

BUSINESS ANALYTICS

Data Analysis and Decision Making

SEVENTH EDITION

S. Christian Albright

*Kelly School of Business,
Indiana University, Emeritus*

Wayne L. Winston

*Kelly School of Business,
Indiana University, Emeritus*



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

BRIEF CONTENTS

Preface xvi

1 Introduction to Business Analytics 1

PART 1 Data Analysis 37

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

PART 2 Probability and Decision Making under Uncertainty 183

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

PART 3 Statistical Inference 293

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

PART 4 Regression Analysis and Time Series Forecasting 411

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

PART 5 Optimization and Simulation Modeling 575

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

PART 6 Advanced Data Analysis 837

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

References 873

Index 875



CONTENTS

Preface xvi

1 Introduction to Business Analytics 1

1-1 Introduction 3

1-2 Overview of the Book 4

1-2a The Methods 4

1-2b The Software 6

1-3 Introduction to Spreadsheet Modeling 8

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

1-3b Cost Projections 12

1-3c Breakeven Analysis 15

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

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

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

1-4 Conclusion 33

PART 1 Data Analysis

37

2 Describing the Distribution of a Variable 38

2-1 Introduction 39

2-2 Basic Concepts 41

2-2a Populations and Samples 41

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

2-2c Data Types 42

2-3 Summarizing Categorical Variables 45

2-4 Summarizing Numeric Variables 49

2-4a Numeric Summary Measures 49

2-4b Charts for Numeric Variables 57

2-5 Time Series Data 62

2-6 Outliers and Missing Values 69

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

2-8 Conclusion 77

Appendix: Introduction to StatTools 83

3 Finding Relationships among Variables 84

3-1 Introduction 85

3-2 Relationships among Categorical Variables 86

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

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

5-5 The Binomial Distribution 214

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

5-6 The Poisson and Exponential Distributions 226

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

5-7 Conclusion 231

6 Decision Making under Uncertainty 242

6-1 Introduction 243

6-2 Elements of Decision Analysis 244

6-3 EMV and Decision Trees 247

6-4 One-Stage Decision Problems 251

6-5 The PrecisionTree Add-In 254

6-6 Multistage Decision Problems 257

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

6-7 The Role of Risk Aversion 274

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

6-8 Conclusion 280

PART 3 Statistical Inference

293

7 Sampling and Sampling Distributions 294

7-1 Introduction 295

7-2 Sampling Terminology 295

7-3 Methods for Selecting Random Samples 297

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

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

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

9-5 Tests for Normality 395

9-6 Chi-Square Test for Independence 401

9-7 Conclusion 404

PART 4 Regression Analysis and Time Series Forecasting 411

10 Regression Analysis: Estimating Relationships 412

10-1 Introduction 413

10-2 Scatterplots: Graphing Relationships 415

10-3 Correlations: Indicators of Linear Relationships 422

10-4 Simple Linear Regression 424

10-4a Least Squares Estimation 424

10-4b Standard Error of Estimate 431

10-4c R-Square 432

10-5 Multiple Regression 435

10-5a Interpretation of Regression Coefficients 436

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

10-6 Modeling Possibilities 442

10-6a Dummy Variables 442

10-6b Interaction Variables 448

10-6c Nonlinear Transformations 452

10-7 Validation of the Fit 461

10-8 Conclusion 463

11 Regression Analysis: Statistical Inference 472

11-1 Introduction 473

11-2 The Statistical Model 474

11-3 Inferences About the Regression Coefficients 477

11-3a Sampling Distribution of the Regression Coefficients 478

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

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

11-4 Multicollinearity 485

11-5 Include/Exclude Decisions 489

11-6 Stepwise Regression 494

11-7 Outliers 499

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

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

14 Optimization Models 630

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

15 Introduction to Simulation Modeling 717

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

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

16 Simulation Models 779

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

PART 6 Advanced Data Analysis

837

17 Data Mining 838

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

17-3 Clustering Methods 860

17-4 Conclusion 870

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

18-1 Introduction 18-2

18-2 One-Way ANOVA 18-5

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

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

18-2c Using a Logarithmic Transformation 18-11

18-3 Using Regression to Perform ANOVA 18-15

18-4 The Multiple Comparison Problem 18-18

18-5 Two-Way ANOVA 18-22

18-5a Confidence Intervals for Contrasts 18-28

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

18-6 More About Experimental Design 18-32

18-6a Randomization 18-32

18-6b Blocking 18-35

18-6c Incomplete Designs 18-38

18-7 Conclusion 18-40

19 Statistical Process Control (MindTap Reader only)

19-1 Introduction 19-2

19-2 Deming's 14 Points 19-3

19-3 Introduction to Control Charts 19-6

19-4 Control Charts for Variables 19-8

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

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

19-4c Rational Subsamples 19-16

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

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

19-5 Control Charts for Attributes 19-26

19-5a *P* Charts 19-26

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

19-6 Process Capability 19-33

19-6a Process Capability Indexes 19-35

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

19-7 Conclusion 19-43

APPENDIX A: Quantitative Reporting (MindTap Reader only)

A-1 Introduction A-1

A-2 Suggestions for Good Quantitative Reporting A-2

A-2a Planning A-2

A-2b Developing a Report A-3

A-2c Be Clear A-4

A-2d Be Concise A-4

A-2e Be Precise A-5

A-3 Examples of Quantitative Reports A-6

A-4 Conclusion A-16

References 873

Index 875