6TH EDITION

Business Analytics: Data Analysis and Decision Making

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

Wayne L. Winston
Kelly School of Business, Indiana University

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

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

Brief Contents

Preface xviii
1 Introduction to Business Analytics 1

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

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

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

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

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

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

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

Part 7 Bonus Online Material* 18-1

18 Importing Data into Excel 18-3

19 Analysis of Variance and Experimental Design 19-1

20 Statistical Process Control 20-1

Appendix A Statistical Reporting A-1

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

References 943

Index 945

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

Contents

Preface xviii

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

PART 1 EXPLORING DATA 17

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

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

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

PART 2 PROBABILITY AND DECISION MAKING UNDER UNCERTAINTY 137

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

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

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

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

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

PART 3 STATISTICAL INFERENCE 277

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

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

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

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

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

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

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

PART 4 REGRESSION ANALYSIS AND TIME SERIES FORECASTING 415

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

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

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

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

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

PART 5 Optimization And Simulation Modeling 597

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

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

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

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

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

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

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

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

PART 6 ADVANCED DATA ANALYSIS 895

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

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

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

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

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

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

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

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

References 943

Index 945

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