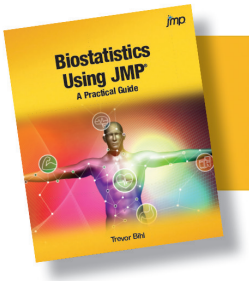

Contents

Dedication ... iii
Acknowledgments ... xi
About This Book ... xiii
About the Author ... xvii
Chapter 1: Introduction ... 1
1.1 Background and Overview ... 1
1.2 Getting Started with JMP ... 2
1.3 General Outline ... 4
1.4 How to Use This Book ... 5
1.5 Reference ... 5
Chapter 2: Data Wrangling: Data Collection ... 7
2.1 Introduction ... 7
2.2 Collecting Data from Files ... 8
2.2.1 JMP Native Files ... 8
2.2.2 SAS Format Files ... 9
2.2.3 Excel Spreadsheets ... 10
2.2.4 Text and CSV Format ... 11
2.3 Extracting Data from Internet Locations ... 14
2.3.1 Opening as Data ... 14
2.3.2 Opening as a Webpage ... 15
2.4 Data Modeling Types ... 17
2.4.1 Incorporating Expression and Contextual Data ... 18
2.5 References ... 19
Chapter 3: Data Wrangling: Data Cleaning ... 21
3.1 Introduction ... 21
3.2 Tables ... 21
3.2.1 Stacking Columns ... 24
3.2.2 Basic Table Organization ... 26
3.2.3 Column Properties ... 31
3.3 The Sorted Array ... 32

3.4 Restructuring Data 34
3.4.1 Combining Columns 35
3.4.2 Separating Out a Column (Text to Columns) 36
3.4.3 Creating Indicator Columns 36
3.4.4 Grouping Inside Columns 38
3.5 References 41
Chapter 4: Initial Data Analysis with Descriptive Statistics 45
4.1 Introduction 45
4.2 Histograms and Distributions 45
4.2.1 Histograms 46
4.2.2 Box Plots 55
4.2.3 Stem-and-Leaf Plots 57
4.2.4 Pareto Charts 58
4.3 Descriptive Statistics 64
4.3.1 Sample Mean and Standard Deviation 66
4.3.2 Additional Statistical Measures 67
4.4 References 69
Chapter 5: Data Visualization Tools 71
5.1 Introduction 71
5.2 Scatter Plots 72
5.2.1 Coloring Points 75
5.2.2 Copying Better-Looking Figures 77
5.2.3 Multiple Scatter Plots 79
5.3 Charts 81
5.4 Multidimensional Plots 84
5.4.1 Parallel Plots 84
5.4.2 Cell Plots 87
5.5 Multivariate and Correlations Tool 89
5.5.1 Correlation Table 91
5.5.2 Correlation Heat Maps 92
5.5.3 Simple Statistics 93
5.5.4 Additional Multivariate Measures 93
5.6 Graph Builder and Custom Figures 94
5.6.1 Graph Builder Custom Colors 96
5.6.2 Incorporating Contextual Data 98
5.7 References 99

Chapter 6: Rates, Proportions, and Epidemiology 101
6.1 Introduction 101
6.2 Rates 101
6.2.1 Crude Rates 101
6.2.2 Adjusted Rates 105
6.3 Geographic Visualizations 108
6.3.1 National Visualizations 108
6.3.2 County and Lower Level Visualizations 116
6.4 References 120
Chapter 7: Statistical Tests and Confidence Intervals 123
7.1 Introduction 123
7.1.1 General Hypothesis Test Background 124
7.1.2 Selecting the Appropriate Method 125
7.2 Testing for Normality 126
7.2.1 Histogram Analysis 126
7.2.2 Normal Quantile/Probability Plot 128
7.2.3 Goodness-of-Fit Tests 131
7.2.4 Goodness-of-Fit for Other Distributions 132
7.3 General Hypothesis Tests 133
7.3.1 Z-Test Hypothesis Test of Mean 133
7.3.2 T-Test Hypothesis Test of Mean 135
7.3.3 Nonparametric Test of Mean (Wilcoxon Signed Rank) 136
7.3.4 Standard Deviation Hypothesis Test 140
7.3.5 Tests of Proportions 141
7.4 Confidence Intervals 144
7.4.1 Mean Confidence Intervals 144
7.4.2 Mean Confidence Intervals with Different Thresholds 144
7.4.3 Confidence Intervals for Proportions 145
7.5 Chi-Squared Analysis of Frequency and Contingency Tables 146
7.6 Two Sample Tests 150
7.6.1 Comparing Two Group Means 150
7.6.2 Paired Comparison, Matched Pairs 154
7.7 References 156
Chapter 8: Analysis of Variance (ANOVA) and Design of Experiments (DoE) 159
8.1 Introduction 159

8.2 One-Way ANOVA 161
8.2.1 One-Way ANOVA with Fit Y by X 161
8.2.2 Means Comparison, LSD Matrix, and Connecting Letters 165
8.2.3 Fit Y by X Changing Significance Levels 168
8.2.4 Multiple Comparisons, Multiple One-Way ANOVAs 169
8.2.5 One-Way ANOVA via Fit Model 171
8.2.6 One-Way ANOVA for Unequal Group Sizes (Unbalanced) 176
8.3 Blocking 179
8.3.1 One-Way ANOVA with Blocking via Fit Y by X 179
8.3.2 One-Way ANOVA with Blocking via Fit Model 182
8.3.3 Note on Blocking 183
8.4 Multiple Factors 183
8.4.1 Experimental Design Considerations 184
8.4.2 Multiple ANOVA 188
8.4.3 Feature Selection and Parsimonious Models 191
8.5 Multivariate ANOVA (MANOVA) and Repeated Measures 196
8.5.1 Repeated Measures MANOVA Background 196
8.5.2 MANOVA in Fit Model 197
8.6 References 201
Chapter 9: Regression and Curve Fitting 205
9.1 Introduction 205
9.2 Simple Linear Regression 206
9.2.1 Fit Y by X for Bivariate Fits (One X and One Y) 206
9.2.2 Special Fitting Tools 208
9.3 Multiple Regression 211
9.3.1 Fit Model 211
9.3.2 Stepwise Feature Selection 214
9.3.3 Analysis of Covariance (ANCOVA) 222
9.4 Nonlinear Curve Fitting and a Nonlinear Platform Example 226
9.5 References 232
Chapter 10: Diagnostic Methods for Regression, Curve Fitting, and ANOVA 233
10.1 Introduction 233
10.2 Computing Residuals with Fit Y by X and Fit Model 234
10.2.1 Fit Y by X 234
10.2.2 Fit Model 234
10.3 Checking for Normality 235

10.4 Checking for Nonconstant Error Variance (Heteroscedasticity) 236
10.5 Checking for Outliers 238
10.6 Checking for Nonindependence 242
10.7 Multiple Factor Diagnostics 243
10.8 Nonlinear Fit Residuals 245
10.9 Developing Appropriate Models 246
10.10 References 247
Chapter 11: Categorical Data Analysis 249
11.1 Introduction 249
11.2 Clustering 250
11.2.1 Hierarchical Clustering 250
11.2.2 K-means Clustering 260
11.3 Classification 263
11.3.1 JMP Data Preliminaries for Classification 265
11.3.2 Example Data Sets 267
11.4 Classification by Logistic Regression 268
11.4.1 Logistic Regression in Fit Y by X 268
11.4.2 Logistic Regression in Fit Model 270
11.5 Classification by Discriminant Analysis 273
11.5.1 Discriminant Analysis Loadings 275
11.5.2 Stepwise Discriminant Analysis 276
11.6 Classification with Tabulated Data 277
11.7 Classifier Performance Verification 280
11.8 References 284
Chapter 12: Advanced Modeling Methods 287
12.1 Introduction 287
12.2 Principal Components and Factor Analysis 288
12.2.1 Principal Components in JMP 288
12.2.2 Dimensionality Assessment 291
12.2.3 Factor Analysis in JMP 293
12.3 Partial Least Squares 296
12.4 Decision Trees 302
12.4.1 Classification Decision Trees in JMP 303
12.4.2 Predictive Decision Trees in JMP 308
12.5 Artificial Neural Networks 310
12.5.1 Neural Network Architecture 311
12.5.2 Classification Neural Networks in JMP 312
12.5.3 Predictive Neural Networks in JMP 315

12.6 Control Charts 317
12.7 References 321
Chapter 13: Survival Analysis 323
13.1 Introduction 323
13.2 Life Distributions 323
13.3 Kaplan-Meier Curves 327
13.3.1 Simple Survival Analysis 327
13.3.2 Multiple Groups 330
13.3.3 Censoring 331
13.3.4 Proportional Hazards 335
13.4 References 336
Chapter 14: Collaboration and Additional Functionality 339
14.1 Introduction 339
14.2 Saving Scripts and SAS Coding 339
14.2.1 Saving Scripts to Data Table 340
14.2.2 SAS Coding Functionality 341
14.3 Collaboration 342
14.3.1 Journals 342
14.3.2 Web Reports 344
14.4 Add-Ins 347
14.4.1 Finding Add-Ins 347
14.4.2 Developing Add-Ins 348
14.4.3 Example Add-In: Forest Plot / Meta-analysis 348
14.4.4 Add-In Version Control 351
14.5 References 352
Index 331