

# Contents

Preface xiii

## 1 Introduction to Data Analysis and Decision Making 1

- 1-1 Introduction 2
- 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 11
  - 1-3b Algebraic Models 11
  - 1-3c Spreadsheet Models 12
  - 1-3d A Seven-Step Modeling Process 13
- 1-4 Conclusion 15

2-7 Excel Tables for Filtering, Sorting, and Summarizing 63

2-8 Conclusion 71

## 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 87
- 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 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 27
- 2-4 Descriptive Measures for Numerical Variables 30
  - 2-4a Numerical Summary Measures 31
  - 2-4b Numerical Summary Measures with StatTools 40
  - 2-4c 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

## 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 178
- 5-3 Applications of the Normal Distribution 178
- 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 194
  - 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 Payoff Tables 225
  - 6-2b Possible Decision Criteria 226
  - 6-2c Expected Monetary Value (EMV) 227
  - 6-2d Sensitivity Analysis 228
  - 6-2e Decision Trees 229
  - 6-2f Risk Profiles 231
- 6-3 The PrecisionTree Add-In 239
- 6-4 Bayes' Rule 252
- 6-5 Multistage Decision Problems and the Value of Information 256
  - 6-5a The Value of Information 260

## 6-6 Risk Aversion and Expected Utility 272

- 6-6a Utility Functions 273
- 6-6b Exponential Utility 273
- 6-6c Certainty Equivalents 277
- 6-6d Is Expected Utility Maximization Used? 279
- 6-7 Conclusion 280

## PART 3 STATISTICAL INFERENCE 299

## 7 Sampling and Sampling Distributions 301

- 7-1 Introduction 302
- 7-2 Sampling Terminology 302
- 7-3 Methods for Selecting Random Samples 304
  - 7-3a Simple Random Sampling 304
  - 7-3b Systematic Sampling 309
  - 7-3c Stratified Sampling 310
  - 7-3d Cluster Sampling 313
  - 7-3e Multistage Sampling Schemes 314
- 7-4 Introduction to Estimation 315
  - 7-4a Sources of Estimation Error 316
  - 7-4b Key Terms in Sampling 317
  - 7-4c Sampling Distribution of the Sample Mean 318
  - 7-4d The Central Limit Theorem 323
  - 7-4e Sample Size Selection 328
  - 7-4f Summary of Key Ideas for Simple Random Sampling 329
- 7-5 Conclusion 331

## 8 Confidence Interval Estimation 335

- 8-1 Introduction 336
- 8-2 Sampling Distributions 338
  - 8-2a The  $t$  Distribution 338
  - 8-2b Other Sampling Distributions 341
- 8-3 Confidence Interval for a Mean 342
- 8-4 Confidence Interval for a Total 348
- 8-5 Confidence Interval for a Proportion 351
- 8-6 Confidence Interval for a Standard Deviation 357

8-7	Confidence Interval for the Difference between Means	360
8-7a	Independent Samples	361
8-7b	Paired Samples	369
8-8	Confidence Interval for the Difference between Proportions	374
8-9	Sample Size Selection	381
8-9a	Sample Size Selection for Estimation of the Mean	382
8-9b	Sample Size Selection for Estimation of Other Parameters	384
8-10	Conclusion	388

## 9 Hypothesis Testing 401

9-1	Introduction	402
9-2	Concepts in Hypothesis Testing	403
9-2a	Null and Alternative Hypotheses	404
9-2b	One-Tailed Versus Two-Tailed Tests	405
9-2c	Types of Errors	405
9-2d	Significance Level and Rejection Region	406
9-2e	Significance from $p$ -values	406
9-2f	Type II Errors and Power	408
9-2g	Hypothesis Tests and Confidence Intervals	409
9-2h	Practical versus Statistical Significance	409
9-3	Hypothesis Tests for a Population Mean	410
9-4	Hypothesis Tests for Other Parameters	418
9-4a	Hypothesis Tests for a Population Proportion	418
9-4b	Hypothesis Tests for Differences between Population Means	421
9-4c	Hypothesis Test for Equal Population Variances	431
9-4d	Hypothesis Tests for Differences between Population Proportions	432
9-5	Tests for Normality	440
9-6	Chi-Square Test for Independence	446
9-7	Conclusion	451

## PART 4

## REGRESSION ANALYSIS AND TIME SERIES FORECASTING 461

## 10 Regression Analysis: Estimating Relationships 463

10-1	Introduction	464
10-2	Scatterplots: Graphing Relationships	467
10-2a	Linear versus Nonlinear Relationships	472
10-2b	Outliers	472
10-2c	Unequal Variance	473
10-2d	No Relationship	473
10-3	Correlations: Indicators of Linear Relationships	474
10-4	Simple Linear Regression	476
10-4a	Least Squares Estimation	476
10-4b	Standard Error of Estimate	483
10-4c	The Percentage of Variation Explained: R-Square	484
10-5	Multiple Regression	487
10-5a	Interpretation of Regression Coefficients	488
10-5b	Interpretation of Standard Error of Estimate and R-Square	491
10-6	Modeling Possibilities	494
10-6a	Dummy Variables	494
10-6b	Interaction Variables	501
10-6c	Nonlinear Transformations	505
10-7	Validation of the Fit	520
10-8	Conclusion	522

## 11 Regression Analysis: Statistical Inference 532

11-1	Introduction	534
11-2	The Statistical Model	534
11-3	Inferences About the Regression Coefficients	538
11-3a	Sampling Distribution of the Regression Coefficients	539

11-3b	Hypothesis Tests for the Regression Coefficients and $p$ -Values	541
11-3c	A Test for the Overall Fit: The ANOVA Table	542
11-4	Multicollinearity	547
11-5	Include/Exclude Decisions	551
11-6	Stepwise Regression	556
11-7	Outliers	561
11-8	Violations of Regression Assumptions	566
11-8a	Nonconstant Error Variance	566
11-8b	Nonnormality of Residuals	567
11-8c	Autocorrelated Residuals	568
11-9	Prediction	570
11-10	Conclusion	576

## 12 Time Series Analysis and Forecasting 590

12-1	Introduction	591
12-2	Forecasting Methods: An Overview	592
12-2a	Extrapolation Models	592
12-2b	Econometric Models	593
12-2c	Combining Forecasts	594
12-2d	Components of Time Series Data	594
12-2e	Measures of Accuracy	597
12-3	Testing for Randomness	599
12-3a	The Runs Test	601
12-3b	Autocorrelation	604
12-4	Regression-Based Trend Models	607
12-4a	Linear Trend	607
12-4b	Exponential Trend	610
12-5	The Random Walk Model	615
12-6	Moving Averages Forecasts	619
12-7	Exponential Smoothing Forecasts	624
12-7a	Simple Exponential Smoothing	624
12-7b	Holt's Model for Trend	629
12-8	Seasonal Models	634
12-8a	Winters' Exponential Smoothing Model	635
12-8b	Deseasonalizing: The Ratio-to-Moving-Averages Method	639
12-8c	Estimating Seasonality with Regression	643
12-9	Conclusion	648

## PART 5

## OPTIMIZATION AND SIMULATION MODELING 655

### 13 Introduction to Optimization Modeling 657

13-1	Introduction	658
13-2	Introduction to Optimization	659
13-3	A Two-Variable Product Mix Model	660
13-4	Sensitivity Analysis	672
13-4a	Solver's Sensitivity Report	673
13-4b	SolverTable Add-In	676
13-4c	Comparison of Solver's Sensitivity Report and SolverTable	682
13-5	Properties of Linear Models	683
13-5a	Proportionality	683
13-5b	Additivity	683
13-5c	Divisibility	684
13-5d	Discussion of Linear Properties	684
13-5e	Linear Models and Scaling	685
13-6	Infeasibility and Unboundedness	685
13-6a	Infeasibility	685
13-6b	Unboundedness	686
13-6c	Comparison of Infeasibility and Unboundedness	686
13-7	A Larger Product Mix Model	688
13-8	A Multiperiod Production Model	696
13-9	A Comparison of Algebraic and Spreadsheet Models	704
13-10	A Decision Support System	705
13-11	Conclusion	707

### 14 Optimization Models 718

14-1	Introduction	719
14-2	Worker Scheduling Models	720
14-3	Blending Models	727
14-4	Logistics Models	733
14-4a	Transportation Models	734
14-4b	Other Logistics Models	742
14-5	Aggregate Planning Models	750
14-6	Financial Models	759
14-7	Integer Optimization Models	770
14-7a	Capital Budgeting Models	770

14-7b	Fixed-Cost Models	775
14-7c	Set-Covering Models	782
14-8	Nonlinear Optimization Models	789
14-8a	Basic Ideas of Nonlinear Optimization	789
14-8b	Managerial Economics Models	790
14-8c	Portfolio Optimization Models	793
14-9	Conclusion	802
<b>15</b>	<b>Introduction to Simulation Modeling</b>	<b>812</b>
15-1	Introduction	813
15-2	Probability Distributions for Input Variables	815
15-2a	Types of Probability Distributions	816
15-2b	Common Probability Distributions	819
15-2c	Using @RISK to Explore Probability Distributions	824
15-3	Simulation and the Flaw of Averages	833
15-4	Simulation with Built-in Excel Tools	836
15-5	Introduction to the @RISK	847
15-5a	@RISK Features	847
15-5b	Loading @RISK	848
15-5c	@RISK Models with a Single Random Input Variable	848
15-5d	Some Limitations of @RISK	857
15-5e	@RISK Models with Several Random Input Variables	858
15-6	The Effects of Input Distributions on Results	863
15-6a	Effect of the Shape of the Input Distribution(s)	863
15-6b	Effect of Correlated Input Variables	866
15-7	Conclusion	871
<b>16</b>	<b>Simulation Models</b>	<b>880</b>
16-1	Introduction	882
16-2	Operations Models	882
16-2a	Bidding for Contracts	882
16-2b	Warranty Costs	886
16-2c	Drug Production with Uncertain Yield	890
16-3	Financial Models	896
16-3a	Financial Planning Models	897

16-3b	Cash Balance Models	901
16-3c	Investment Models	906
16-4	Marketing Models	912
16-4a	Models of Customer Loyalty	913
16-4b	Marketing and Sales Models	921
16-5	Simulating Games of Chance	927
16-5a	Simulating the Game of Craps	927
16-5b	Simulating the NCAA Basketball Tournament	930
16-6	An Automated Template for @Risk Models	934
16-7	Conclusion	935

## PART 6

## ADVANCED DATA ANALYSIS 945

<b>17</b>	<b>Data Mining</b>	<b>947</b>
17-1	Introduction	948
17-2	Data Exploration and Visualization	949
17-2a	Online Analytical Processing (OLAP)	950
17-2b	PowerPivot and Power View in Excel 2013	954
17-2c	Visualization Software	962
17-3	Microsoft Data Mining Add-Ins for Excel	962
17-4	Classification Methods	964
17-4a	Logistic Regression	965
17-4b	Neural Networks	971
17-4c	Classification Trees	976
17-4d	Classification and Lift	977
17-4e	Classification with Rare Events	978
17-5	Clustering	980
17-6	Conclusion	982

## PART 7

## BONUS ONLINE MATERIAL

<b>18</b>	<b>Importing Data into Excel</b>	<b>18-3</b>
18-1	Introduction	18-4
18-2	Rearranging Excel Data	18-5

18-3	Importing Text Data	18-9	20-3	Introduction to Control Charts	20-7
18-4	Importing Relational Database Data	18-15	20-4	Control Charts for Variables	20-9
18-4a	Introduction to Relational Databases	18-15	20-4a	Control Charts and Hypothesis Testing	20-14
18-4b	Using Microsoft Query	18-16	20-4b	Other Out-of-Control Indications	20-16
18-4c	SQL Statements	18-28	20-4c	Rational Subsamples	20-17
18-5	Web Queries	18-30	20-4d	Deming's Funnel Experiment and Tampering	20-19
18-6	Cleansing Data	18-34	20-4e	Control Charts in the Service Industry	20-23
18-7	Conclusion	18-41	20-5	Control Charts for Attributes	20-27
19	Analysis of Variance and Experimental Design	19-1	20-5a	The $p$ Chart	20-27
19-1	Introduction	19-2	20-5b	The Red Bead Experiment	20-31
19-2	One-Way ANOVA	19-5	20-6	Process Capability	20-34
19-2a	The Equal-Means Test	19-5	20-6a	Process Capability Indexes	20-37
19-2b	Confidence Intervals for Differences between Means	19-8	20-6b	More on Motorola and 6-sigma	20-42
19-2c	Using a Logarithmic Transformation	19-11	20-7	Conclusion	20-45
19-3	Using Regression to Perform ANOVA	19-17	Appendix A: Statistical Reporting	A-1	
19-4	The Multiple Comparison Problem	19-20	A-1	Introduction	A-1
19-5	Two-Way ANOVA	19-24	A-2	Suggestions for Good Statistical Reporting	A-2
19-5a	Confidence Intervals for Contrasts	19-31	A-2a	Planning	A-2
19-5b	Assumptions of Two-Way ANOVA	19-34	A-2b	Developing a Report	A-3
19-6	More About Experimental Design	19-35	A-2c	Be Clear	A-4
19-6a	Randomization	19-36	A-2d	Be Concise	A-5
19-6b	Blocking	19-38	A-2e	Be Precise	A-5
19-6c	Incomplete Designs	19-42	A-3	Examples of Statistical Reports	A-7
19-7	Conclusion	19-45	A-4	Conclusion	A-18
20	Statistical Process Control	20-1	References	985	
20-1	Introduction	20-3	Index	987	
20-2	Deming's 14 Points	20-4			