Contents

Acknowledgments vii

Chapter 1 - Regression Concepts 1

1.1 Statistical Background 1

1.2 Performing a Regression with the IML Procedure 9

1.3 Regression with the SAS System 12

Chapter 2 - Using the REG Procedure 15

2.1 Introduction 15

2.2 A Model with One Independent Variable 17

2.3 A Model with Several Independent Variables 20

2.4 Various MODEL Statement Options 24

2.5 Further Examination of Model Parameters 36

2.6 Plotting Results 43

2.7 Creating Data I: the OUTPUT Statement 52

2.8 Creating Data II: Other Data Sets 55

2.9 Creating Data III: ODS Output 57

2.10 Predicting to a Different Set of Data 57

2.11 Exact Collinearity: Linear Dependency 60

2.12 Summary 62

Chapter 3 - Observations 63

3.1 Introduction 63

3.2 Outlier Detection 64

3.3 Specification Errors 77

3.4 Heterogeneous Variances 81

3.5 Correlated Errors 86

3.6 Summary 94

Chapter 4 - Multicollinearity: Detection and Remedial Measures 95

4.1 Introduction 95

4.2 Detecting Multicollinearity 97

4.3 Model Restructuring 101

4.4 Variable Selection 108

4.5 Biased Estimation 120

4.6 Summary 125

Chapter 5 - Curve Fitting 127

5.1 Introduction 127

5.2 Polynomial Models with One Independent Variable 128

5.3 Polynomial Plots 132

5.4 Polynomial Models with Several Variables 134

5.5 Response Surface Plots 138

5.6 A Three-Factor Response Surface Experiment 142

5.7 Curve Fitting without a Model 148

5.8 Summary 155

Chapter 6 - Special Applications of Linear Models 157

6.1 Introduction 157

6.2 Multiplicative Models 158

6.3 Spline Models 164

6.4 Indicator Variables 169

6.5 Binary Response Variable: Logistic Regression 173

6.6 Summary 184

Chapter 7 - Nonlinear Models 185

7.1 Introduction 185

7.2 Estimating the Exponential Decay Model 186

7.3 Fitting a Growth Curve with the NLIN Procedure 192

7.4 Fitting Splines with Unknown Knots 198

7.5 Additional Comments on the NLIN Procedure 203

7.6 Summary 204

Chapter 8 - Using SAS/INSIGHT Software for Regression 207

8.1 Introduction 207

8.2 Multiple Linear Regression: the BOQ Data 207

8.3 A Polynomial Response Surface: the FISH Data 215

8.4 Logistic Regression: the DYSTRO Data 218

8.5 Nonparametric Smoothing: the Barbados Data 221

8.6 Summary 225

References 227

Index 229