correlation Tool 183
Outliers 184
Using Descriptive Statistics to Analyze Survey Data 186
Statistical Thinking in Business Decisions 187
Variability in Samples 188
Analytics in Practice: Applying Statistical Thinking to Detecting Financial Problems 190

Key Terms 191 • Chapter 4 Technology Help 192 • Problems and Exercises 193 • Case: Drout Advertising Research Project 198 • Case: Performance Lawn Equipment 198

Appendix A4: Additional Charts for Descriptive Statistics in Excel for Windows 199

Problems and Exercises 200

Chapter 5: Probability Distributions and Data Modeling 201

Learning Objectives 201
Basic Concepts of Probability 203
Experiments and Sample Spaces 203 • Combinations and Permutations 203 • Probability Definitions 205 • Probability Rules and Formulas 207 • Joint and Marginal Probability 208 • Conditional Probability 210
Random Variables and Probability Distributions 213
Discrete Probability Distributions 215
Expected Value of a Discrete Random Variable 216 • Using Expected Value in Making Decisions 217 • Variance of a Discrete Random Variable 219 • Bernoulli Distribution 219 • Binomial Distribution 220 • Poisson Distribution 221
Analytics in Practice: Using the Poisson Distribution for Modeling Bids on Priceline 223

Continuous Probability Distributions 224
Properties of Probability Density Functions 224 • Uniform Distribution 225 • Normal Distribution 227 • The NORM.INV Function 228 • Standard Normal Distribution 229 • Using Standard Normal Distribution Tables 230 • Exponential Distribution 231 • Triangular Distribution 232
Data Modeling and Distribution Fitting 233
Goodness of Fit: Testing for Normality of an Empirical Distribution 235
Analytics in Practice: The Value of Good Data Modeling in Advertising 237
Key Terms 238 • Chapter 5 Technology Help 238 • Problems and Exercises 239 • Case: Performance Lawn Equipment 245

Chapter 6: Sampling and Estimation 247

Learning Objectives 247
Statistical Sampling 248
Sampling Methods 249
Analytics in Practice: Using Sampling Techniques to Improve Distribution 251
Estimating Population Parameters 252
Unbiased Estimators 252 • Errors in Point Estimation 253 • Understanding Sampling Error 254
Sampling Distributions 256
Sampling Distribution of the Mean 256 • Applying the Sampling Distribution of the Mean 257
Interval Estimates 257
Confidence Intervals 258 • Confidence Interval for the Mean with Known Population Standard Deviation 259 • The *t*-Distribution 260 • Confidence Interval for the Mean with Unknown Population Standard Deviation 261 • Confidence Interval for a Proportion 261 • Additional Types of Confidence Intervals 263
Using Confidence Intervals for Decision Making 263
Data Visualization for Confidence Interval Comparison 264
Prediction Intervals 265
Confidence Intervals and Sample Size 266
Key Terms 268 • Chapter 6 Technology Help 268 • Problems and Exercises 269 • Case: Drout Advertising Research Project 272 • Case: Performance Lawn Equipment 273

Chapter 7: Statistical Inference 275

Learning Objectives 275
Hypothesis Testing 276
Hypothesis-Testing Procedure 276
One-Sample Hypothesis Tests 277
Understanding Potential Errors in Hypothesis Testing 278 • Selecting the Test Statistic 279 • Finding Critical Values and Drawing a Conclusion 280 • Two-Tailed Test of Hypothesis for the Mean 282 • Summary of One-Sample Hypothesis Tests for the Mean 283 • *p*-Values 284 • One-Sample Tests for Proportions 285 • Confidence Intervals and Hypothesis Tests 286 • An Excel Template for One-Sample Hypothesis Tests 286
Two-Sample Hypothesis Tests 287

Two-Sample Tests for Differences in Means 288 • Two-Sample Test for Means with Paired Samples 290 • Two-Sample Test for Equality of Variances 292
Analysis of Variance (ANOVA) 294
Assumptions of ANOVA 296
Chi-Square Test for Independence 297
Cautions in Using the Chi-Square Test 299 • Chi-Square Goodness of Fit Test 300
Analytics in Practice: Using Hypothesis Tests and Business Analytics in a Help Desk Service Improvement Project 301

Key Terms 302 • Chapter 7 Technology Help 302 • Problems and Exercises 304 • Case: Drout Advertising Research Project 309 • Case: Performance Lawn Equipment 309

Part 3: Predictive Analytics

Chapter 8: Trendlines and Regression Analysis 311

Learning Objectives 311
Modeling Relationships and Trends in Data 313
Analytics in Practice: Using Predictive Trendline Models at Procter & Gamble 317
Simple Linear Regression 317
Finding the Best-Fitting Regression Line 319 • Using Regression Models for Prediction 319 • Least-Squares Regression 320 • Simple Linear Regression with Excel 322 • Regression as Analysis of Variance 324 • Testing Hypotheses for Regression Coefficients 325 • Confidence Intervals for Regression Coefficients 325
Residual Analysis and Regression Assumptions 326
Checking Assumptions 327
Multiple Linear Regression 329
Analytics in Practice: Using Linear Regression and Interactive Risk Simulators to Predict Performance at Aramark 332
Building Good Regression Models 334
Correlation and Multicollinearity 336 • Practical Issues in Trendline and Regression Modeling 338
Regression with Categorical Independent Variables 338
Categorical Variables with More Than Two Levels 341
Regression Models with Nonlinear Terms 343

Key Terms 345 • Chapter 8 Technology Help 345 • Problems and Exercises 346 • Case: Performance Lawn Equipment 350

Chapter 9: Forecasting Techniques 353

Learning Objectives 353
Analytics in Practice: Forecasting Call-Center Demand at L.L.Bean 354
Qualitative and Judgmental Forecasting 355
Historical Analogy 327 • The Delphi Method 355 • Indicators and Indexes 356
Statistical Forecasting Models 357
Forecasting Models for Stationary Time Series 359
Moving Average Models 359 • Error Metrics and Forecast Accuracy 361 • Exponential Smoothing Models 363
Forecasting Models for Time Series with a Linear Trend 366
Double Exponential Smoothing 366 • Regression-Based Forecasting for Time Series with a Linear Trend 368

Forecasting Time Series with Seasonality 369
Regression-Based Seasonal Forecasting Models 369 • Holt-Winters Models for Forecasting Time Series with Seasonality and No Trend 371 • Holt-Winters Models for Forecasting Time Series with Seasonality and Trend 373 • Selecting Appropriate Time-Series-Based Forecasting Models 376
Regression Forecasting with Causal Variables 376
The Practice of Forecasting 377
Analytics in Practice: Forecasting at NBCUniversal 378
Key Terms 379 • Chapter 9 Technology Help 380 • Problems and Exercises 380 • Case: Performance Lawn Equipment 382

Chapter 10: Introduction to Data Mining 383

Learning Objectives 383
The Scope of Data Mining 384
Cluster Analysis 386
Measuring Distance Between Objects 387 • Normalizing Distance Measures 388 • Clustering Methods 388
Classification 390
An Intuitive Explanation of Classification 391 • Measuring Classification Performance 392 • Classification Techniques 393
Association 398
Cause-and-Effect Modeling 400
Analytics In Practice: Successful Business Applications of Data Mining 402
Key Terms 402 • Chapter 10 Technology Help 403 • Problems and Exercises 403 • Case: Performance Lawn Equipment 404

Chapter 11: Spreadsheet Modeling and Analysis 405

Learning Objectives 405
Analytics in Practice: Using Spreadsheet Modeling and Analysis at Nestlé 407
Model-Building Strategies 407
Building Models Using Logic and Business Principles 407 • Building Models Using Influence Diagrams 408 • Building Models Using Historical Data 409 • Model Assumptions, Complexity, and Realism 410
Implementing Models on Spreadsheets 410
Spreadsheet Design 411 • Spreadsheet Quality 412 • Data Validation 414
Analytics in Practice: Spreadsheet Engineering at Procter & Gamble 416
Descriptive Spreadsheet Models 416
Staffing Decisions 417 • Single-Period Purchase Decisions 418 • Overbooking Decisions 420
Analytics in Practice: Using an Overbooking Model at a Student Health Clinic 421
Retail Markdown Decisions 421
Predictive Spreadsheet Models 423
New Product Development Model 423 • Cash Budgeting 425 • Retirement Planning 426 • Project Management 426
Prescriptive Spreadsheet Models 429
Portfolio Allocation 429 • Locating Central Facilities 430 • Job Sequencing 432

Analyzing Uncertainty and Model Assumptions 434
What-If Analysis 434 • Data Tables 434 • Scenario Manager 437 • Goal Seek 438
Key Terms 440 • Chapter 11 Technology Help 441 • Problems and Exercises 442 • Case: Performance Lawn Equipment 449

Chapter 12: Simulation and Risk Analysis 451

Learning Objectives 451
Monte Carlo Simulation 453
Random Sampling from Probability Distributions 455
Generating Random Variates using Excel Functions 457
Discrete Probability Distributions 457 • Uniform Distributions 458 • Exponential Distributions 459 • Normal Distributions 459 • Binomial Distributions 461 • Triangular Distributions 461
Monte Carlo Simulation in Excel 463
Profit Model Simulation 463 • New Product Development 466 • Retirement Planning 468 • Single-Period Purchase Decisions 469 • Overbooking Decisions 472 • Project Management 472
Analytics in Practice: Implementing Large-Scale Monte Carlo Spreadsheet Models 474
Dynamic Systems Simulation 475
Simulating Waiting Lines 477
Analytics in Practice: Using Systems Simulation for Agricultural Product Development 480
Key Terms 481 • Chapter 12 Technology Help 481 • Problems and Exercises 481 • Case: Performance Lawn Equipment 491

Part 4: Prescriptive Analytics

Chapter 13: Linear Optimization 493

Learning Objectives 493
Optimization Models 494
Analytics in Practice: Using Optimization Models for Sales Planning at NBC 496
Developing Linear Optimization Models 497
Identifying Decision Variables, the Objective, and Constraints 498 • Developing a Mathematical Model 499 • More About Constraints 500 • Implementing Linear Optimization Models on Spreadsheets 502 • Excel Functions to Avoid in Linear Optimization 503
Solving Linear Optimization Models 504
Solver Answer Report 506 • Graphical Interpretation of Linear Optimization with Two Variables 507
How *Solver* Works 513
How *Solver* Creates Names in Reports 514
Solver Outcomes and Solution Messages 515
Unique Optimal Solution 515 • Alternative (Multiple) Optimal Solutions 515 • Unbounded Solution 515 • Infeasibility 517

Applications of Linear Optimization 519
Blending Models 519 • Dealing with Infeasibility 520 • Portfolio Investment Models 521 • Scaling Issues in Using *Solver* 523 • Transportation Models 526 • Multiperiod Production Planning Models 529 • Multiperiod Financial Planning Models 533
Analytics in Practice: Linear Optimization in Bank Financial Planning 536
Key Terms 537 • Chapter 13 Technology Help 537 • Problems and Exercises 538 • Case: Performance Lawn Equipment 550

Chapter 14: Integer and Nonlinear Optimization 551

Learning Objectives 551
Integer Linear Optimization Models 552
Models with General Integer Variables 553 • Workforce-Scheduling Models 556 • Alternative Optimal Solutions 559
Models with Binary Variables 561
Using Binary Variables to Model Logical Constraints 562 • Applications in Supply Chain Optimization 563
Analytics in Practice: Supply Chain Optimization at Procter & Gamble 567
Nonlinear Optimization Models 567
A Nonlinear Pricing Decision Model 567 • Quadratic Optimization 571 • Practical Issues Using *Solver* for Nonlinear Optimization 572
Analytics in Practice: Applying Nonlinear Optimization at Prudential Securities 573
Non-Smooth Optimization 574
Evolutionary Solver 574 • *Evolutionary Solver* for Sequencing and Scheduling Models 577 • The Traveling Salesperson Problem 579
Key Terms 581 • Chapter 14 Technology Help 581 • Problems and Exercises 582 • Case: Performance Lawn Equipment 591

Chapter 15: Optimization Analytics 593

Learning Objectives 593
What-If Analysis for Optimization Models 594
Solver Sensitivity Report 595 • Using the Sensitivity Report 600 • Degeneracy 601 • Interpreting *Solver* Reports for Nonlinear Optimization Models 601
Models with Bounded Variables 603
Auxiliary Variables for Bound Constraints 606
What-If Analysis for Integer Optimization Models 609
Visualization of *Solver* Reports 611
Using Sensitivity Information Correctly 618
Key Terms 622 • Chapter 15 Technology Help 622 • Problems and Exercises 622 • Case: Performance Lawn Equipment 629

Part 5: Making Decisions

Chapter 16: Decision Analysis 631

Learning Objectives 631
Formulating Decision Problems 633
Decision Strategies Without Outcome Probabilities 634

Decision Strategies for a Minimize Objective 634 • Decision Strategies for a Maximize Objective 636 • Decisions with Conflicting Objectives 636

Decision Strategies with Outcome Probabilities 638

Average Payoff Strategy 638 • Expected Value Strategy 638 • Evaluating Risk 639

Decision Trees 640

Decision Trees and Risk 642 • Sensitivity Analysis in Decision Trees 645

The Value of Information 646

Decisions with Sample Information 647 • Bayes's Rule 648

Utility and Decision Making 649

Constructing a Utility Function 650 • Exponential Utility Functions 653

Analytics in Practice: Using Decision Analysis in Drug Development 654

Key Terms 655 • *Chapter 16 Technology Help 655* • *Problems and Exercises 656* • *Case: Performance Lawn Equipment 660*

Online Supplements: Information about how to access and use Analytic Solver Basic are available for download at www.pearsonglobaleditions.com.

Getting Started with Analytic Solver
Using Advanced Regression Techniques in Analytic Solver
Using Forecasting Techniques in Analytic Solver
Using Data Mining in Analytic Solver
Model Analysis in Analytic Solver
Using Monte Carlo Simulation in Analytic Solver
Using Linear Optimization in Analytic Solver
Using Integer and Nonlinear Optimization in Analytic Solver
Using Optimization Parameter Analysis in Analytic Solver
Using Decision Trees in Analytic Solver

Appendix A 661
Glossary 685
Index 693

This page intentionally left blank

Preface

In 2007, Thomas H. Davenport and Jeanne G. Harris wrote a groundbreaking book, *Competing on Analytics: The New Science of Winning* (Boston: Harvard Business School Press). They described how many organizations are using analytics strategically to make better decisions and improve customer and shareholder value. Over the past several years, we have seen remarkable growth in analytics among all types of organizations. The Institute for Operations Research and the Management Sciences (INFORMS) noted that analytics software as a service is predicted to grow at three times the rate of other business segments in upcoming years.[1] In addition, the *MIT Sloan Management Review* in collaboration with the IBM Institute for Business Value surveyed a global sample of nearly 3,000 executives, managers, and analysts.[2] This study concluded that top-performing organizations use analytics five times more than lower performers, that improvement of information and analytics was a top priority in these organizations, and that many organizations felt they were under significant pressure to adopt advanced information and analytics approaches. Since these reports were published, the interest in and the use of analytics has grown dramatically.

In reality, business analytics has been around for more than a half-century. Business schools have long taught many of the core topics in business analytics—statistics, data analysis, information and decision support systems, and management science. However, these topics have traditionally been presented in separate and independent courses and supported by textbooks with little topical integration. This book is uniquely designed to present the emerging discipline of business analytics in a unified fashion consistent with the contemporary definition of the field.

About the Book

This book provides undergraduate business students and introductory graduate students with the fundamental concepts and tools needed to understand the role of modern business analytics in organizations, to apply basic business analytics tools in a spreadsheet environment, and to communicate with analytics professionals to effectively use and interpret analytic models and results for making better business decisions. We take a balanced, holistic approach in viewing business analytics from descriptive, predictive, and prescriptive perspectives that define the discipline.

[1]Anne Robinson, Jack Levis, and Gary Bennett, INFORMS News: INFORMS to Officially Join Analytics Movement. http://www.informs.org/ORMS-Today/Public-Articles/October-Volume-37-Number-5/INFORMS-News-INFORMS-to-Officially-Join-Analytics-Movement.

[2]"Analytics: The New Path to Value," *MIT Sloan Management Review* Research Report, Fall 2010.

This book is organized in five parts.

1. Foundations of Business Analytics

   The first two chapters provide the basic foundations needed to understand business analytics and to manipulate data using Microsoft Excel. Chapter 1 provides an introduction to business analytics and its key concepts and terminology, and includes an appendix that reviews basic Excel skills. Chapter 2, Database Analytics, is a unique chapter that covers intermediate Excel skills, Excel template design, and PivotTables.

2. Descriptive Analytics

   Chapters 3 through 7 cover fundamental tools and methods of data analysis and statistics. These chapters focus on data visualization, descriptive statistical measures, probability distributions and data modeling, sampling and estimation, and statistical inference. We subscribe to the American Statistical Association's recommendations for teaching introductory statistics, which include emphasizing statistical literacy and developing statistical thinking, stressing conceptual understanding rather than mere knowledge of procedures, and using technology for developing conceptual understanding and analyzing data. We believe these goals can be accomplished without introducing every conceivable technique into an 800–1,000 page book as many mainstream books currently do. In fact, we cover all essential content that the state of Ohio has mandated for undergraduate business statistics across all public colleges and universities.

3. Predictive Analytics

   In this section, Chapters 8 through 12 develop approaches for applying trendlines and regression analysis, forecasting, introductory data mining techniques, building and analyzing models on spreadsheets, and simulation and risk analysis.

4. Prescriptive Analytics

   Chapters 13 and 14 explore linear, integer, and nonlinear optimization models and applications. Chapter 15, Optimization Analytics, focuses on what-if and sensitivity analysis in optimization, and visualization of Solver reports.

5. Making Decisions

   Chapter 16 focuses on philosophies, tools, and techniques of decision analysis.

Changes to the Third Edition

The third edition represents a comprehensive revision that includes many significant changes. The book now relies only on native Excel, and is independent of platforms, allowing it to be used easily by students with either PC or Mac computers. These changes provide students with enhanced Excel skills and basic understanding of fundamental concepts. *Analytic Solver* is no longer integrated directly in the book, but is illustrated in online supplements to facilitate revision as new software updates may occur. These supplements plus information regarding how to access *Analytic Solver* may be accessed at http://www.pearsonglobaleditions.com.

Key changes to this edition are as follows:

- Also available for purchase (separately) is MyLab Statistics, a teaching and learning platform that empowers you to reach every student. By combining trusted author content with digital tools and a flexible platform, MyLab personalizes the learning experience and improves results for each student. For example, new Excel and StatCrunch Projects help students develop business decision-making skills.

- Each chapter now includes a short section called *Technology Help*, which provides useful summaries of key Excel functions and procedures, and the use of supplemental software including *StatCrunch* and *Analytic Solver Basic*.
- Chapter 1 includes an Appendix reviewing basic Excel skills, which will be used throughout the book.
- Chapter 2, Database Analytics, is a new chapter derived from the second edition that focuses on applications of Excel functions and techniques for dealing with spreadsheet data, including a new section on Excel template design.
- Chapter 3, Data Visualization, includes a new Appendix illustrating Excel tools for Windows and a brief overview of Tableau.
- Chapter 5, Probability Distributions and Data Modeling, includes a new section on Combinations and Permutations.
- Chapter 6, Sampling and Estimation, provides a discussion of using data visualization for confidence interval comparison.
- Chapter 9, Forecasting Techniques, now includes Excel approaches for double exponential smoothing and Holt-Winters models for seasonality and trend.
- Chapter 10, Introduction to Data Mining, has been completely rewritten to illustrate simple data mining techniques that can be implemented on spreadsheets using Excel.
- Chapter 11, Spreadsheet Modeling and Analysis, is now organized along the analytic classification of descriptive, predictive, and prescriptive modeling.
- Chapter 12 has been rewritten to apply Monte-Carlo simulation using only Excel, with an additional section of systems simulation concepts and approaches.
- Optimization topics have been reorganized into two chapters—Chapter 13, Linear Optimization, and Chapter 14, Integer and Nonlinear Optimization, which rely only on the Excel-supplied *Solver*.
- Chapter 15 is a new chapter called Optimization Analytics, which focuses on what-if and sensitivity analysis, and visualization of *Solver* reports; it also includes a discussion of how *Solver* handles models with bounded variables.

In addition, we have carefully checked, and revised as necessary, the text and problems for additional clarity. We use major section headings in each chapter and tie these clearly to the problems and exercises, which have been revised and updated throughout the book. At the end of each section we added several "Check Your Understanding" questions that provide a basic review of fundamental concepts to improve student learning. Finally, new Analytics in Practice features have been incorporated into several chapters.

Features of the Book

- **Chapter Section Headings**—with "Check Your Understanding" questions that provide a means to review fundamental concepts.
- **Numbered Examples**—numerous, short examples throughout all chapters illustrate concepts and techniques and help students learn to apply the techniques and understand the results.
- **"Analytics in Practice"**—at least one per chapter, this feature describes real applications in business.
- **Learning Objectives**—lists the goals the students should be able to achieve after studying the chapter.

- **Key Terms**—bolded within the text and listed at the end of each chapter, these words will assist students as they review the chapter and study for exams. Key terms and their definitions are contained in the glossary at the end of the book.
- **End-of-Chapter Problems and Exercises**—clearly tied to sections in each chapter, these help to reinforce the material covered through the chapter.
- **Integrated Cases**—allow students to think independently and apply the relevant tools at a higher level of learning.
- **Data Sets and Excel Models**—used in examples and problems and are available to students at www.pearsonglobaleditions.com.

Software Support

Technology Help sections in each chapter provide additional support to students for using Excel functions and tools, Tableau, and StatCrunch.

Online supplements provide detailed information and examples for using *Analytic Solver Basic*, which provides more powerful tools for data mining, Monte-Carlo simulation, optimization, and decision analysis. These can be used at the instructor's discretion, but are not necessary to learn the fundamental concepts that are implemented using Excel. Instructions for obtaining licenses for *Analytic Solver Basic* can be found on the book's website, http://www.pearsonglobaleditions.com.

To the Students

To get the most out of this book, you need to do much more than simply read it! Many examples describe in detail how to use and apply various Excel tools or add-ins. We highly recommend that you *work through these examples* on your computer to replicate the outputs and results shown in the text. You should also *compare mathematical formulas* with spreadsheet formulas and *work through basic numerical calculations by hand*. Only in this fashion will you learn how to use the tools and techniques effectively, gain a better understanding of the underlying concepts of business analytics, and increase your proficiency in using Microsoft Excel, which will serve you well in your future career.

Visit the companion Web site (www.pearsonglobaleditions.com) for access to the following:

- **Online Files:** Data Sets and Excel Models—files for use with the numbered examples and the end-of-chapter problems. (For easy reference, the relevant file names are italicized and clearly stated when used in examples.)
- **Online Supplements for *Analytic Solver Basic*:** Online supplements describing the use of *Analytic Solver* that your instructor might use with selected chapters.

To the Instructors

MyLab Statistics is now available with Evans "Business Analytics" 3e: MyLab™ Statistics is the teaching and learning platform that empowers instructors to reach every student. Teach your course your way with a flexible platform. Collect, crunch, and communicate with data in StatCrunch®, an integrated Web-based statistical software. Empower each learner with personalized and interactive practice. Tailor your course to your students' needs with enhanced reporting features. Available with the complete eText, accessible anywhere with the Pearson eText app.

Instructor's Resource Center—Reached through a link at www.pearsonglobaleditions.com, the Instructor's Resource Center contains the electronic files for the complete Instructor's Solutions Manual, PowerPoint lecture presentations, and the Test Item File.

- **Register, redeem, log in at www.pearsonglobaleditions.com:** instructors can access a variety of print, media, and presentation resources that are available with this book in downloadable digital format.
- *Instructor's Solutions Manual*—The Instructor's Solutions Manual, updated and revised for the second edition by the author, includes Excel-based solutions for all end-of-chapter problems, exercises, and cases.
- *PowerPoint presentations*—The PowerPoint slides, revised and updated by the author, provide an instructor with individual lecture outlines to accompany the text. The slides include nearly all of the figures, tables, and examples from the text. Instructors can use these lecture notes as they are or can easily modify the notes to reflect specific presentation needs.
- *Test Bank*—The TestBank is prepared by Paolo Catasti from Virginia Commonwealth University.
- **Need help?** Pearson Education's dedicated technical support team is ready to assist instructors with questions about the media supplements that accompany this text. The supplements are available to adopting instructors. Detailed descriptions are provided at the Instructor's Resource Center.

Acknowledgments

I would like to thank the staff at Pearson Education for their professionalism and dedication to making this book a reality. In particular, I want to thank Angela Montoya, Kathleen Manley, Karen Wernholm, Kaylee Carlson, Jean Choe, Bob Carroll, and Patrick Barbera. I would also like to thank Gowri Duraiswamy at SPI, and accuracy and solutions checker Jennifer Blue for their outstanding contributions to producing this book. I also want to acknowledge Daniel Fylstra and his staff at Frontline Systems for working closely with me on providing *Analytic Solver Basic* as a supplement with this book. If you have any suggestions or corrections, please contact the author via email at james.evans@uc.edu.

James R. Evans
Department of Operations, Business Analytics, and Information Systems
University of Cincinnati
Cincinnati, Ohio

Global Edition Acknowledgments

Pearson would like to thank Alicia Tan Yiing Fei, Taylor's University Malaysia; Dániel Kehl, University of Pécs; and Roland Baczur, University of Pécs for their contribution to the Global Edition.

Pearson MyLab

Get the Most Out of MyLab Statistics

Statistics courses are continuously evolving to help today's students succeed. It's more challenging than ever to support students with a wide range of backgrounds, learner styles, and math anxieties. The flexibility to build a course that fits instructors' individual course formats—with a variety of content options and multimedia resources all in one place—has made MyLab Statistics the market-leading solution for teaching and learning mathematics since its inception.

78% of students say MyLab Statistics helped them learn their course content.*

Teach your course with a consistent author voice

With market-leading author content options, your course can fit your style. Pearson offers the widest variety of content options, addressing a range of approaches and learning styles, authored by thought leaders across the business and math curriculum. MyLab™ Statistics is tightly integrated with each author's style, offering a range of author-created multimedia resources, so your students have a consistent experience.

Thanks to feedback from instructors and students from more than 10,000 institutions, MyLab Statistics continues to transform—delivering new content, innovative learning resources, and platform updates to support students and instructors, today and in the future.

**Source: 2018 Student Survey, n 31,721*

Pearson MyLab

Resources for Success

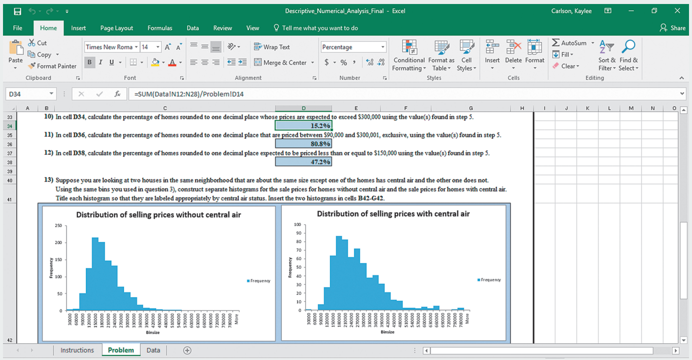
MyLab Statistics Online Course for Business Analytics

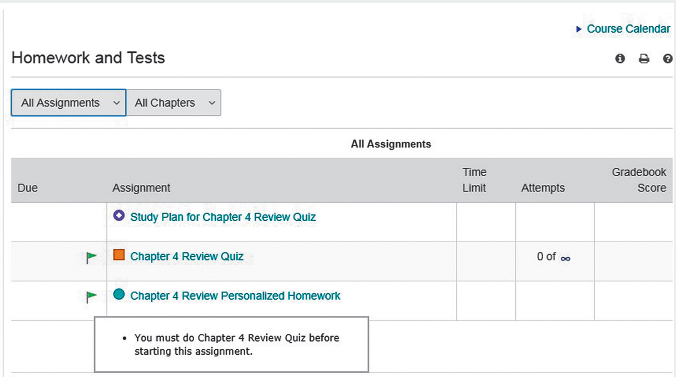
by James R. Evans

MyLab™ Statistics is available to accompany Pearson's market leading text offerings. To give students a consistent tone, voice, and teaching method each text's flavor and approach is tightly integrated throughout the accompanying MyLab Statistics course, making learning the material as seamless as possible.

Enjoy hands off grading with Excel Projects

Using proven, field-tested technology, auto-graded Excel Projects let instructors seamlessly integrate Microsoft Excel content into the course without manually grading spreadsheets. Students can practice important statistical skills in Excel, helping them master key concepts and gain proficiency with the program.



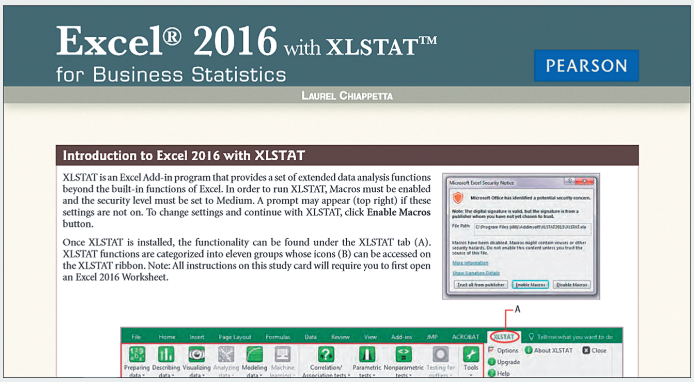


StatCrunch

StatCrunch, Pearson's powerful web-based statistical software, instructors and students can access tens of thousands of data sets including those from the textbook, perform complex analyses, and generate compelling reports. StatCrunch is integrated directly into MyLab Statistics or available as a standalone product. To learn more, go to www.statcrunch.com on any laptop, tablet, or smartphone.

Technology Tutorials and Study Cards

MyLab makes learning and using a variety of statistical software programs as seamless and intuitive as possible. Download data sets from the text and MyLab exercises directly into Excel. Students can also access instructional support tools including tutorial videos, study cards, and manuals for a variety of statistical software programs including StatCrunch, Excel, Minitab, JMP, R, SPSS, and TI83/84 calculators.



pearson.com/mylab/statistics

This page intentionally left blank

About the Author

James R. Evans
Professor Emeritus, University of Cincinnati, Lindner College of Business

James R. Evans is Professor Emeritus in the Department of Operations, Business Analytics, and Information Systems in the College of Business at the University of Cincinnati. He holds BSIE and MSIE degrees from Purdue and a PhD in Industrial and Systems Engineering from Georgia Tech.

Dr. Evans has published numerous textbooks in a variety of business disciplines, including statistics, decision models, and analytics, simulation and risk analysis, network optimization, operations management, quality management, and creative thinking. He has published 100 papers in journals such as *Management Science, IIE Transactions, Decision Sciences, Interfaces*, the *Journal of Operations Management, the Quality Management Journal*, and many others, and wrote a series of columns in *Interfaces* on creativity in management science and operations research during the 1990s. He has also served on numerous journal editorial boards and is a past-president and Fellow of the Decision Sciences Institute. In 1996, he was an INFORMS Edelman Award Finalist as part of a project in supply chain optimization with Procter & Gamble that was credited with helping P&G save over $250,000,000 annually in their North American supply chain, and consulted on risk analysis modeling for Cincinnati 2012's Olympic Games bid proposal.

A recognized international expert on quality management, he served on the Board of Examiners and the Panel of Judges for the Malcolm Baldrige National Quality Award. Much of his research has focused on organizational performance excellence and measurement practices.

This page intentionally left blank

Credits

Text Credits

Chapter 3 *page 129* Prem Thomas, MD and Seth Powsner, MD "Data Presentation for Quality Improvement", 2005, AMIA.

Appendix A *page 661–663* National Institute of Standards and Technology.

Photo Credits

Chapter 1 *page 29* NAN728/Shutterstock *page 56* hans12/Fotolia

Chapter 2 *page 75* NAN728/Shutterstock *page 86* 2jenn/Shutterstock

Chapter 3 *page 113* ESB Professional/Shutterstock

Chapter 4 *page 143* Nataliiap/Shutterstock *page 191* langstrup/123RF

Chapter 5 *page 201* PeterVrabel/Shutterstock *page 223* Fantasista/Fotolia *page 237* Victor Correia/Shutterstock.com

Chapter 6 *page 247* Robert Brown Stock/Shutterstock *page 252* Stephen Finn/Shutterstock.com

Chapter 7 *page 275* Jirsak/Shutterstock *page 301* Hurst Photo/Shutterstock

Chapter 8 *page 311* Luca Bertolli/123RF *page 333* Gunnar Pippel/Shutterstock *page 333* Vector-Illustration/Shutterstock *page 333* Claudio Divizia/Shutterstock *page 333* Nataliia Natykach/Shutterstock

Chapter 9 *page 353* rawpixel/123RF *page 379* Sean Pavone/Shutterstock

Chapter 10 *page 383* Laborant/Shutterstock *page 402* Helder Almeida/Shutterstock

Chapter 11 *page 405* marekuliasz/Shutterstock *page 416* Bryan Busovicki/Shutterstock *page 421* Poprotskiy Alexey/Shutterstock

Chapter 12 *page 451* Stephen Rees/Shutterstock *page 475* Vladitto/Shutterstock

Chapter 13 *page 493* Pinon Road/Shutterstock *page 496* bizoo_n/Fotolia *page 537* 2jenn/Shutterstock

Chapter 14 *page 551* Jirsak/Shutterstock *page 567* Kheng Guan Toh/Shutterstock

Chapter 15 *page 593* Alexander Orlov/Shutterstock

Chapter 16 *page 631* marekuliasz/Shutterstock *page 655* SSokolov/Shutterstock

Front Matter James R. Evans

This page intentionally left blank

CHAPTER

1

Introduction to Business Analytics

NAN728/Shutterstock

LEARNING OBJECTIVES

After studying this chapter, you will be able to:

- Define business analytics.
- Explain why analytics is important in today's business environment.
- State some typical examples of business applications in which analytics would be beneficial.
- Summarize the evolution of business analytics and explain the concepts of business intelligence, operations research and management science, and decision support systems.
- Explain the difference between descriptive, predictive, and prescriptive analytics.
- State examples of how data are used in business.
- Explain the concept of a model and various ways a model can be characterized.
- Define and list the elements of a decision model.
- Illustrate examples of descriptive, predictive, and prescriptive models.
- Explain the difference between uncertainty and risk.
- Define the terms *optimization*, *objective function*, and *optimal solution*.
- Explain the difference between a deterministic and stochastic decision model.
- List and explain the steps in the problem-solving process.

The purpose of this book is to provide you with a basic introduction to the concepts, methods, and models used in business analytics so that you will develop an appreciation not only for its capabilities to support and enhance business decisions, but also for the ability to use business analytics at an elementary level in your work. In this chapter, we introduce you to the field of business analytics and set the foundation for many of the concepts and techniques that you will learn. Let's start with a rather innovative example.

Most of you have likely been to a zoo, seen the animals, had something to eat, and bought some souvenirs. You probably wouldn't think that managing a zoo is very difficult; after all, it's just feeding and taking care of the animals, right? A zoo might be the last place that you would expect to find business analytics being used, but not anymore. The Cincinnati Zoo & Botanical Garden has been an "early adopter" and one of the first organizations of its kind to exploit business analytics.[1]

Despite generating more than two-thirds of its budget through its own fund-raising efforts, the zoo wanted to reduce its reliance on local tax subsidies even further by increasing visitor attendance and revenues from secondary sources such as membership, food, and retail outlets. The zoo's senior management surmised that the best way to realize more value from each visit was to offer visitors a truly transformed customer experience. By using business analytics to gain greater insight into visitors' behavior and tailoring operations to their preferences, the zoo expected to increase attendance, boost membership, and maximize sales.

The project team—which consisted of consultants from IBM and Brightstar Partners, as well as senior executives from the zoo—began translating the organization's goals into technical solutions. The zoo worked to create a business analytics platform that was capable of delivering the desired goals by combining data from ticketing and point-of-sale systems throughout the zoo with membership information and geographical data gathered from the ZIP codes of all visitors. This enabled the creation of reports and dashboards that gave everyone from senior managers to zoo staff access to real-time information that helped them optimize operational management and transform the customer experience.

By integrating weather forecast data, the zoo is now able to compare current forecasts with historic attendance and sales data, supporting better decision making for labor scheduling and inventory planning. Another area where the solution delivers new insight is food service. By opening food outlets at specific times of day when demand is highest (for example, keeping ice cream kiosks open in the

[1] IBM Software Business Analtyics, "Cincinnati Zoo transforms customer experience and boosts profits," © IBM Corporation 2012.

final hour before the zoo closes), the zoo has been able to increase sales significantly. In addition, attendance and revenues have dramatically increased, resulting in annual return on investment of 411%. The business analytics initiative paid for itself within three months and delivers, on average, benefits of $738,212 per year. Specifically,

- The zoo has seen a 4.2% rise in ticket sales by targeting potential visitors who live in specific ZIP codes.
- Food revenues increased 25% by optimizing the mix of products on sale and adapting selling practices to match peak purchase times.
- Eliminating slow-selling products and targeting visitors with specific promotions enabled an 18% increase in merchandise sales.
- The zoo was able to cut its marketing expenditure, saving $40,000 in the first year, and reduce advertising expenditure by 43% by eliminating ineffective campaigns and segmenting customers for more targeted marketing.

Because of the zoo's success, other organizations such as Point Defiance Zoo & Aquarium in Tacoma, Washington, and History Colorado Center, a museum in Denver, have embarked on similar initiatives.

What Is Business Analytics?

Everyone makes decisions. Individuals face personal decisions such as choosing a college or graduate program, making product purchases, selecting a mortgage instrument, and investing for retirement. Managers in business organizations make numerous decisions every day. Some of these decisions include what products to make and how to price them, where to locate facilities, how many people to hire, where to allocate advertising budgets, whether or not to outsource a business function or make a capital investment, and how to schedule production. Many of these decisions have significant economic consequences; moreover, they are difficult to make because of uncertain data and imperfect information about the future.

Managers today no longer make decisions based on pure judgment and experience; they rely on factual data and the ability to manipulate and analyze data to supplement their intuition and experience, and to justify their decisions. What makes business decisions complicated today is the overwhelming amount of available data and information. Data to support business decisions—including those specifically collected by firms as well as through the Internet and social media such as Facebook—are growing exponentially and becoming increasingly difficult to understand and use. As a result, many companies have recently established analytics departments; for instance, IBM reorganized its consulting business and established a new 4,000-person organization focusing on analytics. Companies are increasingly seeking business graduates with the ability to understand and use analytics. The demand for professionals with analytics expertise has skyrocketed, and many universities now have programs in analytics.[2]

[2]Matthew J. Liberatore and Wenhong Luo, "The Analytics Movement: Implications for Operations Research," Interfaces, 40, 4 (July–August 2010): 313–324.

Business analytics, or simply **analytics**, is the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions. Business analytics is "a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving."[3] Business analytics is supported by various tools such as Microsoft Excel and various Excel add-ins, commercial statistical software packages such as SAS or Minitab, and more complex business intelligence suites that integrate data with analytical software.

Using Business Analytics

Tools and techniques of business analytics are used across many areas in a wide variety of organizations to improve the management of customer relationships, financial and marketing activities, human capital, supply chains, and many other areas. Leading banks use analytics to predict and prevent credit fraud. Investment firms use analytics to select the best client portfolios to manage risk and optimize return. Manufacturers use analytics for production planning, purchasing, and inventory management. Retailers use analytics to recommend products to customers and optimize marketing promotions. Pharmaceutical firms use analytics to get life-saving drugs to market more quickly. The leisure and vacation industries use analytics to analyze historical sales data, understand customer behavior, improve Web site design, and optimize schedules and bookings. Airlines and hotels use analytics to dynamically set prices over time to maximize revenue. Even sports teams are using business analytics to determine both game strategy and optimal ticket prices.[4] For example, teams use analytics to decide on ticket pricing, who to recruit and trade, what combinations of players work best, and what plays to run under different situations.

Among the many organizations that use analytics to make strategic decisions and manage day-to-day operations are Caesars Entertainment, the Cleveland Indians baseball, Phoenix Suns basketball, and New England Patriots football teams, Amazon.com, Procter & Gamble, United Parcel Service (UPS), and Capital One bank. It was reported that nearly all firms with revenues of more than $100 million are using some form of business analytics.

Some common types of business decisions that can be enhanced by using analytics include

- pricing (for example, setting prices for consumer and industrial goods, government contracts, and maintenance contracts),
- customer segmentation (for example, identifying and targeting key customer groups in retail, insurance, and credit card industries),
- merchandising (for example, determining brands to buy, quantities, and allocations),
- location (for example, finding the best location for bank branches and ATMs, or where to service industrial equipment),
- supply chain design (for example, determining the best sourcing and transportation options and finding the best delivery routes),

[3]Liberatore and Luo, "The Analytics Movement".

[4]Jim Davis, "8 Essentials of Business Analytics," in "Brain Trust—Enabling the Confident Enterprise with Business Analytics" (Cary, NC: SAS Institute, Inc., 2010): 27–29. www.sas.com/bareport

- staffing (for example, ensuring the appropriate staffing levels and capabilities and hiring the right people—sometimes referred to as “people analytics”),
- health care (for example, scheduling operating rooms to improve utilization, improving patient flow and waiting times, purchasing supplies, predicting health risk factors),

and many others in operations management, finance, marketing, and human resources—in fact, in every discipline of business.[5]

Various research studies have discovered strong relationships between a company’s performance in terms of profitability, revenue, and shareholder return and its use of analytics. Top-performing organizations (those that outperform their competitors) are three times more likely to be sophisticated in their use of analytics than lower performers and are more likely to state that their use of analytics differentiates them from competitors.[6] However, research has also suggested that organizations are overwhelmed by data and struggle to understand how to use data to achieve business results and that most organizations simply don’t understand how to use analytics to improve their businesses. Thus, understanding the capabilities and techniques of analytics is vital to managing in today’s business environment.

So, no matter what your job position in an organization is or will be, the study of analytics will be quite important to your future success. You may find many uses in your everyday work for the Excel-based tools that we will study. You may not be skilled in all the technical nuances of analytics and supporting software, but you will, at the very least, be a consumer of analytics and work with analytics professionals to support your analyses and decisions. For example, you might find yourself on project teams with managers who know very little about analytics and analytics experts such as statisticians, programmers, and economists. Your role might be to ensure that analytics is used properly to solve important business problems.

Impacts and Challenges

The benefits of applying business analytics can be significant. Companies report reduced costs, better risk management, faster decisions, better productivity, and enhanced bottom-line performance such as profitability and customer satisfaction. For example, 1-800-Flowers.com used analytic software to target print and online promotions with greater accuracy; change prices and offerings on its Web site (sometimes hourly); and optimize its marketing, shipping, distribution, and manufacturing operations, resulting in a $50 million cost savings in one year.[7]

Business analytics is changing how managers make decisions.[8] To thrive in today’s business world, organizations must continually innovate to differentiate themselves from competitors, seek ways to grow revenue and market share, reduce costs, retain existing customers and acquire new ones, and become faster and leaner. IBM suggests that traditional management

[5]Thomas H. Davenport, “How Organizations Make Better Decisions,” edited excerpt of an article distributed by the International Institute for Analytics published in “Brain Trust—Enabling the Confident Enterprise with Business Analytics” (Cary, NC: SAS Institute, Inc., 2010): 8–11. www.sas.com/bareport

[6]Thomas H. Davenport and Jeanne G. Harris, *Competing on Analytics* (Boston: Harvard Business School Press, 2007): 46; Michael S. Hopkins, Steve LaValle, Fred Balboni, Nina Kruschwitz, and Rebecca Shockley, “10 Data Points: Information and Analytics at Work,” *MIT Sloan Management Review*, 52, 1 (Fall 2010): 27–31.

[7]Jim Goodnight, “The Impact of Business Analytics on Performance and Profitability,” in “Brain Trust—Enabling the Confident Enterprise with Business Analytics” (Cary, NC: SAS Institute, Inc., 2010): 4–7. www.sas.com/bareport

[8]*Analytics: The New Path to Value*, a joint MIT Sloan Management Review and IBM Institute for Business Value study.

approaches are evolving in today's analytics-driven environment to include more fact-based decisions as opposed to judgment and intuition, more prediction rather than reactive decisions, and the use of analytics by everyone at the point where decisions are made rather than relying on skilled experts in a consulting group.[9] Nevertheless, organizations face many challenges in developing analytics capabilities, including lack of understanding of how to use analytics, competing business priorities, insufficient analytical skills, difficulty in getting good data and sharing information, and not understanding the benefits versus perceived costs of analytics studies. Successful application of analytics requires more than just knowing the tools; it requires a high-level understanding of how analytics supports an organization's competitive strategy and effective execution that crosses multiple disciplines and managerial levels.

In 2011, a survey by Bloomberg Businessweek Research Services and SAS concluded that business analytics was still in the "emerging stage" and was used only narrowly within business units, not across entire organizations. The study also noted that many organizations lacked analytical talent, and those that did have analytical talent often didn't know how to apply the results properly. While analytics was used as part of the decision-making process in many organizations, most business decisions are still based on intuition.[10] Today, business analytics has matured in many organizations, but many more opportunities still exist. These opportunities are reflected in the job market for analytics professionals, or "data scientists," as some call them. McKinsey & Company suggested that there is a shortage of qualified data scientists.[11]

CHECK YOUR UNDERSTANDING

1. Explain why analytics is important in today's business environment.
2. Define business analytics.
3. State three examples of how business analytics is used in organizations.
4. What are the key benefits of using business analytics?
5. What challenges do organizations face in using analytics?

Evolution of Business Analytics

Analytical methods, in one form or another, have been used in business for more than a century. The core of business analytics consists of three disciplines: business intelligence and information systems, statistics, and modeling and optimization.

Analytic Foundations

The modern evolution of analytics began with the introduction of computers in the late 1940s and their development through the 1960s and beyond. Early computers provided the ability to store and analyze data in ways that were either very difficult or impossible to do manually. This facilitated the collection, management, analysis, and reporting of data, which

[9]"Business Analytics and Optimization for the Intelligent Enterprise" (April 2009). www.ibm.com/qbs/intelligent-enterprise

[10]Bloomberg Businessweek Research Services and SAS, "The Current State of Business Analytics: Where Do We Go From Here?" (2011).

[11]Andrew Jennings, "What Makes a Good Data Scientist?" *Analytics Magazine* (July–August 2013): 8–13. www.analytics-magazine.org

is often called **business intelligence (BI)**, a term that was coined in 1958 by an IBM researcher, Hans Peter Luhn.[12] Business intelligence software can answer basic questions such as "How many units did we sell last month?" "What products did customers buy and how much did they spend?" "How many credit card transactions were completed yesterday?" Using BI, we can create simple rules to flag exceptions automatically; for example, a bank can easily identify transactions greater than $10,000 to report to the Internal Revenue Service.[13] BI has evolved into the modern discipline we now call **information systems (IS)**.

Statistics has a long and rich history, yet only rather recently has it been recognized as an important element of business, driven to a large extent by the massive growth of data in today's world. Google's chief economist noted that statisticians surely have one of the best jobs.[14] Statistical methods allow us to gain a richer understanding of data that goes beyond business intelligence reporting by not only summarizing data succinctly but also finding unknown and interesting relationships among the data. Statistical methods include the basic tools of description, exploration, estimation, and inference, as well as more advanced techniques like regression, forecasting, and data mining.

Much of modern business analytics stems from the analysis and solution of complex decision problems using mathematical or computer-based models—a discipline known as operations research, or management science. *Operations research (OR)* was born from efforts to improve military operations prior to and during World War II. After the war, scientists recognized that the mathematical tools and techniques developed for military applications could be applied successfully to problems in business and industry. A significant amount of research was carried on in public and private think tanks during the late 1940s and through the 1950s. As the focus on business applications expanded, the term *management science (MS)* became more prevalent. Many people use the terms *operations research* and *management science* interchangeably, so the field became known as **Operations Research/Management Science (OR/MS)**. Many OR/MS applications use **modeling and optimization**—techniques for translating real problems into mathematics, spreadsheets, or various computer languages, and using them to find the best ("optimal") solutions and decisions. INFORMS, the Institute for Operations Research and the Management Sciences, is the leading professional society devoted to OR/MS and analytics and publishes a bimonthly magazine called *Analytics* (http://analytics-magazine.org/). Digital subscriptions may be obtained free of charge at the Web site.

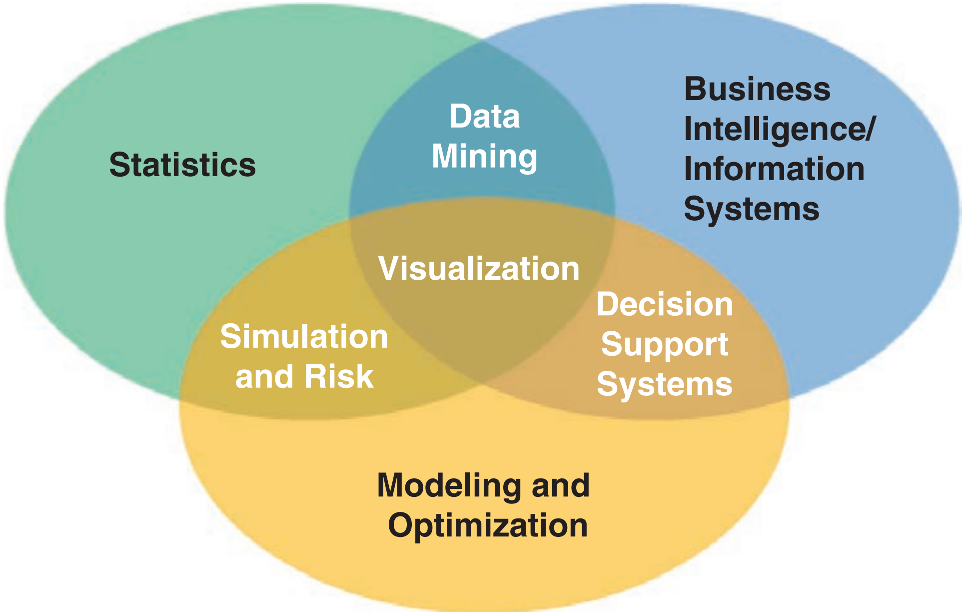
Modern Business Analytics

Modern business analytics can be viewed as an integration of BI/IS, statistics, and modeling and optimization, as illustrated in Figure 1.1. While these core topics are traditional and have been used for decades, the uniqueness lies in their intersections. For example, **data mining** is focused on better understanding characteristics and patterns among variables in large databases using a variety of statistical and analytical tools. Many standard statistical tools as well as more advanced ones are used extensively in data mining. **Simulation and risk analysis** relies on spreadsheet models and statistical analysis to examine the impacts of uncertainty in estimates and their potential interaction with one another on the output variable of interest.

[12] H. P. Luhn, "A Business Intelligence System." *IBM Journal* (October 1958).
[13] Jim Davis, "Business Analytics: Helping You Put an Informed Foot Forward," in "Brain Trust—Enabling the Confident Enterprise with Business Analytics," (Cary, NC: SAS Institute, Inc., 2010): 4–7. www.sas.com/bareport
[14] James J. Swain, "Statistical Software in the Age of the Geek," *Analytics Magazine* (March -April 2013): 48–55.

Figure 1.1

A Visual Perspective of Business Analytics



Decision support systems (DSSs) began to evolve in the 1960s by combining business intelligence concepts with OR/MS models to create analytical-based computer systems to support decision making. DSSs include three components:

1. *Data management*. The data management component includes databases for storing data and allows the user to input, retrieve, update, and manipulate data.
2. *Model management*. The model management component consists of various statistical tools and management science models and allows the user to easily build, manipulate, analyze, and solve models.
3. *Communication system*. The communication system component provides the interface necessary for the user to interact with the data and model management components.[15]

DSSs have been used for many applications, including pension fund management, portfolio management, work-shift scheduling, global manufacturing and facility location, advertising-budget allocation, media planning, distribution planning, airline operations planning, inventory control, library management, classroom assignment, nurse scheduling, blood distribution, water pollution control, ski-area design, police-beat design, and energy planning.[16]

A key feature of a DSS is the ability to perform **what-if analysis**—how specific combinations of inputs that reflect key assumptions will affect model outputs. What-if analysis is also used to assess the sensitivity of optimization models to changes in data inputs and provide better insight for making good decisions.

Perhaps the most useful component of business analytics, which makes it truly unique, is the center of Figure 1.1—**visualization**. Visualizing data and results of analyses provides a way of easily communicating data at all levels of a business and can reveal surprising patterns and relationships. Software such as IBM's Cognos system exploits data visualization

[15]William E. Leigh and Michael E. Doherty, *Decision Support and Expert Systems* (Cincinnati, OH: South-Western Publishing Co., 1986).
[16]H. B. Eom and S. M. Lee, "A Survey of Decision Support System Applications (1971–April 1988)," *Interfaces*, 20, 3 (May–June 1990): 65–79.

for query and reporting, data analysis, dashboard presentations, and scorecards linking strategy to operations. The Cincinnati Zoo, for example, has used this on an iPad to display hourly, daily, and monthly reports of attendance, food and retail location revenues and sales, and other metrics for prediction and marketing strategies. UPS uses telematics to capture vehicle data and display them to help make decisions to improve efficiency and performance. You may have seen a **tag cloud** (see the graphic at the beginning of this chapter), which is a visualization of text that shows words that appear more frequently with larger fonts.

Software Support and Spreadsheet Technology

Many companies, such as IBM, SAS, and Tableau Software, have developed a variety of software and hardware solutions to support business analytics. For example, IBM's Cognos Express, an integrated business intelligence and planning solution designed to meet the needs of midsize companies, provides reporting, analysis, dashboard, scorecard, planning, budgeting, and forecasting capabilities. It is made up of several modules, including Cognos Express Reporter, for self-service reporting and ad hoc query; Cognos Express Advisor, for analysis and visualization; and Cognos Express Xcelerator, for Excel-based planning and business analysis. Information is presented to users in a context that makes it easy to understand; with an easy-to-use interface, users can quickly gain the insight they need from their data to make the right decisions and then take action for effective and efficient business optimization and outcome. SAS provides a variety of software that integrate data management, business intelligence, and analytics tools. SAS Analytics covers a wide range of capabilities, including predictive modeling and data mining, visualization, forecasting, optimization and model management, statistical analysis, text analytics, and more. Tableau Software provides simple drag and drop tools for visualizing data from spreadsheets and other databases. We encourage you to explore many of these products as you learn the basic principles of business analytics in this book.

Although commercial software often have powerful features and capabilities, they can be expensive, generally require advanced training to understand and apply, and may work only on specific computer platforms. Spreadsheet software, on the other hand, is widely used across all areas of business and used by nearly everyone. Spreadsheets are an effective platform for manipulating data and developing and solving models; they support powerful commercial add-ins and facilitate communication of results. Spreadsheets provide a flexible modeling environment and are particularly useful when the end user is not the designer of the model. Teams can easily use spreadsheets and understand the logic upon which they are built. Information in spreadsheets can easily be copied from spreadsheets into other documents and presentations. A recent survey identified more than 180 commercial spreadsheet products that support analytics efforts, including data management and reporting, data- and model-driven analytical techniques, and implementation.[17] Many organizations have used spreadsheets extremely effectively to support decision making in marketing, finance, and operations. Some illustrative applications include the following:[18]

- Analyzing supply chains (Hewlett-Packard)
- Determining optimal inventory levels to meet customer service objectives (Procter & Gamble)

[17]Thomas A. Grossman, "Resources for Spreadsheet Analysts," *Analytics Magazine* (May/June 2010): 8. www.analytics-magazine.org

[18]Larry J. LeBlanc and Thomas A. Grossman, "Introduction: The Use of Spreadsheet Software in the Application of Management Science and Operations Research," *Interfaces*, 38, 4 (July–August 2008): 225–227.

- Selecting internal projects (Lockheed Martin Space Systems)
- Planning for emergency clinics in response to a sudden epidemic or bioterrorism attack (Centers for Disease Control)
- Analyzing the default risk of a portfolio of real estate loans (Hypo International)
- Assigning medical residents to on-call and emergency rotations (University of Vermont College of Medicine)
- Performance measurement and evaluation (American Red Cross)

Some optional software packages for statistical applications that your instructor might use are SAS, Minitab, *XLSTAT* and *StatCrunch*. These provide many powerful procedures as alternatives or supplements to Excel.

Spreadsheet technology has been influential in promoting the use and acceptance of business analytics. Spreadsheets provide a convenient way to manage data, calculations, and visual graphics simultaneously, using intuitive representations instead of abstract mathematical notation. Although the early applications of spreadsheets were primarily in accounting and finance, spreadsheets have developed into powerful general-purpose managerial tools for applying techniques of business analytics. The power of analytics in a personal computing environment was noted decades ago by business consultants Michael Hammer and James Champy, who said, "When accessible data is combined with easy-to-use analysis and modeling tools, frontline workers—when properly trained—suddenly have sophisticated decision-making capabilities."[19]

ANALYTICS IN PRACTICE: Social Media Analytics

One of the emerging applications of analytics is helping businesses learn from social media and exploit social media data for strategic advantage.[20] Using analytics, firms can integrate social media data with traditional data sources such as customer surveys, focus groups, and sales data; understand trends and customer perceptions of their products; and create informative reports to assist marketing managers and product designers.

Social media analytics is useful in decision making in many business domains to understand how user-generated content spreads and influences user interactions, how information is transmitted, and how it influences decisions. A review of research published in social media analytics provides numerous examples:[21]

- The analysis of public responses from social media before, during, and after disasters, such as the 2010 Haiti earthquake and Hurricane Sandy in New York City in 2012, has the potential to improve situational knowledge in emergency and disaster management practices.
- Social media platforms enable citizens' engagement with politicians, governments, and other citizens. Studies have examined how voters discuss the candidates during an election, how candidates are adopting Twitter for campaigning and influencing conversations in the public space, and how presidential candidates in the United States used Twitter to engage people and identify the topics mentioned by candidates during their campaigns. Others have used analytics to track political preference by monitoring online popularity.
- In the entertainment industry, one study analyzed viewer ratings to predict the impact on revenue for upcoming movies. Another developed a web intelligence application to aggregate the news about popular TV serials and identify emerging storylines.

[19] Michael Hammer and James Champy, *Reengineering the Corporation* (New York: HarperBusiness, 1993): 96.

[20] Jim Davis, "Convergence—Taking Social Media from Talk to Action," SASCOM (First Quarter 2011): 17.

[21] Ashish K. Rathore, Arpan K. Kar, and P. Vigneswara Ilavarasana, "Social Media Analytics: Literature Review and Directions for Future Research," *Decision Analysis*, 14, 4 (December 2017): 229–249.

- Retail organizations monitor and analyze social media data about their own products and services and also about their competitors' products and services to stay competitive. For instance, one study analyzed different product features based on rankings from users' online reviews.
- The integration of social media application and health care leads to better patient management and empowerment. One researcher classified various online health communities, such as a diabetes patients' community, using posts from WebMD.com. Another analyzed physical activity–related tweets for a better understanding of physical activity behaviors. To predict the spread of influenza, one researcher developed a forecasting approach using flu-related tweets.

In this book, we use Microsoft Excel as the primary platform for implementing analytics. In the Chapter 1 Appendix, we review some key Excel skills that you should have before moving forward in this book.

The main chapters in this book are designed using Excel 2016 for Windows or Excel 2016 for Mac. Earlier versions of Excel do not have all the capabilities that we use in this book. In addition, some key differences exist between Windows and Mac versions that we will occasionally point out. Thus, some Excel tools that we will describe in chapter appendixes require you to use Excel for Windows, Office 365, or Google Sheets, and will not run on Excel for Mac; these are optional to learn, and are not required for any examples or problems. Your instructor may use optional software, such as XLSTAT and StatCrunch, which are provided by the publisher (Pearson), or Analytic Solver, which is described in online supplements to this book.

CHECK YOUR UNDERSTANDING

1. Provide two examples of questions that business intelligence can address.
2. How do statistical methods enhance business intelligence reporting?
3. What is operations research/management science?
4. How does modern business analytics integrate traditional disciplines of BI, statistics, and modeling/optimization?
5. What are the components of a decision support system?

Descriptive, Predictive, and Prescriptive Analytics

Business analytics begins with the collection, organization, and manipulation of data and is supported by three major components:[22]

1. *Descriptive analytics*. Most businesses start with **descriptive analytics**—the use of data to understand past and current business performance and make informed decisions. Descriptive analytics is the most commonly used and most well-understood type of analytics. These techniques categorize, characterize, consolidate, and classify data to convert them into useful information for the purposes of understanding and analyzing business performance. Descriptive

[22]Parts of this section are adapted from Irv Lustig, Brenda Dietric, Christer Johnson, and Christopher Dziekan, "The Analytics Journey," *Analytics* (November/December 2010). http://analytics-magazine.org/novemberdecember-2010-table-of-contents/

analytics summarizes data into meaningful charts and reports, for example, about budgets, sales, revenues, or cost. This process allows managers to obtain standard and customized reports and then drill down into the data and make queries to understand the impact of an advertising campaign, such as reviewing business performance to find problems or areas of opportunity, and identifying patterns and trends in data. Typical questions that descriptive analytics helps answer are "How much did we sell in each region?" "What was our revenue and profit last quarter?" "How many and what types of complaints did we resolve?" "Which factory has the lowest productivity?" Descriptive analytics also helps companies to classify customers into different segments, which enables them to develop specific marketing campaigns and advertising strategies.

2. *Predictive analytics*. **Predictive analytics** seeks to predict the future by examining historical data, detecting patterns or relationships in these data, and then extrapolating these relationships forward in time. For example, a marketer might wish to predict the response of different customer segments to an advertising campaign, a commodities trader might wish to predict short-term movements in commodities prices, or a skiwear manufacturer might want to predict next season's demand for skiwear of a specific color and size. Predictive analytics can predict risk and find relationships in data not readily apparent with traditional analyses. Using advanced techniques, predictive analytics can help detect hidden patterns in large quantities of data, and segment and group data into coherent sets to predict behavior and detect trends. For instance, a bank manager might want to identify the most profitable customers, predict the chances that a loan applicant will default, or alert a credit card customer to a potential fraudulent charge. Predictive analytics helps to answer questions such as "What will happen if demand falls by 10% or if supplier prices go up 5%?" "What do we expect to pay for fuel over the next several months?" "What is the risk of losing money in a new business venture?"
3. *Prescriptive analytics*. Many problems, such as aircraft or employee scheduling and supply chain design, simply involve too many choices or alternatives for a human decision maker to effectively consider. **Prescriptive analytics** uses optimization to identify the best alternatives to minimize or maximize some objective. Prescriptive analytics is used in many areas of business, including operations, marketing, and finance. For example, we may determine the best pricing and advertising strategy to maximize revenue, the optimal amount of cash to store in ATMs, or the best mix of investments in a retirement portfolio to manage risk. Prescriptive analytics addresses questions such as "How much should we produce to maximize profit?" "What is the best way of shipping goods from our factories to minimize costs?" "Should we change our plans if a natural disaster closes a supplier's factory, and if so, by how much?" The mathematical and statistical techniques of predictive analytics can also be combined with prescriptive analytics to make decisions that take into account the uncertainty in the data.

A wide variety of tools are used to support business analytics. These include

- Database queries and analysis
- "Dashboards" to report key performance measures
- Data visualization
- Statistical methods
- Spreadsheets and predictive models

ANALYTICS IN PRACTICE: Analytics in the Home Lending and Mortgage Industry[23]

Sometime during their lives, most Americans will receive a mortgage loan for a house or condominium. The process starts with an application. The application contains all pertinent information about the borrower that the lender will need. The bank or mortgage company then initiates a process that leads to a loan decision. It is here that key information about the borrower is provided by third-party providers. This information includes a credit report, verification of income, verification of assets, verification of employment, and an appraisal of the property. The result of the processing function is a complete loan file that contains all the information and documents needed to underwrite the loan, which is the next step in the process. Underwriting is where the loan application is evaluated for its risk. Underwriters evaluate whether the borrower can make payments on time, can afford to pay back the loan, and has sufficient collateral in the property to back up the loan. In the event the borrower defaults on their loan, the lender can sell the property to recover the amount of the loan. But if the amount of the loan is greater than the value of the property, then the lender cannot recoup their money. If the underwriting process indicates that the borrower is creditworthy and has the capacity to repay the loan and the value of the property in question is greater than the loan amount, then the loan is approved and will move to closing. Closing is the step where the borrower signs all the appropriate papers, agreeing to the terms of the loan.

In reality, lenders have a lot of other work to do. First, they must perform a quality control review on a sample of the loan files that involves a manual examination of all the documents and information gathered. This process is designed to identify any mistakes that may have been made or information that is missing from the loan file. Because lenders do not have unlimited money to lend to borrowers, they frequently sell the loan to a third party so that they have fresh capital to lend to others. This occurs in what is called the secondary market. Freddie Mac and Fannie Mae are the two largest purchasers of mortgages in the secondary market. The final step in the process is servicing. Servicing includes all the activities associated with providing the customer service on the loan, like processing payments, managing property taxes held in escrow, and answering questions about the loan.

In addition, the institution collects various operational data on the process to track its performance and efficiency, including the number of applications, loan types and amounts, cycle times (time to close the loan), bottlenecks in the process, and so on. Many different types of analytics are used:

Descriptive analytics—This focuses on historical reporting, addressing such questions as

- How many loan applications were taken in each of the past 12 months?
- What was the total cycle time from application to close?
- What was the distribution of loan profitability by credit score and loan-to-value (LTV), which is the mortgage amount divided by the appraised value of the property?

Predictive analytics—Predictive modeling uses mathematical, spreadsheet, and statistical models and addresses questions such as

- What impact on loan volume will a given marketing program have?
- How many processors or underwriters are needed for a given loan volume?
- Will a given process change reduce cycle time?

Prescriptive analytics—This involves the use of simulation or optimization to drive decisions. Typical questions include

- What is the optimal staffing to achieve a given profitability constrained by a fixed cycle time?
- What is the optimal product mix to maximize profit constrained by fixed staffing?

The mortgage market has become much more dynamic in recent years due to rising home values, falling interest rates, new loan products, and an increased desire by home owners to utilize the equity in their homes as a financial resource. This has increased the complexity and variability of the mortgage process and created an opportunity for lenders to proactively use the data that are available to them as a tool for managing their business. To ensure that the process is efficient, effective, and performed with quality, data and analytics are used every day to track what is done, who is doing it, and how long it takes.

[23]Contributed by Craig Zielazny, BlueNote Analytics, LLC.

- Scenario and "what-if" analyses
- Simulation
- Forecasting
- Data and text mining
- Optimization
- Social media, Web, and text analytics

Although the tools used in descriptive, predictive, and prescriptive analytics are different, many applications involve all three. Here is a typical example in retail operations.

EXAMPLE 1.1 Retail Markdown Decisions[24]

As you probably know from your shopping experiences, most department stores and fashion retailers clear their seasonal inventory by reducing prices. The key question they face is what prices should they set—and when should they set them—to meet inventory goals and maximize revenue? For example, suppose that a store has 100 bathing suits of a certain style that go on sale on April 1 and wants to sell all of them by the end of June. Over each week of the 12-week selling season, they can make a decision to discount the price. They face two decisions: When to reduce the price, and by how much. This results in 24 decisions to make. For a major national chain that may carry thousands of products, this can easily result in millions of decisions that store managers have to make. Descriptive analytics can be used to examine historical data for similar products, such as the number of units sold, price at each point of sale, starting and ending inventories, and special promotions, newspaper ads, direct marketing ads, and so on, to understand what the results of past decisions achieved. Predictive analytics can be used to predict sales based on pricing decisions. Finally, prescriptive analytics can be applied to find the best set of pricing decisions to maximize the total revenue.

CHECK YOUR UNDERSTANDING

1. Define descriptive analytics and provide two examples.
2. Define predictive analytics and provide two examples.
3. Define prescriptive analytics and provide two examples.

Data for Business Analytics

Since the dawn of the electronic age and the Internet, both individuals and organizations have had access to an enormous wealth of data and information. Most data are collected through some type of measurement process, and consist of numbers (e.g., sales revenues) or textual data (e.g., customer demographics such as gender). Other data might be extracted from social media, online reviews, and even audio and video files. *Information* comes from analyzing data—that is, extracting meaning from data to support evaluation and decision making.

Data are used in virtually every major function in a business. Modern organizations—which include not only for-profit businesses but also nonprofit organizations—need good data to support a variety of company purposes, such as planning, reviewing company performance, improving operations, and comparing company performance with competitors'

[24]Inspired by a presentation by Radhika Kulkarni, SAS Institute, "Data-Driven Decisions: Role of Operations Research in Business Analytics," INFORMS Conference on Business Analytics and Operations Research, April 10–12, 2011.

or best-practice benchmarks. Some examples of how data are used in business include the following:

- Annual reports summarize data about companies' profitability and market share both in numerical form and in charts and graphs to communicate with shareholders.
- Accountants conduct audits to determine whether figures reported on a firm's balance sheet fairly represent the actual data by examining samples (that is, subsets) of accounting data, such as accounts receivable.
- Financial analysts collect and analyze a variety of data to understand the contribution that a business provides to its shareholders. These typically include profitability, revenue growth, return on investment, asset utilization, operating margins, earnings per share, economic value added (EVA), shareholder value, and other relevant measures.
- Economists use data to help companies understand and predict population trends, interest rates, industry performance, consumer spending, and international trade. Such data are often obtained from external sources such as Standard & Poor's Compustat data sets, industry trade associations, or government databases.
- Marketing researchers collect and analyze extensive customer data. These data often consist of demographics, preferences and opinions, transaction and payment history, shopping behavior, and much more. Such data may be collected by surveys, personal interviews, or focus groups, or from shopper loyalty cards.
- Operations managers use data on production performance, manufacturing quality, delivery times, order accuracy, supplier performance, productivity, costs, and environmental compliance to manage their operations.
- Human resource managers measure employee satisfaction, training costs, turnover, market innovation, training effectiveness, and skills development.

Data may be gathered from primary sources such as internal company records and business transactions, automated data-capturing equipment, and customer market surveys and from secondary sources such as government and commercial data sources, custom research providers, and online research.

Perhaps the most important source of data today is data obtained from the Web. With today's technology, marketers collect extensive information about Web behaviors, such as the number of page views, visitor's country, time of view, length of time, origin and destination paths, products they searched for and viewed, products purchased, and what reviews they read. Using analytics, marketers can learn what content is being viewed most often, what ads were clicked on, who the most frequent visitors are, and what types of visitors browse but don't buy. Not only can marketers understand what customers have done, but they can better predict what they intend to do in the future. For example, if a bank knows that a customer has browsed for mortgage rates and homeowner's insurance, they can target the customer with homeowner loans rather than credit cards or automobile loans. Traditional Web data are now being enhanced with social media data from Facebook, cell phones, and even Internet-connected gaming devices.

As one example, a home furnishings retailer wanted to increase the rate of sales for customers who browsed their Web site. They developed a large data set that covered more than 7,000 demographic, Web, catalog, and retail behavioral attributes for each customer. They used predictive analytics to determine how well a customer would respond to different e-mail marketing offers and customized promotions to individual customers. This not only helped them to determine where to most effectively spend marketing resources but also doubled the response rate compared to previous marketing campaigns, with a projected and multimillion dollar increase in sales.[25]

Big Data

Today, nearly all data are captured digitally. As a result, data have been growing at an overwhelming rate, being measured by terabytes (10^{12} bytes), petabytes (10^{15} bytes), exabytes (10^{18} bytes), and even by higher-dimensional terms. Just think of the amount of data stored on Facebook, Twitter, or Amazon servers, or the amount of data acquired daily from scanning items at a national grocery chain such as Kroger and its affiliates. Walmart, for instance, has over one million transactions each hour, yielding more than 2.5 petabytes of data. Analytics professionals have coined the term **big data** to refer to massive amounts of business data from a wide variety of sources, much of which is available in real time. IBM calls these characteristics *volume*, *variety*, and *velocity*. Most often, big data revolve around customer behavior and customer experiences. Big data provide an opportunity for organizations to gain a competitive advantage—if the data can be understood and analyzed effectively to make better business decisions.

The volume of data continues to increase; what is considered "big" today will be even bigger tomorrow. In one study of information technology (IT) professionals in 2010, nearly half of survey respondents ranked data growth among their top three challenges. Big data are captured using sensors (for example, supermarket scanners), click streams from the Web, customer transactions, e-mails, tweets and social media, and other ways. Big data sets are unstructured and messy, requiring sophisticated analytics to integrate and process the data and understand the information contained in them. Because much big data are being captured in real time, they must be incorporated into business decisions at a faster rate. Processes such as fraud detection must be analyzed quickly to have value. In addition to *volume*, *variety*, and *velocity*, IBM proposed a fourth dimension: *veracity*—the level of reliability associated with data. Having high-quality data and understanding the uncertainty in data are essential for good decision making. Data veracity is an important role for statistical methods.

Big data can help organizations better understand and predict customer behavior and improve customer service. A study by the McKinsey Global Institute noted that, "The effective use of big data has the potential to transform economies, delivering a new wave of productivity growth and consumer surplus. Using big data will become a key basis of competition for existing companies, and will create new competitors who are able to attract employees that have the critical skills for a big data world."[26] However, understanding big data requires advanced analytics tools such as data mining and text analytics, and new technologies such as cloud computing, faster multi-core processors, large memory spaces, and solid-state drives.

Data Reliability and Validity

Poor data can result in poor decisions. In one situation, a distribution system design model relied on data obtained from the corporate finance department. Transportation costs were

[25]Based on a presentation by Bill Franks of Teradata, "Optimizing Customer Analytics: How Customer Level Web Data Can Help," INFORMS Conference on Business Analytics and Operations Research, April 10–12, 2011.

[26]James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey & Company May 2011.

determined using a formula based on the latitude and longitude of the locations of plants and customers. But when the solution was represented on a geographic information system (GIS) mapping program, one of the customers was located in the Atlantic Ocean.

Thus, data used in business decisions need to be reliable and valid. **Reliability** means that data are accurate and consistent. **Validity** means that data correctly measure what they are supposed to measure. For example, a tire pressure gauge that consistently reads several pounds of pressure below the true value is not reliable, although it is valid because it does measure tire pressure. The number of calls to a customer service desk might be counted correctly each day (and thus is a reliable measure), but it is not valid if it is used to assess customer dissatisfaction, as many calls may be simple queries. Finally, a survey question that asks a customer to rate the quality of the food in a restaurant may be neither reliable (because different customers may have conflicting perceptions) nor valid (if the intent is to measure customer satisfaction, as satisfaction generally includes other elements of service besides food).

CHECK YOUR UNDERSTANDING

1. State three examples of how data are used in different business functions.
2. How are data obtained from the Web used in marketing and business?
3. Define big data and list the four characteristics of big data.
4. Explain the concepts of data reliability and validity.

Models in Business Analytics

To make an informed decision, we must be able to specify the decision alternatives that represent the choices that can be made and criteria for evaluating the alternatives. Specifying decision alternatives might be very simple; for example, you might need to choose one of three corporate health plan options. Other situations can be more complex; for example, in locating a new distribution center, it might not be possible to list just a small number of alternatives. The set of potential locations might be anywhere in the United States or even across the globe. Decision criteria might be to maximize discounted net profits, customer satisfaction, or social benefits or to minimize costs, environmental impact, or some measure of loss.

Many decision problems can be formalized using a model. A **model** is an abstraction or representation of a real system, idea, or object. Models capture the most important features of a problem and present them in a form that is easy to interpret. A model can be as simple as a written or verbal description of some phenomenon, a visual representation such as a graph or a flowchart, or a mathematical or spreadsheet representation. Example 1.2 illustrates three ways to express a model.

A **decision model** is a logical or mathematical representation of a problem or business situation that can be used to understand, analyze, or facilitate making a decision. Decision models can be represented in various ways, most typically with mathematical functions and spreadsheets. Spreadsheets are ideal vehicles for implementing decision models because of their versatility in managing data, evaluating different scenarios, and presenting results in a meaningful fashion. We will focus on spreadsheet models beginning with Chapter 11.

EXAMPLE 1.2 Three Forms of a Model

Models are usually developed from theory or observation and establish relationships between actions that decision makers might take and results that they might expect, thereby allowing the decision makers to evaluate scenarios or to predict what might happen. For example, the sales of a new product, such as a first-generation iPad, Android phone, or 3-D television, often follow a common pattern. We might represent this in one of the three following ways:

1. A simple verbal description of sales might be: *The rate of sales starts small as early adopters begin to evaluate a new product and then begins to grow at an increasing rate over time as positive customer feedback spreads. Eventually, the market begins to become saturated, and the rate of sales begins to decrease.*
2. A sketch of sales as an S-shaped curve over time, as shown in Figure 1.2, is a visual model that conveys this phenomenon.
3. Finally, analysts might identify a mathematical model that characterizes this curve. Several different mathematical functions do this; one is called a Gompertz curve and has the formula: $S = ae^{be^{ct}}$, where S = sales, t = time, e is the base of natural logarithms, and a, b, and c are constants. Of course, you would not be expected to know this; that's what analytics professionals do.

Decision models typically have three types of input:

1. *Data*, which are assumed to be constant for purposes of the model. Some examples are costs, machine capacities, and intercity distances.
2. *Uncontrollable inputs*, which are quantities that can change but cannot be directly controlled by the decision maker. Some examples are customer demand, inflation rates, and investment returns. Often, these variables are uncertain.
3. *Decision options*, which are controllable and can be selected at the discretion of the decision maker. Some examples are production quantities, staffing levels, and investment allocations. Decision options are often called **decision variables**.

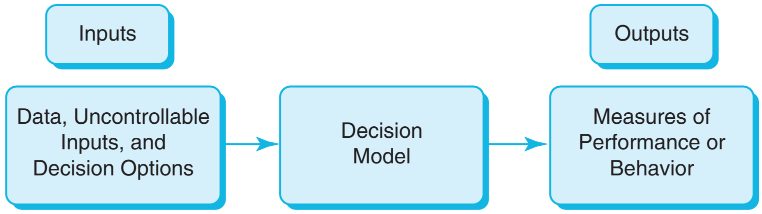
Decision models characterize the relationships among these inputs and the outputs of interest to the decision maker (see Figure 1.3). In this way, the user can manipulate the decision options and understand how they influence outputs, make predictions for the future, or use analytical tools to find the best decisions. Thus, decision models can be descriptive, predictive, or prescriptive and therefore are used in a wide variety of business analytics applications.

Figure 1.2

New Product Sales over Time

Figure 1.3

The Nature of Decision Models



Decision models complement decision makers' intuition and often provide insights that intuition cannot. For example, one early application of analytics in marketing involved a study of sales operations. Sales representatives had to divide their time between large and small customers and between acquiring new customers and keeping old ones. The problem was to determine how the representatives should best allocate their time. Intuition suggested that they should concentrate on large customers and that it was much harder to acquire a new customer than to keep an old one. However, intuition could not tell whether they should concentrate on the 100 largest or the 1,000 largest customers, or how much effort to spend on acquiring new customers. Models of sales force effectiveness and customer response patterns provided the insight to make these decisions. However, it is important to understand that all models are only representations of the real world and, as such, cannot capture every nuance that decision makers face in reality. Decision makers must often modify the policies that models suggest to account for intangible factors that they might not have been able to incorporate into the model.

Descriptive Models

Descriptive models explain behavior and allow users to evaluate potential decisions by asking "what-if?" questions. The following example illustrates a simple descriptive mathematical model.

EXAMPLE 1.3 Gasoline Usage Model

Automobiles have different fuel economies (miles per gallon), and commuters drive different distances to work or school. Suppose that a state Department of Transportation (DOT) is interested in measuring the average monthly fuel consumption of commuters in a certain city. The DOT might sample a group of commuters and collect information on the number of miles driven per day, the number of driving days per month, the fuel economy of their vehicles, and additional miles driven per month for leisure and household activities. We may develop a simple descriptive model for calculating the amount of gasoline consumed, using the following symbols for the data:

G = gallons of fuel consumed per month
m = miles driven per day to and from work or school
d = number of driving days per month
f = fuel economy in miles per gallon (mpg)
a = additional miles for leisure and household activities per month

When developing mathematical models, it is very important to use the dimensions of the variables to ensure logical consistency. In this example, we see that

$$(m \text{ miles/day}) \times (d \text{ days/month}) = m \times d \text{ miles/month}$$

Thus, the total number of miles driven per month $= m \times d + a$. If the vehicle gets f miles/gallon, then the total number of gallons consumed per month is

$$\begin{aligned} G &= (m \times d + a \text{ miles/month})/(f \text{ miles/gallon}) \\ &= (m \times d + a)/f \text{ gallons/month} \qquad \textbf{(1.1)} \end{aligned}$$

Suppose that a commuter drives 30 miles round trip to work for 20 days each month, achieves a fuel economy of 34 mpg, and drives an additional 250 miles each month. Using formula (1.1), the number of gallons consumed is

$$G = (30 \times 20 + 250)/34 = 25.0 \text{ gallons/month}$$

In the previous example, we have no decision options; the model is purely descriptive, but allows us to evaluate "what-if?" questions, for example, "What if we purchase a hybrid vehicle with a fuel economy of 45 miles/gallon?" "What if leisure and household activity driving increases to 400 miles/month?" Most of the models we will be using include decision options. As an example, suppose that a manufacturer has the option of producing a part in house or outsourcing it from a supplier (the decision options). Should the firm produce the part or outsource it? The decision depends on the costs of manufacturing and outsourcing, as well as the anticipated volume of demand (the uncontrollable inputs). By developing a model to evaluate the total cost of both alternatives (the outputs), the best decision can be made.

EXAMPLE 1.4 An Outsourcing Decision Model

Suppose that a manufacturer can produce a part for \$125/unit with a fixed cost of \$50,000. The alternative is to outsource production to a supplier at a unit cost of \$175. The total manufacturing and outsourcing costs can be expressed by simple mathematical formulas, where Q is the production volume:

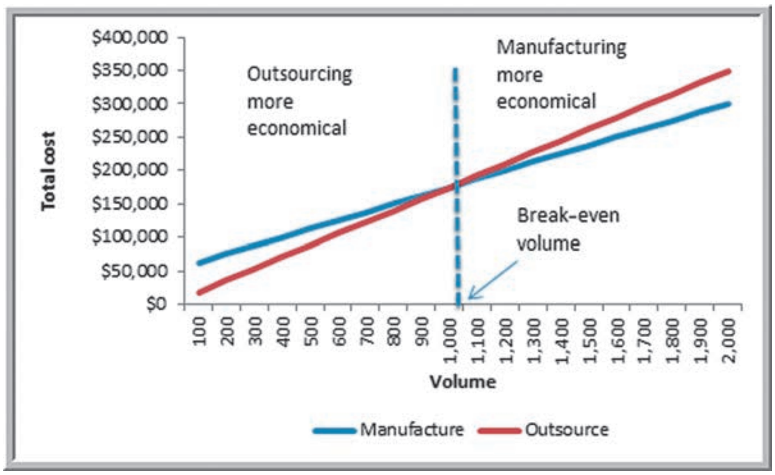
$$TC\text{ (manufacturing)} = \$50{,}000 + \$125 \times Q \quad \textbf{(1.2)}$$
$$TC\text{ (outsourcing)} = \$175 \times Q \quad \textbf{(1.3)}$$

These formulas comprise the decision model, which simply describes what the costs of manufacturing and outsourcing are for any level of production volume. Thus, if the anticipated production volume is 1,500 units, the cost of manufacturing will be $\$50{,}000 + \$125 \times 1{,}500 = \$237{,}500$, and the cost of outsourcing would be $\$175 \times 1{,}500 = \$262{,}500$; therefore, manufacturing would be the best decision. On the other hand, if the anticipated production volume is only 800 units, the cost of manufacturing will be $\$50{,}000 + \$125 \times 800 = \$150{,}000$ and the cost of outsourcing would be $\$175 \times 800 = \$140{,}000$, and the best decision would be to outsource. If we graph the two total cost formulas, we can easily see how the costs compare for different values of Q. This is shown graphically in Figure 1.4. The point at which the total costs of manufacturing and outsourcing are equal is called the break-even volume. This can easily be found by setting TC (manufacturing) = TC (outsourcing) and solving for Q:

$$\$50{,}000 + \$125 \times Q = \$175 \times Q$$
$$\$50{,}000 = 50 \times Q$$
$$Q = 1{,}000$$

▸ **Figure 1.4**

Graphical Illustration of Break-Even Analysis



Predictive Models

Predictive models focus on what will happen in the future. Many predictive models are developed by analyzing historical data and assuming that the past is representative of the future. Example 1.5 shows how historical data might be used to develop a model that can be used to predict the impact of pricing and promotional strategies in the grocery industry.[27]

EXAMPLE 1.5 A Predictive Sales-Promotion Model

In the grocery industry, managers typically need to know how best to use pricing, coupons, and advertising strategies to influence sales. Grocers often study the relationship of sales volume to these strategies by conducting controlled experiments. That is, they implement different combinations of pricing, coupons, and advertising, observe the sales that result, and use analytics to develop predictive models of sales as a function of these decision strategies.

For example, suppose that a grocer who operates three stores in a small city varied the price, coupons (yes = 1, no = 0), and advertising expenditures in a local newspaper over a 16-week period and observed the following sales:

Week	Price ($)	Coupon (0,1)	Advertising ($)	Store1 Sales (Units)	Store 2 Sales (Units)	Store 3 Sales (Units)
1	6.99	0	0	501	510	481
2	6.99	0	150	772	748	775
3	6.99	1	0	554	528	506
4	6.99	1	150	838	785	834
5	6.49	0	0	521	519	500
6	6.49	0	150	723	790	723
7	6.49	1	0	510	556	520
8	6.49	1	150	818	773	800
9	7.59	0	0	479	491	486
10	7.59	0	150	825	822	757
11	7.59	1	0	533	513	540
12	7.59	1	150	839	791	832
13	5.49	0	0	484	480	508
14	5.49	0	150	686	683	708
15	5.49	1	0	543	531	530
16	5.49	1	150	767	743	779

To better understand the relationships among price, coupons, and advertising, an analyst might have developed the following model using business analytics tools (we will see how to do this in Chapter 8):

$$\text{Total Sales} = 1105.55 + 56.18 \times \text{Price} + 123.88 \times \text{Coupon} + 5.24 \times \text{Advertising} \quad \textbf{(1.4)}$$

In this example, the uncontrollable inputs are the sales at each store. The decision options are price, coupons, and advertising. The numerical values in the model are estimated from the data obtained from the experiment. They reflect the impact on sales of changing the decision options. For example, an increase in price of $1 results in a 56.18-unit

(continued)

[27]Roger J. Calantone, Cornelia Droge, David S. Litvack, and C. Anthony di Benedetto. “Flanking in a Price War,” *Interfaces*, 19, 2 (1989): 1–12.

increase in weekly sales; using coupons (that is, setting Coupon = 1 in the model) results in a 123.88-unit increase in weekly sales. The output of the model is the predicted total sales units of the product. For example, if the price is \$6.99, no coupons are offered, and no advertising is done (the experiment corresponding to week 1), the model estimates sales as

$$\text{Total Sales} = 1{,}105.55 + 56.18 \times 6.99 + 123.88 \times 0 + 5.24 \times 0 = 1{,}498.25 \text{ units}$$

We see that the actual total sales in the three stores for week 1 was 1,492. Thus, this model appears to provide good estimates for sales using the historical data. We would hope that this model would also provide good predictions of future sales. So if the grocer decides to set the price at \$5.99, does not use coupons, and spends \$100 in advertising, the model would predict sales to be

$$\text{Total Sales} = 1{,}105.55 + 56.18 \times 5.99 + 123.88 \times 0 + 5.24 \times 100 = 1{,}966.07 \text{ units}$$

Prescriptive Models

A prescriptive decision model helps decision makers to identify the best solution to a decision problem. **Optimization** is the process of finding a set of values for decision options that minimize or maximize some quantity of interest—profit, revenue, cost, time, and so on—called the **objective function**. Any set of decision options that optimizes the objective function is called an **optimal solution**. In a highly competitive world, where one percentage point can mean a difference of hundreds of thousands of dollars or more, knowing the best solution can mean the difference between success and failure.

EXAMPLE 1.6 A Prescriptive Model for Pricing

To illustrate an example of a prescriptive model, suppose that a firm wishes to determine the best pricing for one of its products to maximize revenue over the next year. A market research study has collected data that estimate the expected annual sales for different levels of pricing. Analysts determined that sales can be expressed by the following model:

$$\text{Sales} = -2.9485 \times \text{Price} + 3{,}240.9 \quad \textbf{(1.5)}$$

Because revenue equals price × sales, a model for total revenue is

$$\begin{aligned}\text{Total Revenue} &= \text{Price} \times \text{Sales}\\ &= \text{Price} \times (-2.9485 \times \text{Price} + 3{,}240.9)\\ &= -2.9485 \times \text{Price}^2 + 3{,}240.9 \times \text{Price}\end{aligned} \quad \textbf{(1.6)}$$

The firm would like to identify the price that maximizes the total revenue. One way to do this would be to try different prices and search for the one that yields the highest total revenue. This would be quite tedious to do by hand or even with a calculator; however, as we will see in later chapters, spreadsheet models make this much easier.

Although the pricing model did not, most optimization models have **constraints**—limitations, requirements, or other restrictions that are imposed on any solution, such as "Do not exceed the allowable budget" or "Ensure that all demand is met." For instance, a consumer products company manager would probably want to ensure that a specified level of customer service is achieved with the redesign of the distribution system. The presence of constraints makes modeling and solving optimization problems more challenging; we address constrained optimization problems later in this book, starting in Chapter 13.

For some prescriptive models, analytical solutions—closed-form mathematical expressions or simple formulas—can be obtained using such techniques as calculus or other types of mathematical analyses. In most cases, however, some type of computer-based procedure is needed to find an optimal solution. An **algorithm** is a systematic procedure that finds a solution to a problem. Researchers have developed effective algorithms to solve many types of optimization problems. For example, Microsoft Excel has a built-in add-in called *Solver* that allows you to find optimal solutions to optimization problems formulated as spreadsheet models. We use *Solver* in later chapters. However, we will not be concerned with the detailed mechanics of these algorithms; our focus will be on the use of the algorithms to solve and analyze the models we develop.

If possible, we would like to ensure that an algorithm such as the one *Solver* uses finds the best solution. However, some models are so complex that it is impossible to solve them optimally in a reasonable amount of computer time because of the extremely large number of computations that may be required or because they are so complex that finding the best solution cannot be guaranteed. In these cases, analysts use **search algorithms**—solution procedures that generally find good solutions without guarantees of finding the best one. Powerful search algorithms exist to obtain good solutions to extremely difficult optimization problems. One of these is discussed in Chapter 14.

Model Assumptions

All models are based on assumptions that reflect the modeler's view of the "real world." Some assumptions are made to simplify the model and make it more tractable, that is, able to be easily analyzed or solved. Other assumptions might be made to better characterize historical data or past observations. The task of the modeler is to select or build an appropriate model that best represents the behavior of the real situation. For example, economic theory tells us that demand for a product is negatively related to its price. Thus, as prices increase, demand falls, and vice versa (a phenomenon that you may recognize as price elasticity—the ratio of the percentage change in demand to the percentage change in price). Different mathematical models can describe this phenomenon. In the following examples, we illustrate two of them.

EXAMPLE 1.7 A Linear Demand Prediction Model

A simple model to predict demand as a function of price is the linear model

$$D = a - bP \quad \textbf{(1.7)}$$

where D is the demand, P is the unit price, a is a constant that estimates the demand when the price is zero, and b is the slope of the demand function. This model is most applicable when we want to predict the effect of small changes around the current price. For example, suppose we know that when the price is \$100, demand is 19,000 units and that demand falls by 10 for each dollar of price increase. Using simple algebra, we can determine that $a = 20{,}000$ and $b = 10$. Thus, if the price is \$80, the predicted demand is

$$D = 20{,}000 - 10(80) = 19{,}200 \text{ units}$$

If the price increases to \$90, the model predicts demand as

$$D = 20{,}000 - 10(90) = 19{,}100 \text{ units}$$

If the price is \$100, demand would be

$$D = 20{,}000 - 10(100) = 19{,}000 \text{ units}$$

and so on. A graph of demand as a function of price is shown in Figure 1.5 as price varies between \$80 and \$120. We see that there is a constant decrease in demand for each \$10 increase in price, a characteristic of a linear model.

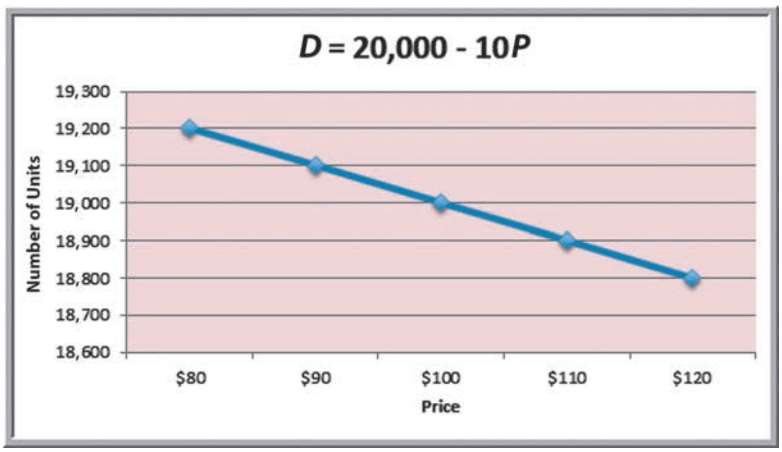


Figure 1.5

Graph of Linear Demand Model $D = a - bP$

EXAMPLE 1.8 A Nonlinear Demand Prediction Model

An alternative model assumes that price elasticity is constant. In this case, the appropriate model is

$$D = cP^{-d} \tag{1.8}$$

where c is the demand when the price is 0 and $d > 0$ is the price elasticity. To be consistent with Example 1.7, we assume that when the price is zero, demand is 20,000. Therefore, $c = 20{,}000$. We will also, as in Example 1.7, assume that when the price is \$100, $D = 19{,}000$.

Using these values in equation (1.8), we can determine the value for d as 0.0111382 (we can do this mathematically using logarithms, but we'll see how to do this very easily using Excel in Chapter 11). Thus, if the price is \$80, then the predicted demand is

$$D = 20{,}000(80)^{-0.0111382} = 19{,}047$$

If the price is 90, the demand would be

$$D = 20{,}000(90)^{-0.0111382} = 19{,}022$$

If the price is 100, demand is

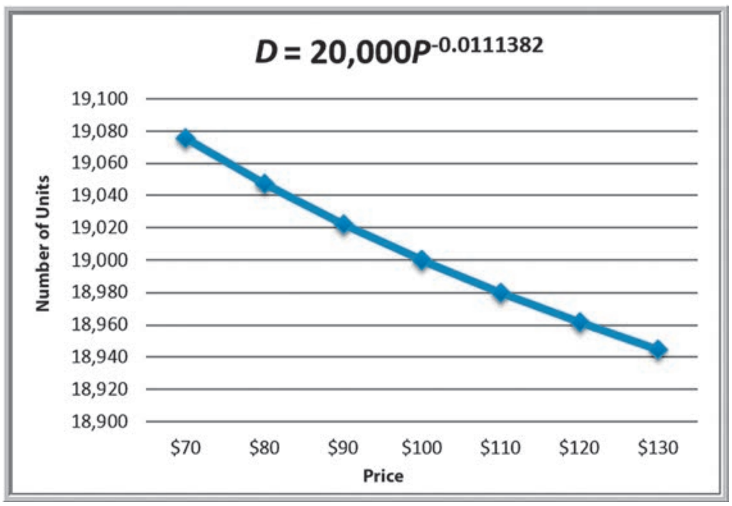
$$D = 20{,}000(100)^{-0.0111382} = 19{,}000$$

A graph of demand as a function of price is shown in Figure 1.6. The predicted demand falls in a slight nonlinear fashion as price increases. For example, demand decreases by 25 units when the price increases from \$80 to \$90, but only by 22 units when the price increases from \$90 to \$100. If the price increases to \$110, you would see a smaller decrease in demand. Therefore, we see a nonlinear relationship, in contrast to Example 1.7.

Both models in Examples 1.7 and 1.8 make different predictions of demand for different prices (other than \$90). Which model is best? The answer may be neither. First of all, the development of realistic models requires many price point changes within a carefully designed experiment. Second, it should also include data on competition and customer disposable income, both of which are hard to determine. Nevertheless, it is possible to develop price elasticity models with limited price ranges and narrow customer segments. A good starting point would be to create a historical database with detailed information on all past pricing actions. Unfortunately, practitioners have observed that such models are not widely used in retail marketing, suggesting ample opportunity to apply business analytics.[28]

Figure 1.6

Graph of Nonlinear Demand Model $D = cP^{-d}$



[28]Zhang, Clay Duan, and Arun Muthupalaniappan, "Analytics Applications in Consumer Credit and Retail Marketing," *Analytics Magazine* (November - December 2011): 27–33.

Uncertainty and Risk

As we all know, the future is always uncertain. Thus, many predictive models incorporate uncertainty and help decision makers analyze the risks associated with their decisions. **Uncertainty** is imperfect knowledge of what will happen; **risk** is associated with the consequences and likelihood of what might happen. For example, the change in the stock price of Apple on the next day of trading is uncertain. If you own Apple stock, then you face the risk of losing money if the stock price falls. If you don't own any stock, the price is still uncertain, although you would not have any risk. Risk is evaluated by the magnitude of the consequences and the likelihood that they would occur. For example, a 10% drop in a stock price would incur a higher risk if you own $1 million worth of that stock than if you only owned $1,000 worth of that stock. Similarly, if the chances of a 10% drop were 1 in 5, the risk would be higher than if the chances were only 1 in 100.

The importance of risk in business has long been recognized. The renowned management writer Peter Drucker observed in 1974:

> *To try to eliminate risk in business enterprise is futile. Risk is inherent in the commitment of present resources to future expectations. Indeed, economic progress can be defined as the ability to take greater risks. The attempt to eliminate risks, even the attempt to minimize them, can only make them irrational and unbearable. It can only result in the greatest risk of all: rigidity.*[29]

Consideration of risk is a vital element of decision making. For instance, you would probably not choose an investment simply on the basis of the return you might expect because, typically, higher returns are associated with higher risk. Therefore, you have to make a tradeoff between the benefits of greater rewards and the risks of potential losses. Analytic models can help assess risk. A model in which some of the model input information is uncertain is often called a **stochastic**, or **probabilistic**, **model**. In contrast, a **deterministic model** is one in which all model input information is either known or assumed to be known with certainty. For instance, suppose that customer demand is an important element of some model. We can make the assumption that the demand is known with certainty; say, 5,000 units per month. In this case, we would be dealing with a deterministic model. On the other hand, suppose we have evidence to indicate that demand is uncertain, with an average value of 5,000 units per month, but which typically varies between 3,200 and 6,800 units. If we make this assumption, we would be dealing with a stochastic model. Stochastic models are useful in analyzing uncertainty in real-world situations, and we will discuss these later in this book.

CHECK YOUR UNDERSTANDING

1. Define a model and state three common forms of a model.
2. Explain the elements of a decision model.
3. Explain how decision models are used for descriptive, predictive, and prescriptive applications.
4. Define optimization and the characteristics of optimization models.
5. Explain the importance of assumptions in building decision models.
6. What is the difference between uncertainty and risk?

---

[29] P. F. Drucker, *The Manager and the Management Sciences in Management: Tasks, Responsibilities, Practices* (London: Harper and Row, 1974).

Problem Solving with Analytics

The fundamental purpose of analytics is to help managers solve problems and make decisions. The techniques of analytics represent only a portion of the overall problem-solving and decision-making process. Problem solving is the activity associated with defining, analyzing, and solving a problem and selecting an appropriate solution that solves a problem.

Problem solving consists of several phases:

1. Recognizing a problem
2. Defining the problem
3. Structuring the problem
4. Analyzing the problem
5. Interpreting results and making a decision
6. Implementing the solution

Recognizing a Problem

Managers at different organizational levels face different types of problems. In a manufacturing firm, for instance, top managers face decisions regarding allocating financial resources, building or expanding facilities, determining product mix, and strategically sourcing production. Middle managers in operations develop distribution plans, production and inventory schedules, and staffing plans. Finance managers analyze risks, determine investment strategies, and make pricing decisions. Marketing managers develop advertising plans and make sales force allocation decisions. In manufacturing operations, problems involve the size of daily production runs, individual machine schedules, and worker assignments. Whatever the problem, the first step is to realize that it exists.

How are problems recognized? Problems exist when there is a gap between what is happening and what we think should be happening. For example, a consumer products manager might feel that distribution costs are too high. This recognition might result from comparing performance with that of a competitor, or observing an increasing trend compared to previous years.

Defining the Problem

The second step in the problem-solving process is to clearly define the problem. Finding the real problem and distinguishing it from symptoms that are observed is a critical step. For example, high distribution costs might stem from inefficiencies in routing trucks, poor location of distribution centers, or external factors such as increasing fuel costs. The problem might be defined as improving the routing process, redesigning the entire distribution system, or optimally hedging fuel purchases.

Defining problems is not a trivial task. The complexity of a problem increases when the following occur:

- The number of potential courses of action is large.
- The problem belongs to a group rather than to an individual.
- The problem solver has several competing objectives.
- External groups or individuals are affected by the problem.
- The problem solver and the true owner of the problem—the person who experiences the problem and is responsible for getting it solved—are not the same.
- Time limitations are important.

These factors make it difficult to develop meaningful objectives and characterize the range of potential decisions. In defining problems, it is important to involve all people who make the decisions or who may be affected by them.

Structuring the Problem

This usually involves stating goals and objectives, characterizing the possible decisions, and identifying any constraints or restrictions. For example, if the problem is to redesign a distribution system, decisions might involve new locations for manufacturing plants and warehouses (where?), new assignments of products to plants (which ones?), and the amount of each product to ship from different warehouses to customers (how much?). The goal of cost reduction might be measured by the total delivered cost of the product. The manager would probably want to ensure that a specified level of customer service—for instance, being able to deliver orders within 48 hours—is achieved with the redesign. This is an example of a constraint. Structuring a problem often involves developing a formal model.

Analyzing the Problem

Here is where analytics plays a major role. Analysis involves some sort of experimentation or solution process, such as evaluating different scenarios, analyzing risks associated with various decision alternatives, finding a solution that meets certain goals, or determining an optimal solution. Analytics professionals have spent decades developing and refining a variety of approaches to address different types of problems. Much of this book is devoted to helping you understand these techniques and gain a basic facility in using them.

Interpreting Results and Making a Decision

Interpreting the results from the analysis phase is crucial in making good decisions. Models cannot capture every detail of the real problem, and managers must understand the limitations of models and their underlying assumptions and often incorporate judgment into making a decision. For example, in locating a facility, we might use an analytical procedure to find a "central" location; however, many other considerations must be included in the decision, such as highway access, labor supply, and facility cost. Thus, the location specified by an analytical solution might not be the exact location the company actually chooses.

Implementing the Solution

This simply means making the solution work in the organization, or translating the results of a model back to the real world. This generally requires providing adequate resources, motivating employees, eliminating resistance to change, modifying organizational policies, and developing trust. Problems and their solutions affect people: customers, suppliers, and employees. All must be an important part of the problem-solving process. Sensitivity to political and organizational issues is an important skill that managers and analytical professionals alike must possess when solving problems.

In each of these steps, good communication is vital. Analytics professionals need to be able to communicate with managers and clients to understand the business context of the problem and be able to explain results clearly and effectively. Such skills as constructing good visual charts and spreadsheets that are easy to understand are vital to users of analytics. We emphasize these skills throughout this book.

ANALYTICS IN PRACTICE: Developing Effective Analytical Tools at Hewlett-Packard[30]

Hewlett-Packard (HP) uses analytics extensively. Many applications are used by managers with little knowledge of analytics. These require that analytical tools be easily understood. Based on years of experience, HP analysts compiled some key lessons. Before creating an analytical decision tool, HP asks three questions:

1. *Will analytics solve the problem?* Will the tool enable a better solution? Should other, nonanalytical solutions be used? Are there organizational or other issues that must be resolved? Often, what may appear to be an analytical problem may actually be rooted in problems of incentive misalignment, unclear ownership and accountability, or business strategy.
2. *Can we leverage an existing solution?* Before "reinventing the wheel," can existing solutions address the problem? What are the costs and benefits?
3. *Is a decision model really needed?* Can simple decision guidelines be used instead of a formal decision tool?

hans12/Fotolia

Once a decision is made to develop an analytical tool, several guidelines are used to increase the chances of successful implementation:

- *Use prototyping*, a quick working version of the tool designed to test its features and gather feedback.
- *Build insight, not black boxes.* A "black box" tool is one that generates an answer, but may not provide confidence to the user. Interactive tools that create insights to support a decision provide better information.
- *Remove unneeded complexity.* Simpler is better. A good tool can be used without expert support.
- *Partner with end users in discovery and design.* Decision makers who will actually use the tool should be involved in its development.
- *Develop an analytic champion.* Someone (ideally, the actual decision maker) who is knowledgeable about the solution and close to it must champion the process.

CHECK YOUR UNDERSTANDING

1. List the major phases of problem solving and explain each.
2. What lessons did Hewlett-Packard learn about using analytics?

[30]Based on Thomas Olavson and Chris Fry, "Spreadsheet Decision-Support Tools: Lessons Learned at Hewlett-Packard," *Interfaces*, 38, 4, July–August 2008: 300–310.

KEY TERMS

Algorithm
Big data
Business analytics (analytics)
Business intelligence (BI)
Constraint
Data mining
Decision model
Decision options (decision variables)
Decision support systems (DSS)
Descriptive analytics
Deterministic model
Information systems (IS)
Model
Modeling and optimization
Objective function
Operations Research/Management Science (OR/MS)
Optimal solution
Optimization
Predictive analytics
Prescriptive analytics
Price elasticity
Problem solving
Reliability
Risk
Search algorithm
Simulation and risk analysis
Statistics
Stochastic (probabilistic) model
Tag cloud
Uncertainty
Validity
Visualization
What-if analysis

CHAPTER 1 TECHNOLOGY HELP

Useful Excel Functions (*see Appendix A1*)

MIN(*range*) Finds the smallest value in a range of cells.

MAX(*range*) Finds the largest value in a range of cells.

SUM(*range*) Finds the sum of values in a range of cells.

AVERAGE(*range*) Finds the average of the values in a range of cells.

COUNT(*range*) Finds the number of cells in a range that contain numbers.

COUNTIF(*range, criteria*) Finds the number of cells within a range that meet a specified criterion.

NPV(*rate, value1, value2, . . .*) Calculates the net present value of an investment by using a discount rate and a series of future payments (negative values) and income (positive values).

DATEDIF(*startdate*, *enddate*, *time unit*) Computes the number of whole years, months, or days between two dates.

PROBLEMS AND EXERCISES

What Is Business Analytics?

1. Discuss the use of business analytics in sports, such as tennis, cricket, swimming, and football. Identify as many opportunities as you can for each.
2. How might analytics be used in the following situations?
   a. Planning a business
   b. Understanding customer behavior
   c. Solving human resource problems
3. Swissotel, London, is working on improving customer satisfaction and engagement. At the same time, Swissotel wants to ensure that the staff members are also satisfied. It has noticed that there are several areas that will need to be analyzed. The problem facing the hotel management is identifying what information is necessary to analyze a situation and improve customer and staff satisfaction levels. What data would Swissotel need to collect from its guests to facilitate good decisions? How might this data and business analytics help the hotel management?

Descriptive, Predictive, and Prescriptive Analytics

4. For each of the following scenarios, state whether descriptive, predictive, or prescriptive analytics tools would most likely be used.

- **a.** An insurance firm wants to analyze fraudulent claims to predict their future losses.
- **b.** A store manager has to review a store's historical operations, sales, financials, and customer reports to improve the store's business.
- **c.** The Met Office would like to forecast weather for nine to 10 days more accurately than the 24-hour forecasts, compared to the forecasting done 40 years ago.
- **d.** Autonomous or self-driving car programmers have to make millions of calculations for each trip to determine directions and speed.
- **e.** A secretary needs to summarize past events such as regional sales, customer attrition, and success of marketing campaigns for her manager to understand their company's progression.
- **f.** A fuel producer wants to identify the factors affecting the price of oil and gas to get the best terms and hedge risks.
- **g.** A mechanical engineer needs to predict the failure rate of the equipment in the next 6 months to help reduce maintenance costs and improve power availability.
- **h.** A travel agency is setting the price of their travel packages to attract more customers and wants to know the trends in popular travel destinations.

Models in Business Analytics

5. A firm installs 1,500 air conditioners that need to be checked every six months. The firm can hire a team from its logistics department at a fixed cost of €6,000. This team will check each unit for €15.00. The firm can also outsource this at a cost of €17.00 inclusive of all charges.

- **a.** For the given number of units, compute the firm's total cost of checking for both options. Which is a better decision?
- **b.** Find the break-even volume and characterize the range of volumes for which it is more economical to outsource.

6. Use the model developed in Example 1.5 to predict the total sales for weeks 2 through 16, and compare the results to the observed sales. Does the accuracy of the model seem to be different when coupons are used or not? When advertising is used or not?

7. The yearly government tax revenue (in \$ millions) depends on customer income, business profits, and capital gains defined as tax $= -2.556 + 3.799 \times$ income $+ 1.583 \times$ profit $+ 0.739 \times$ capital.

- **a.** Interpret the numbers in this model.
- **b.** What is the government tax revenue if a customer earns \$1.2 million, the business' profits amount to \$8.63 million, and the capital gains are \$5.76 million?

8. In a toy manufacturing company, the manufacturer needs to pay rent (p) each month and monthly electricity (q). To produce each toy, the company incurs costs for plastic (r) and for cloth (s). The variables p, q, r, and s are positive constants. Given the cost model as

$$\text{cost} = \text{fixed cost} + \text{variable cost}$$

- **a.** Define the cost model with the given unknowns.
- **b.** How does plastic and cloth affect the cost model?
- **c.** Will the variable plastic influence the cloth variable?

9. An insurance company is estimating the risk model based on customer age is define as

$$R = ae^{bx}$$

where a is a fixed constant and b is a constant that depends on the customer age. They categorize the age into 5 categories as age < 10 with $b = 0.05$, $10 <$ age ≤ 20 with $b = 0.02$, $20 <$ age ≤ 30 with $b = 0.01$, $30 <$ age ≤ 40 with $b = 0.03$ and age > 40 with $b = 0.08$. Sketch the graphs for each model and interpret them. Are the graphs reasonable?

10. A tablet manufacturer is preparing to set the price of a new model. It examines the demand and cost structure, and determines the following model to represent the relationship between demand and price

$$D = 50 + 2P$$

The finance department estimates that the total costs is represented by

$$C = 100 + 0.3D$$

Develop a model for the total profit in terms of the price, P.

Problem Solving with Analytics

11. In this chapter, we noted the importance of defining and analyzing a problem prior to attempting to find a solution. Consider this example: One of the earliest operations research groups during World War II was conducting a study on the optimum utilization of Spitfire and Hurricane aircraft during the Battle of Britain. Whenever one of these planes returned from battle, the locations of the bullet holes on it were carefully plotted. By repeatedly recording these data over time, and studying the clusters of data, the group was able to estimate the regions of the aircraft most likely to be hit by enemy gunfire, with the objective of reinforcing these regions with special armor. What difficulties do you see here?

12. A part-time professor of a business program at a university in Germany always tells her students that one of the first questions they should ask their new employers is whether the organization has a good customer program. Why do you think the professor encourages her students to ask this specific question?

CASE: PERFORMANCE LAWN EQUIPMENT

In each chapter of this book, we use a a fictitious company, Performance Lawn Equipment (PLE), within a case exercise for applying the tools and techniques introduced in the chapter.[31] To put the case in perspective, we first provide some background about the company, so that the applications of business analytic tools will be more meaningful.

PLE, headquartered in St. Louis, Missouri, is a privately owned designer and producer of traditional lawn mowers used by homeowners. In the past ten years, PLE has added another key product, a medium-size diesel power lawn tractor with front and rear power takeoffs, Class I three-point hitches, four-wheel drive, power steering, and full hydraulics. This equipment is built primarily for a niche market consisting of large estates, including golf and country clubs, resorts, private estates, city parks, large commercial complexes, lawn care service providers, private homeowners with five or more acres, and government (federal, state, and local) parks, building complexes, and military bases. PLE provides most of the products to dealerships, which, in turn, sell directly to end users. PLE employs 1,660 people worldwide. About half the workforce is based in St. Louis; the remainder is split among their manufacturing plants.

In the United States, the focus of sales is on the eastern seaboard, California, the Southeast, and the south central states, which have the greatest concentration of customers. Outside the United States, PLE's sales include a European market, a growing South American market, and developing markets in the Pacific Rim and China.

Both end users and dealers have been established as important customers for PLE. Collection and analysis of end-user data showed that satisfaction with the products depends on high quality, easy attachment/dismount of implements, low maintenance, price value, and service. For dealers, key requirements are high quality, parts and feature availability, rapid restock, discounts, and timeliness of support.

PLE has several key suppliers: Mitsitsiu, Inc., the sole source of all diesel engines; LANTO Axles, Inc., which provides tractor axles; Schorst Fabrication, which provides subassemblies; Cuberillo, Inc, supplier of transmissions; and Specialty Machining, Inc., a supplier of precision machine parts.

To help manage the company, PLE managers have developed a "balanced scorecard" of measures. These data, which are summarized shortly, are stored in the form of a Microsoft Excel workbook (*Performance Lawn Equipment*) accompanying this book. The database contains various measures captured on a monthly or quarterly basis and is used by various managers to evaluate business performance. Data for each of the key measures are stored in a separate worksheet. A summary of these worksheets is given next:

- *Dealer Satisfaction*, measured on a scale of 1–5 (1 = poor, 2 = less than average, 3 = average, 4 = above average, and 5 = excellent). Each year, dealers in each region are surveyed about their overall satisfaction with PLE. The worksheet contains summary data from surveys for the past five years.
- *End-User Satisfaction*, measured on the same scale as dealers. Each year, 100 users from each region are surveyed. The worksheet contains summary data for the past five years.
- *Customer Survey*, results from a survey for customer ratings of specific attributes of PLE tractors: quality, ease of use, price, and service on the same 1–5 scale. This sheet contains 200 observations of customer ratings.

[31]The case scenario was based on *Gateway Estate Lawn Equipment Co. Case Study*, used for the 1997 Malcolm Baldrige National Quality Award Examiner Training course. This material is in the public domain. The database, however, was developed by the author.

- *Complaints*, which shows the number of complaints registered by all customers each month in each of PLE's five regions (North America, South America, Europe, the Pacific, and China).
- *Mower Unit Sales* and *Tractor Unit Sales*, which provide sales by product by region on a monthly basis. Unit sales for each region are aggregated to obtain world sales figures.
- *Industry Mower Total Sales* and *Industry Tractor Total Sales*, which list the number of units sold by all producers by region.
- *Unit Production Costs*, which provides monthly accounting estimates of the variable cost per unit for manufacturing tractors and mowers over the past five years.
- *Operating and Interest Expenses*, which provides monthly administrative, depreciation, and interest expenses at the corporate level.
- *On-Time Delivery*, which provides the number of deliveries made each month from each of PLE's major suppliers, the number on time, and the percent on time.
- *Defects After Delivery*, which shows the number of defects in supplier-provided material found in all shipments received from suppliers.
- *Time to Pay Suppliers*, which provides measurements in days from the time the invoice is received until payment is sent.
- *Response Time*, which gives samples of the times taken by PLE customer-service personnel to respond to service calls by quarter over the past two years.
- *Employee Satisfaction*, which provides data for the past four years of internal surveys of employees to determine their overall satisfaction with their jobs, using the same scale used for customers. Employees are surveyed quarterly, and results are stratified by employee category: design and production, managerial, and sales/administrative support.

In addition to these business measures, the PLE database contains worksheets with data from special studies:

- *Engines*, which lists 50 samples of the time required to produce a lawn mower blade using a new technology.
- *Transmission Costs*, which provides the results of 30 samples each for the current process used to produce tractor transmissions and two proposed new processes.
- *Blade Weight*, which provides samples of mower blade weights to evaluate the consistency of the production process.
- *Mower Test*, which lists test results of mower functional performance after assembly for 30 samples of 100 units each.
- *Employee Retention*, data from a study of employee duration (length of hire) with PLE. The 40 subjects were identified by reviewing hires from ten years prior and identifying those who were involved in managerial positions (either hired into management or promoted into management) at some time in this ten-year period.
- *Shipping Cost*, which gives the unit shipping cost for mowers and tractors from existing and proposed plants for a supply chain design study.
- *Fixed Cost*, which lists the fixed cost to expand existing plants or build new facilities, also as part of the supply chain design study.
- *Purchasing Survey*, which provides data obtained from a third-party survey of purchasing managers of customers of Performance Lawn Care.

Elizabeth Burke has recently joined the PLE management team to oversee production operations. She has reviewed the types of data that the company collects and has assigned you the responsibility to be her chief analyst in the coming weeks. She has asked you to do some preliminary analysis of the data for the company.

1. First, she would like you to edit the worksheets *Dealer Satisfaction* and *End-User Satisfaction* to display the total number of responses to each level of the survey scale across all regions for each year.
2. Second, she wants a count of the number of failures in the worksheet *Mower Test*.
3. Next, Elizabeth has provided you with prices for PLE products for the past five years:

Year	Mower Price	Tractor Price
2014	$150	$3,250
2015	$175	$3,400
2016	$180	$3,600
2017	$185	$3,700
2018	$190	$3,800

   Create a new worksheet to compute gross revenues by month and region, as well as worldwide totals, for each product using the data in *Mower Unit Sales* and *Tractor Unit Sales*.
4. Finally, she wants to know the market share for each product and region by month based on the PLE and industry sales data, and the average market share by region over the five years.

Summarize all your findings in a report to Ms. Burke.

APPENDIX

A1

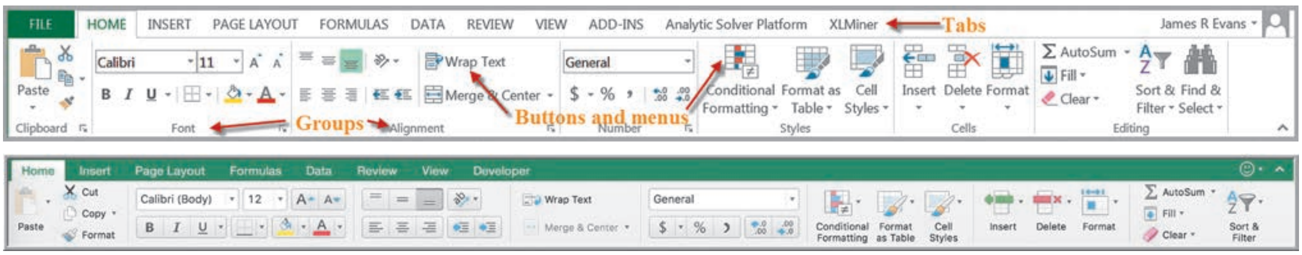
Basic Excel Skills

To be able to apply the procedures and techniques that you will learn in this book, it is necessary for you to be relatively proficient in using Excel. We assume that you are familiar with the most elementary spreadsheet concepts and procedures, such as

- opening, saving, and printing files;
- using workbooks and worksheets;
- moving around a spreadsheet;
- selecting cells and ranges;
- inserting/deleting rows and columns;
- entering and editing text, numerical data, and formulas in cells;
- formatting data (number, currency, decimal places, etc.);
- working with text strings;
- formatting data and text; and
- modifying the appearance of the spreadsheet using borders, shading, and so on.

Menus and commands in Excel reside in the “ribbon” shown in Figure A1.1. All Excel discussions in this book will be based on Excel 2016 for Windows; if you use Excel 2016 for Mac, some differences may exist, and we may point these out as appropriate. Menus and commands are arranged in logical *groups* under different *tabs* (*Home, Insert, Formulas,* and so on); small triangles pointing downward indicate *menus* of additional choices. We often refer to certain commands or options and where they may be found in the ribbon. For instance, in the Mac version, groups are not specified.

Excel provides an add-in called the *Analysis Toolpak*, which contains a variety of tools for statistical computation, and *Solver*, which is used for optimization. They will be found in the *Data* tab ribbon; you should ensure that these are activated. To activate them in Windows, click the *File* tab and then *Options* in the left column. Choose *Add-Ins* from the left column. At the bottom of the dialog, make sure *Excel Add-ins* is selected in the *Manage:* box and click *Go*. In the *Add-Ins* dialog, if *Analysis Toolpak*, *Analysis Toolpak VBA*, and *Solver Add-in* are not checked, simply check the boxes and click *OK*. You will not have to repeat this procedure every time you run Excel in the future. On Excel 2016 for Mac, go to *Tools* > *Excel Add-ins* and select both *Analysis Toolpak* and *Solver*.



▲ **Figure A1.1**

Excel Ribbons for Windows and Mac

Excel Formulas and Addressing

Formulas in Excel use common mathematical operators:

- addition (+)
- subtraction (−)
- multiplication (*)
- division (/)

Exponentiation uses the ^ symbol; for example, 2^5 is written as 2^5 in an Excel formula.

Cell references in formulas can be written either with *relative addresses* or *absolute addresses*. A **relative address** uses just the row and column label in the cell reference (for example, A4 or C21); an **absolute address** uses a dollar sign ($ sign) before either the row or column label or both (for example, $A2, C$21, or $B$15). Which one we choose makes a critical difference if you copy the cell formulas. If only relative addressing is used, then copying a formula to another cell changes the cell references by the number of rows or columns in the direction that the formula is copied. So, for instance, if we would use a formula in cell B8, =B4−B5*A8, and copy it to cell C9 (one column to the right and one row down), all the cell references are increased by one and the formula would be changed to =C5−C6*B9.

Using a $ sign before a row label (for example, B$4) keeps the reference fixed to row 4 but allows the column reference to change if the formula is copied to another cell. Similarly, using a $ sign before a column label (for example, $B4) keeps the reference to column B fixed but allows the row reference to change. Finally, using a $ sign before both the row and column labels (for example, $B$4) keeps the reference to cell B4 fixed no

EXAMPLE A1.1 Implementing Price-Demand Models in Excel

In Chapter 1, we described two models for predicting demand as a function of price:

$$D = a - bP$$

and

$$D = cP^{-d}$$

Figure A1.2 shows a spreadsheet (Excel file *Demand Prediction Models*) for calculating demand for different prices using each of these models. For example, to calculate the demand in cell B8 for the linear model, we use the formula

=$B$4 – $B$5*A8

To calculate the demand in cell E8 for the nonlinear model, we use the formula

=$E$4*D8^ – $E$5

Note how the absolute addresses are used so that as these formulas are copied down, the demand is computed correctly.

▸ **Figure A1.2**

Excel Models for Demand Prediction

	A	B	C	D	E
1	Demand Prediction Models				
2					
3	Linear Model			Nonlinear Model	
4	a	20,000		c	20,000
5	b	10		d	0.0111382
6					
7	Price	Demand		Price	Demand
8	$80.00	$19,200		$70.00	$19,075.63
9	$90.00	$19,100		$80.00	$19,047.28
10	$100.00	$19,000		$90.00	$19,022.31
11	$110.00	$18,900		$100.00	$19,000.00
12	$120.00	$18,800		$110.00	$18,979.84
13				$120.00	$18,961.45
14				$130.00	$18,944.56

matter where the formula is copied. You should be very careful to use relative and absolute addressing appropriately in your models, especially when copying formulas.

Copying Formulas

Excel provides several ways of copying formulas to different cells. This is extremely useful in building decision models, because many models require replication of formulas for different periods of time, similar products, and so on. The easiest way is to select the cell with the formula to be copied, and then press Ctrl-C on your Windows keyboard or Command-C on a Mac, click on the cell you wish to copy to, and then press Ctrl-V in Windows or Command-V on a Mac. You may also enter a formula directly in a range of cells without copying and pasting by selecting the range, typing in the formula, and pressing Ctrl-Enter in Windows or Command-Enter on a Mac.

To copy a formula from a single cell or range of cells down a column or across a row, first select the cell or range, click and hold the mouse on the small square in the lower right-hand corner of the cell (the "fill handle"), and drag the formula to the "target" cells to which you wish to copy.

Useful Excel Tips

- ***Split Screen.*** You may split the worksheet horizontally and/or vertically to view different parts of the worksheet at the same time. The vertical splitter bar is just to the right of the bottom scroll bar, and the horizontal splitter bar is just above the right-hand scroll bar. Position your cursor over one of these until it changes shape, click, and drag the splitter bar to the left or down.
- ***Column and Row Widths.*** Many times a cell contains a number that is too large to display properly because the column width is too small. You may change the column width to fit the largest value or text string anywhere in the column by positioning the cursor to the right of the column label so that it changes to a cross with horizontal arrows and then double-clicking. You may also move the arrow to the left or right to manually change the column width. You may change the row heights in a similar fashion by moving the cursor below the row number label. This can be especially useful if you have a very long formula to display. To break a formula within a cell, position the cursor at the break point in the formula bar and press Alt-Enter.
- ***Displaying Formulas in Worksheets.*** Choose *Show Formulas* in the *Formulas* tab. You may also press Ctrl ~ in either Windows or Mac to toggle formulas on and off. You often need to change the column width to display the formulas properly.
- ***Displaying Grid Lines and Row and Column Headers for Printing.*** Check the *Print* boxes for gridlines and headings in the *Sheet Options* group under the *Page Layout* tab. Note that the *Print* command can be found by clicking on the *Office* button in Windows or under the *File* menu in Mac.
- ***Filling a Range with a Series of Numbers.*** Suppose you want to build a worksheet for entering 100 data values. It would be tedious to have to enter the numbers from 1 to 100 one at a time. Simply fill in the first few values in the series and highlight them. Then click and drag the small square (fill handle) in the lower right-hand corner down (Excel will show a small pop-up window that tells you the last value in the range) until you have filled in the column to 100; then release the mouse.

Excel Functions

Functions are used to perform special calculations in cells and are used extensively in business analytics applications. All Excel functions require an equal sign and a function name followed by parentheses, in which you specify arguments for the function.

Basic Excel Functions

Some of the more common functions that we will use in applications include the following:

MIN(*range*)—finds the smallest value in a range of cells
MAX(*range*)—finds the largest value in a range of cells
SUM(*range*)—finds the sum of values in a range of cells
AVERAGE(*range*)—finds the average of the values in a range of cells
COUNT(*range*)—finds the number of cells in a range that contain numbers
COUNTIF(*range, criteria*)—finds the number of cells within a range that meet a specified criterion

Logical functions, such as IF, AND, OR, and VLOOKUP will be discussed in Chapter 2.

The COUNTIF function counts the number of cells within a range that meet a criterion you specify. For example, you can count all the cells that start with a certain letter, or you can count all the cells that contain a number that is larger or smaller than a number you specify. Examples of criteria are 100, ">100", a cell reference such as A4, and a text string such as "Facebook." Note that text and logical formulas must be enclosed in quotes. See Excel Help for other examples.

Excel has other useful COUNT-type functions: COUNTA counts the number of non-blank cells in a range, and COUNTBLANK counts the number of blank cells in a range. In addition, COUNTIFS(*range1, criterion1, range2, criterion2, . . . , range_n, criterion_n*) finds the number of cells within multiple ranges that meet specific criteria for each range.

We illustrate these functions using the *Purchase Orders* data set in Example A1.2.

EXAMPLE A1.2 Using Basic Excel Functions

In the *Purchase Orders* data set, we will find the following:

- Smallest and largest quantity of any item ordered
- Total order costs
- Average number of months per order for accounts payable
- Number of purchase orders placed
- Number of orders placed for O-rings
- Number of orders with A/P (accounts payable) terms shorter than 30 months
- Number of O-ring orders from Spacetime Technologies
- Total cost of all airframe fasteners
- Total cost of airframe fasteners purchased from Alum Sheeting

The results are shown in Figure A1.3. In this figure, we used the split-screen feature in Excel to reduce the number of rows shown in the spreadsheet. To find the smallest and largest quantity of any item ordered, we use the MIN and MAX functions for the data in column F. Thus, the formula in cell B99 is =MIN(F4:F97) and the formula in cell B100 is =MAX(F4:F97). To find the total order costs, we sum the data in column G using the SUM function: =SUM(G4:G97); this is the formula in cell B101. To find the average number of A/P months, we use the AVERAGE function for the data in column H. The formula in cell B102 is =AVERAGE(H4:H97). To find the number of purchase orders placed, use the COUNT function. Note that the COUNT function counts only the number of cells in a range that contain numbers, so we could not use it in columns A, B, or D; however, any other column would be acceptable. Using the item numbers in column C, the formula in cell B103 is =COUNT(C4:C97). To find the number of orders placed for O-rings, we use the COUNTIF function. For this example, the formula used in cell B104 is =COUNTIF(D4:D97, "O-Ring"). We could have also used the cell reference for any cell containing the text O-Ring, such as =COUNTIF(D4:D97, D12). To find the number of orders with A/P terms less than 30 months, we use the formula =COUNTIF(H4:H97, "<30") in cell B105. Finally, to count the number of O-ring orders for Spacetime Technologies, we use =COUNTIFS(D4:D97, "O-Ring", A4:A97, "Spacetime Technologies").

IF-type functions are also available for other calculations. For example, the functions SUMIF, AVERAGEIF, SUMIFS, and AVERAGEIFS can be used to embed IF logic within mathematical functions. For instance, the syntax of SUMIF is SUMIF(*range, criterion, [sum range]*); *sum range* is an optional argument that allows you to add cells in a different range. Thus, in the *Purchase Orders* database, to find the total cost of all airframe fasteners, we would use

=SUMIF(D4:D97, "Airframe fasteners", G4:G97)

This function looks for airframe fasteners in the range D4:D97, but then sums the associated values in column G (cost per order). The arguments for SUMIFS and AVERAGEIFS are *(sumrange, range1, criterion1, range2, criterion2, . . . ,rangeN, criterionN)*. For example, the function

=SUMIFS(F4:F97,A4:A97,"Alum Sheeting", D4:D97, "Airframe fasteners")

will find the total quantity (from the *sumrange* in column F) of all airframe fasteners purchased from Alum Sheeting.

Functions for Specific Applications

Excel has a wide variety of other functions for statistical, financial, and other applications, many of which we introduce and use throughout the text. For instance, some financial models that we develop require the calculation of net present value (NPV). **Net present value** (also called **discounted cash flow**) measures the worth of a stream of cash flows, taking into account the time value of money. That is, a cash flow of F dollars t time periods in the future is worth $F/(1+i)^t$ dollars today, where i is the **discount rate.** The discount rate reflects the opportunity costs of spending funds now versus achieving a return through another investment, as well as the risks associated with not receiving returns until a later

	A	B	C	D	E	F	G	H	I	J
1	**Purchase Orders**									
2										
3	**Supplier**	**Order No.**	**Item No.**	**Item Description**	**Item Cost**	**Quantity**	**Cost per order**	**A/P Terms (Months)**	**Order Date**	**Arrival Date**
4	Hulkey Fasteners	Aug11001	1122	Airframe fasteners	$ 4.25	19,500	$ 82,875.00	30	08/05/11	08/13/11
5	Alum Sheeting	Aug11002	1243	Airframe fasteners	$ 4.25	10,000	$ 42,500.00	30	08/08/11	08/14/11
6	Fast-Tie Aerospace	Aug11003	5462	Shielded Cable/ft.	$ 1.05	23,000	$ 24,150.00	30	08/10/11	08/15/11
7	Fast-Tie Aerospace	Aug11004	5462	Shielded Cable/ft.	$ 1.05	21,500	$ 22,575.00	30	08/15/11	08/22/11
8	Steelpin Inc.	Aug11005	5319	Shielded Cable/ft.	$ 1.10	17,500	$ 19,250.00	30	08/20/11	08/31/11
9	Fast-Tie Aerospace	Aug11006	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/20/11	08/26/11
10	Steelpin Inc.	Aug11007	4312	Bolt-nut package	$ 3.75	4,250	$ 15,937.50	30	08/25/11	09/01/11
11	Durrable Products	Aug11008	7258	Pressure Gauge	$ 90.00	100	$ 9,000.00	45	08/25/11	08/28/11
12	Fast-Tie Aerospace	Aug11009	6321	O-Ring	$ 2.45	1,300	$ 3,185.00	30	08/25/11	09/04/11
96	Steelpin Inc.	Nov11009	5677	Side Panel	$ 195.00	110	$ 21,450.00	30	11/05/11	11/17/11
97	Manley Valve	Nov11010	9955	Door Decal	$ 0.55	125	$ 68.75	30	11/05/11	11/10/11
98										
99	Minimum Quantity	90								
100	Maximum Quantity	25,000								
101	Total Order Costs	$ 2,471,760.00								
102	Average Number of A/P Months	30.63829787								
103	Number of Purchase Orders	94								
104	Number of O-ring Orders	12								
105	Number of A/P Terms < 30	17								
106	Number of O-ring Orders Spacetime	3								

▲ **Figure A1.3**

Application of Basic Excel Functions to Purchase Orders *Data* time. The sum of the present values of all cash flows over a stated time horizon is the net present value:

$$\text{NPV} = \sum_{t=0}^{n} \frac{F_t}{(1+i)^t} \tag{A1.1}$$

where F_t = cash flow in period t. A positive NPV means that the investment will provide added value because the projected return exceeds the discount rate.

The Excel function NPV(*rate, value1, value2, . . .*) calculates the net present value of an investment by using a discount rate and a series of future payments (negative values) and income (positive values). *Rate* is the value of the discount rate i over the length of one period, and *value1*, *value2*, . . . are 1 to 29 arguments representing the payments and income for each period. The values must be equally spaced in time and are assumed to occur at the end of each period. The NPV investment begins one period before the date of the *value1* cash flow and ends with the last cash flow in the list. The NPV calculation is based on future cash flows. If the first cash flow (such as an initial investment or fixed cost) occurs at the beginning of the first period, then it must be added to the NPV result and *not* included in the function arguments.

EXAMPLE A1.3 Using the NPV Function

A company is introducing a new product. The fixed cost for marketing and distribution is $25,000 and is incurred just prior to launch. The forecasted net sales revenues for the first six months are shown in Figure A1.4. The formula in cell B8 computes the net present value of these cash flows as =NPV(B6, C4:H4)−B5. Note that the fixed cost is not a future cash flow and is not included in the NPV function arguments.

Insert Function

The easiest way to locate a particular function is to select a cell and click on the *Insert Function* button $[f_x]$, which can be found under the ribbon next to the formula bar and also in the *Function Library* group in the *Formulas* tab. You may either type in a description in the search field, such as "net present value," or select a category, such as "financial," from the drop-down box.

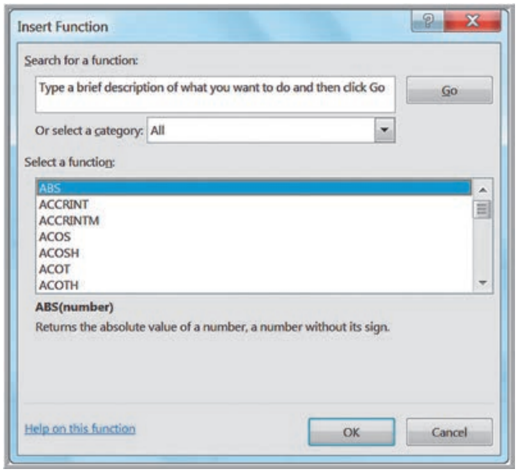
This feature is particularly useful if you know what function to use but are not sure of what arguments to enter because it will guide you in entering the appropriate data for the function arguments. Figure A1.5 shows the dialog from which you may select the function you wish to use. For example, if we would choose the COUNTIF function, the dialog in Figure A1.6 appears. When you click in an input cell, a description of the argument is shown. Thus, if you are not sure what to enter for the range, the explanation in Figure A1.6

▶ **Figure A1.4**

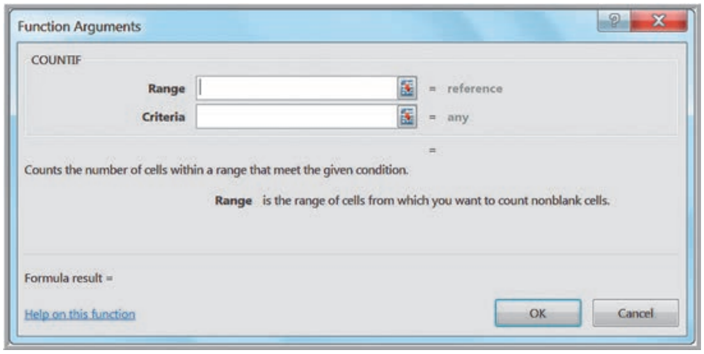
Net Present Value Calculation

	A	B	C	D	E	F	G	H
1	Net Present Value							
2								
3		Month	January	February	March	April	May	June
4		Sales Revenue Forecast	$2,500	$4,000	$5,000	$8,000	$10,000	$12,500
5	Fixed Cost	$25,000.00						
6	Discount Rate	3%						
7								
8	NPV	$11,975.81						

▶ **Figure A1.5**
Insert Function Dialog



▶ **Figure A1.6**
Function Arguments Dialog for COUNTIF



will help you. For further information, you could click on the *Help* button in the lower left-hand corner.

Date and Time Functions

In many analytics applications, a database might contain dates, such as when an order is placed or when an employee was hired. Excel can display a date in a variety of formats, such as 2/14/17 or 14-Feb-17. You may choose the standard date format (for example, 2/14/17) by selecting *Date* in the *Number* formatting box or select a custom format by selecting *Custom* in the *Number* box. Excel stores dates in a serial format. January 1, 1900 is day 1, and each subsequent day is numbered sequentially. Both the current date and January 1, 1900 are included in the count. So 2/14/17 is 42,780 when expressed in this format. This means there are 42,780 days between January 1, 1900 and February 14, 2017 (including both days). Therefore, to determine the number of days between two dates, you can simply subtract them.

Another useful date function is DATEDIF (which surprisingly doesn't appear in the *Insert Function* list!), which can compute the number of whole years, months, or days between two dates. The syntax of DATEDIF is

DATEDIF(*startdate, enddate, time unit*)

The time unit can be "y," "m," or "d." For instance, DATEDIF(4/26/89, 2/14/17, "y") will return 27 (years), while DATEDIF(4/26/89, 2/14/17, "m") will return 333 (months).

EXAMPLE A1.4 Computing Lead Times

In the *Purchase Orders* database, we will compute the lead time for each order, that is, the number of days between the order date and arrival date. Figure A1.7 shows the use of the DATEDIF function. Alternatively, we could have simply subtracted the values: for example, in cell K4, use = J4−I4.

Other useful date functions are

- YEAR(date)
- MONTH(date)
- DAY(date)

These functions simply extract the year, month, and day from a date or cell reference that contains a date. The function TODAY( ) displays the current date.

Similar to date functions, times can be formatted in a variety of ways, such as 12:26 PM, Hours:Minutes:Seconds, or in military time. The function NOW( ) displays the current time and date.

Miscellaneous Excel Functions and Tools

In this section, we will illustrate a few miscellaneous Excel functions and tools that support analytics applications.

Range Names

A range name is a descriptive label assigned to a cell or range of cells. Range names can help to facilitate building models on spreadsheets and understanding the formulas on which models are based. There are several ways to create range names in Excel.

EXAMPLE A1.5 Using the *Name* Box to Create a Range Name

Suppose that we create a simple spreadsheet for computing total cost (which is the fixed cost plus unit variable cost times quantity produced) shown in Figure A1.8. We will define range names for each of the numerical cells that correspond to the labels on the left. That is, we will name cell B3 Fixed cost, cell B4 Unit variable cost, and so on. Click on cell B3; in the *Name* box, type the name Fixed_cost (note the underscore), and then press Enter. Figure A1.8 shows that the name for cell B3 is displayed in the *Name* box. Repeat this process for each of the other numerical cells.

K4 =DATEDIF(I4,J4,"d")

	A	B	C	D	E	F	G	H	I	J	K
1	Purchase Orders										
2											
3	Supplier	Order No.	Item No.	Item Description	Item Cost	Quantity	Cost per order	A/P Terms (Months)	Order Date	Arrival Date	Lead Time
4	Hulkey Fasteners	Aug11001	1122	Airframe fasteners	$ 4.25	19,500	$ 82,875.00	30	08/05/11	08/13/11	8
5	Alum Sheeting	Aug11002	1243	Airframe fasteners	$ 4.25	10,000	$ 42,500.00	30	08/08/11	08/14/11	6
6	Fast-Tie Aerospace	Aug11003	5462	Shielded Cable/ft.	$ 1.05	23,000	$ 24,150.00	30	08/10/11	08/15/11	5
7	Fast-Tie Aerospace	Aug11004	5462	Shielded Cable/ft.	$ 1.05	21,500	$ 22,575.00	30	08/15/11	08/22/11	7
8	Steelpin Inc.	Aug11005	5319	Shielded Cable/ft.	$ 1.10	17,500	$ 19,250.00	30	08/20/11	08/31/11	11
9	Fast-Tie Aerospace	Aug11006	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/20/11	08/26/11	6

▲ **Figure A1.7**

Using DATEDIF to Compute Lead Times

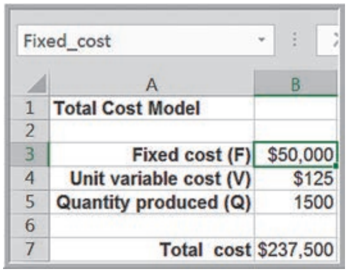


Figure A1.8

Defined Range Name for Cell B3 Using the Name Box

1. Use the *Name* box. The name box is at the upper left of a spreadsheet between the ribbon and the column headers. Usually, it displays the cell reference that is selected. To name a range, first select a cell or range and enter the range name in the *Name* box. Range names cannot contain any spaces, so it is common to use an underscore between words.

2. Use *Create from Selection*. This option is particularly useful when the names you want to use are listed in the column immediately to the right or left of the cell, or in the row immediately above or below the cell or range. The next example illustrates this procedure.

EXAMPLE A1.6 Using *Create from Selection* to Define Range Names

In the *Total Cost Model* spreadsheet, we will use the text labels to the left of the numerical inputs as the range names. First, highlight the range A3:B5. Then, on the *Formulas* tab, choose *Create from Selection*. The box for the left column will automatically be checked. Click OK. If you select any numerical cell, you will see the range name in the *Name* box as shown in Figure A1.9.

3. Use *Define Name*. This allows you to enter the name but also provides an option to restrict the name to the current worksheet or allow it to be used in any worksheet of the workbook.

EXAMPLE A1.7 Using *Define Name* to Create a Range Name

In the *Total Cost Model* spreadsheet, select cell B3. Click *Define Name* on the *Formulas* tab. This will bring up a dialog that allows you to enter a range name. Figure A1.10 illustrates this. Click OK.

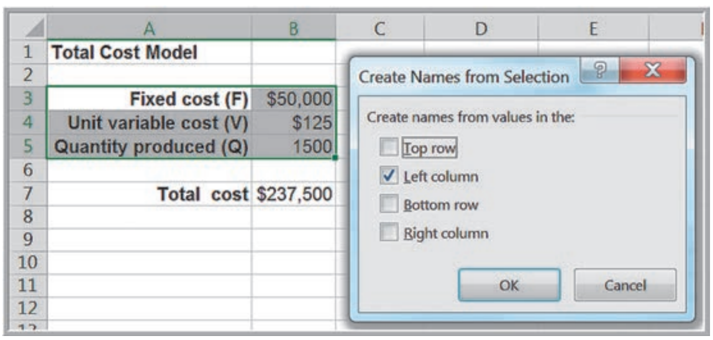
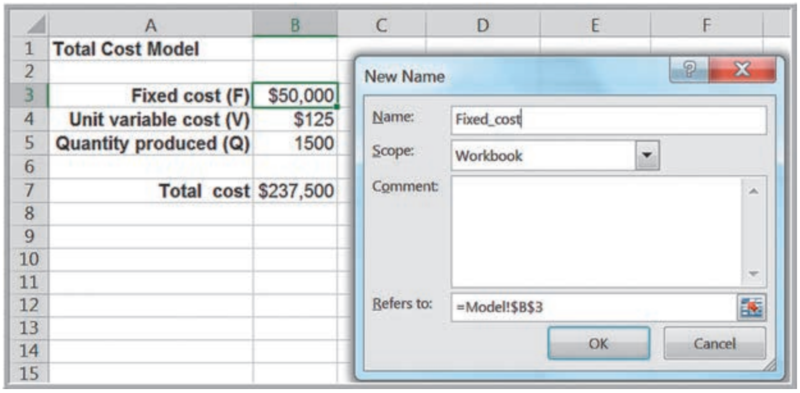


Figure A1.9

Defined Range Name for Cell B4 Using Create from Selection

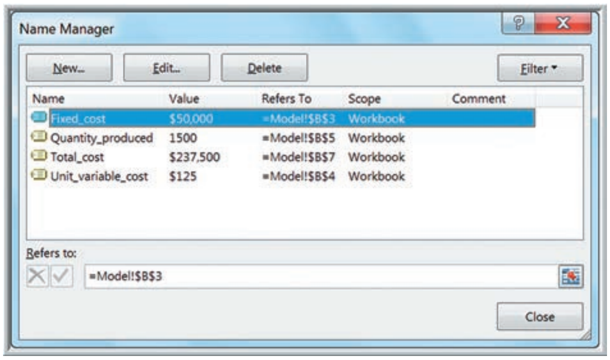


▶ **Figure A1.10**

Defined Range Name for Cell B3 Using Define Name

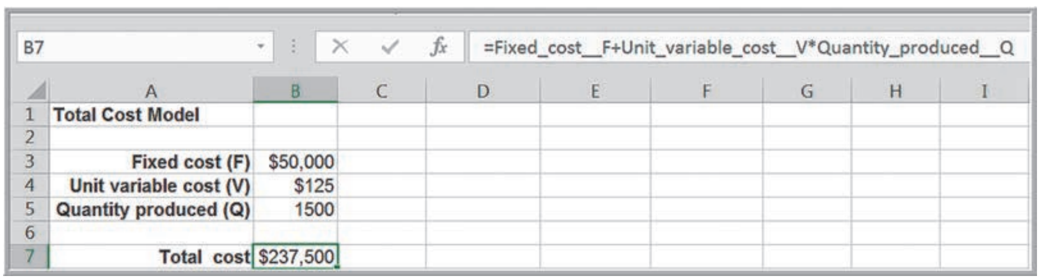
After range names have been defined, we can display a summary by clicking the *Name Manager* in the *Formulas* tab. Figure A1.11 shows this. (Note: The *Name Manager* button in the *Formulas* tab is only available in Windows. On a Mac, you can click *Define Name* and see a list of your range names, and then modify, add, or delete them.) This allows you to easily add, edit, or delete range names.

Finally, you can apply range names to the formulas in the spreadsheet. This replaces the cell references by the names, making it easy to understand the formulas. To do this, click on the drop-down arrow next to *Define Name* and select *Apply Names*. . . . In the dialog box, select all the names you wish to use and click OK. Figure A1.12 shows the result for the total cost cell, B7. We see that the original formula = B3 + B4*B5 now displays the names.



▶ **Figure A1.11**

Name Manager *Dialog (Windows)*



▲ **Figure A1.12**

Using Range Names in Formulas