Contents

Preface ix

Chapter 1 Introduction to SAS Enterprise Guide 1

1.1 What Is SAS Enterprise Guide? 2
1.2 Using This Book 3
1.3 The SAS Enterprise Guide Interface 4
1.3.1 SAS Enterprise Guide Projects 5
1.3.2 The User Interface 5
1.3.3 The Process Flow 6
1.3.4 The Active Data Set 8
1.4 Creating a Project 9
1.4.1 Opening a SAS Data Set 9
1.4.2 Importing Data 10
1.5 Modifying Data 15
1.5.1 Modifying Variables: Using Queries 15
1.5.2 Recoding Variables 18
1.5.3 Splitting Data Sets: Using Filters 20
1.5.4 Concatenating and Merging Data Sets: Appends and Joins 21
1.5.5 Names of Data Sets and Variables in SAS and SAS Enterprise Guide 26
1.5.6 Storing SAS Data Sets: Libraries 27
1.6 Statistical Analysis Tasks 28
1.7 Graphs 30
1.8 Running Parts of the Process Flow 30

Chapter 2 Data Description and Simple Inference 31

2.1 Introduction 32

2.2 Example: Guessing the Width of a Room: Analysis of Room Width Guesses 32

2.2.1 Initial Analysis of Room Width Guesses Using Simple Summary Statistics and Graphics 33

2.2.2 Guessing the Width of a Room: Is There Any Difference in Guesses Made in Feet and in Meters? 40

2.2.3 Checking the Assumptions Made When Using Student's *t*-Test and Alternatives to the *t*-Test 47

2.3 Example: Wave Power and Mooring Methods 49

2.3.1 Initial Analysis of Wave Energy Data Using Box Plots 50

2.3.2 Wave Power and Mooring Methods: Do Two Mooring Methods Differ in Bending Stress? 54

2.3.3 Checking the Assumptions of the Paired *t*-Tests 56

2.4 Exercises 57

Chapter 3 Dealing with Categorical Data 61

3.1 Introduction 61

3.2 Example: Horse Race Winners 62

3.2.1 Looking at Horse Race Winners Using Some Simple Graphics: Bar Charts and Pie Charts 62

3.2.2 Horse Race Winners: Does Starting Stall Position Predict Horse Race Winners? 66

3.3 Example: Brain Tumors 68

3.3.1 Tabulating the Brain Tumor Data into a Contingency Table 69

3.3.2 Do Different Types of Brain Tumors Occur More Frequently at Particular Sites? The Chi-Square Test 70

3.4 Example: Suicides and Baiting Behavior 71

3.4.1 How Is Baiting Behavior at Suicides Affected by Season? Fisher's Exact Test 71

3.5 Example: Juvenile Felons 74

3.5.1 Juvenile Felons: Where Should They Be Tried? McNemar's Test 75

3.6 Exercises 74

Chapter 4 Dealing with Bivariate Data 79

4.1 Introduction 80
4.2 Example: Heights and Resting Pulse Rates 80
4.2.1 Plotting Heights and Resting Pulse Rates: The Scatterplot 81
4.2.2 Quantifying the Relationship between Resting Pulse Rate and Height: The Correlation Coefficient 82
4.2.3 Heights and Resting Pulse Rates: Simple Linear Regression 85
4.3 Example: An Experiment in Kinesiology 90
4.3.1 Oxygen Uptake and Expired Ventilation: The Scatterplot 91
4.3.2 Expired Ventilation and Oxygen Uptake: Is Simple Linear Regression Appropriate? 93
4.4 Example: U.S. Birthrates in the 1940s 95
4.4.1 Plotting the Birthrate Data: The Aspect Ratio of a Scatterplot 95
4.5 Exercises 102

Chapter 5 Analysis of Variance 107

5.1 Introduction 108
5.2 Example: Teaching Arithmetic 108
5.2.1 Initial Examination of the Teaching Arithmetic Data with Summary Statistics and Box Plots 109
5.2.2 Teaching Arithmetic: Are Some Teaching Methods for Teaching Arithmetic Better Than Others? 112
5.3 Example: Weight Gain in Rats 116
5.3.1 A First Look at the Rat Weight Gain Data Using Box Plots and Numerical Summaries 116
5.3.2 Weight Gain in Rats: Do Rats Gain More Weight on a Particular Diet? 119
5.4 Example: Mother's Post-Natal Depression and Child's IQ 124
5.4.1 Summarizing the Post-Natal Depression Data 125
5.4.2 How Is a Child's IQ Affected by Post-Natal Depression in the Mother? 128
5.5 Exercises 133

Chapter 6 Multiple Linear Regression 139

6.1 Introduction 140
6.2 Example: Consuming Ice Cream 140
6.2.1 The Ice Cream Data: An Initial Analysis Using Scatterplots 141
6.2.2 Ice Cream Sales: Are They Most Affected by Price or Temperature? How to Tell Using Multiple Regression 143
6.2.3 Diagnosing the Multiple Regression Model Fitted to the Ice Cream Consumption Data: The Use of Residuals 146
6.3 Example: Making It Rain by Cloud Seeding 152
6.3.1 The Cloud Seeding Data: Initial Examination of the Data Using Box Plots and Scatterplots 154
6.3.2 When Is Cloud Seeding Best Carried Out? How to Tell Using Multiple Regression Models Containing Interaction Terms 158
6.3.3 Diagnosing the Fitted Model for the Cloud Seeding Data Using Residuals 164
6.4 Exercises 166

Chapter 7 Logistic Regression 171

7.1 Introduction 172
7.2 Example: Myocardial Infarctions 172
7.2.1 Myocardial Infarctions: What Predicts a Past History of Myocardial Infarctions? Answering the Question Using Logistic Regression 174
7.2.2 Odds 174
7.2.3 Applying the Logistic Regression Model with a Single Explanatory Variable 175
7.2.4 Interpreting the Regression Coefficient in the Fitted Logistic Regression Model 179
7.2.5 Applying the Logistic Regression Model Using SAS Enterprise Guide 180
7.3 Exercises 186

Chapter 8 Survival Analysis 191

8.1 Introduction 192

8.2 Example: Gastric Cancer 192

8.2.1 Gastric Cancer Patients: Summarizing and Displaying Their Survival Experience Using the Survival Function 193

8.2.2 Plotting Survival Functions Using SAS Enterprise Guide 194

8.2.3 Testing the Equality of Two Survival Functions: The Log-Rank Test 202

8.3 Example: Myeloblastic Leukemia 204

8.3.1 What Affects Survival in Patients with Leukemia? The Hazard Function and Cox Regression 207

8.3.2 Applying Cox Regression Using SAS Enterprise Guide 209

8.4 Exercises 213

References 215

Index 217