

Business Analytics

Third Edition



Jeffrey D. Camm
Wake Forest University

James J. Cochran
University of Alabama

Michael J. Fry
University of Cincinnati

Jeffrey W. Ohlmann
University of Iowa

David R. Anderson
University of Cincinnati

Dennis J. Sweeney
University of Cincinnati

Thomas A. Williams
Rochester Institute
of Technology



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

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

Copyright 2019 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Brief Contents

ABOUT THE AUTHORS XIX
PREFACE XXIII

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

Contents

ABOUT THE AUTHORS XIX

PREFACE XXIII

CHAPTER 1 Introduction 2

1.1 Decision Making 4

1.2 Business Analytics Defined 5

1.3 A Categorization of Analytical Methods and Models 6

 Descriptive Analytics 6

 Predictive Analytics 6

 Prescriptive Analytics 7

1.4 Big Data 7

 Volume 9

 Velocity 9

 Variety 9

 Veracity 9

1.5 Business Analytics in Practice 11

 Financial Analytics 11

 Human Resource (HR) Analytics 12

 Marketing Analytics 12

 Health Care Analytics 12

 Supply-Chain Analytics 13

 Analytics for Government and Nonprofits 13

 Sports Analytics 13

 Web Analytics 14

Summary 14

Glossary 15

CHAPTER 2 Descriptive Statistics 18

Analytics in Action: U.S. Census Bureau 19

2.1 Overview of Using Data: Definitions and Goals 19

2.2 Types of Data 21

 Population and Sample Data 21

 Quantitative and Categorical Data 21

 Cross-Sectional and Time Series Data 21

 Sources of Data 21

2.3 Modifying Data in Excel 24

 Sorting and Filtering Data in Excel 24

 Conditional Formatting of Data in Excel 27

2.4 Creating Distributions from Data 29

 Frequency Distributions for Categorical Data 29

 Relative Frequency and Percent Frequency Distributions 30

 Frequency Distributions for Quantitative Data 31

 Histograms 34

 Cumulative Distributions 37

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

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

CHAPTER 4 Descriptive Data Mining 138

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

Available in the MindTap Reader:

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

Appendix 4.7 *k*-Means Clustering with JMP Pro

Appendix 4.8 Association Rules with JMP Pro

Appendix 4.9 Text Mining with JMP Pro

CHAPTER 5 Probability: An Introduction to Modeling Uncertainty 166

Analytics in Action: National Aeronautics and Space Administration 167

5.1 Events and Probabilities 168

5.2 Some Basic Relationships of Probability 169

Complement of an Event 169

Addition Law 170

5.3 Conditional Probability 172

Independent Events 177

Multiplication Law 177

Bayes' Theorem 178

5.4 Random Variables 180

Discrete Random Variables 180

Continuous Random Variables 181

5.5 Discrete Probability Distributions 182

Custom Discrete Probability Distribution 182

Expected Value and Variance 184

Discrete Uniform Probability Distribution 187

Binomial Probability Distribution 188

Poisson Probability Distribution 191

5.6 Continuous Probability Distributions 194

Uniform Probability Distribution 194

Triangular Probability Distribution 196

Normal Probability Distribution 198

Exponential Probability Distribution 203

Summary 207

Glossary 207

Problems 209

Case Problem: Hamilton County Judges 218

CHAPTER 6 Statistical Inference 220

Analytics in Action: John Morrell & Company 221

6.1 Selecting a Sample 223

Sampling from a Finite Population 223

Sampling from an Infinite Population 224

6.2 Point Estimation 227

Practical Advice 229

6.3 Sampling Distributions 229

Sampling Distribution of \bar{x} 232

Sampling Distribution of \bar{p} 237

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

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

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

CHAPTER 9 Predictive Data Mining 422

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

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

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

Available in the MindTap Reader:

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

CHAPTER 12 Linear Optimization Models 556

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

CHAPTER 13 Integer Linear Optimization Models 606

Analytics in Action: Petrobras 607

13.1 Types of Integer Linear Optimization Models 607

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

13.4 Applications Involving Binary Variables 616

Capital Budgeting 616

Fixed Cost 618

Bank Location 621

Product Design and Market Share Optimization 623

13.5 Modeling Flexibility Provided by Binary Variables 626

Multiple-Choice and Mutually Exclusive Constraints 626

 k Out of n Alternatives Constraint 627

Conditional and Corequisite Constraints 627

13.6 Generating Alternatives in Binary Optimization 628

Summary 630

Glossary 631

Problems 632

Case Problem: Applecore Children's Clothing 643

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

Analytics in Action: InterContinental Hotels 647

14.1 A Production Application: Par, Inc. Revisited 647

An Unconstrained Problem 647

A Constrained Problem 648

Solving Nonlinear Optimization Models Using Excel Solver 650

Sensitivity Analysis and Shadow Prices in Nonlinear Models 651

14.2 Local and Global Optima 652

Overcoming Local Optima with Excel Solver 655

14.3 A Location Problem 657

14.4 Markowitz Portfolio Model 658

14.5 Forecasting Adoption of a New Product 663

Summary 666

Glossary 667

Problems 667

Case Problem: Portfolio Optimization with Transaction Costs 675

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

CHAPTER 15 Decision Analysis 678

Analytics in Action: Phytopharm 679

15.1 Problem Formulation 680

Payoff Tables 681

Decision Trees 681

15.2 Decision Analysis without Probabilities 682

Optimistic Approach 682

Conservative Approach 683

Minimax Regret Approach 683

15.3 Decision Analysis with Probabilities 685

Expected Value Approach 685

Risk Analysis 687

Sensitivity Analysis 688

15.4 Decision Analysis with Sample Information 689

Expected Value of Sample Information 694

Expected Value of Perfect Information 694

15.5 Computing Branch Probabilities with Bayes' Theorem 695

15.6 Utility Theory 698

Utility and Decision Analysis 699

Utility Functions 703

Exponential Utility Function 706

Summary 708

Glossary 708

Problems 710

Case Problem: Property Purchase Strategy 721

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