Contents

1 Introduction **1**

1.1 Relationships 1
1.2 Determining Relationships: A Specific Problem 2
1.3 The Model 5
1.4 Least Squares 7
1.5 Another Example and a Special Case 10
1.6 When Is Least Squares a Good Method? 11
1.7 A Measure of Fit for Simple Regression 13
1.8 Mean and Variance of b_0 and b_1 14
1.9 Confidence Intervals and Tests 17
1.10 Predictions 18
Appendix to Chapter 1 20
Problems 23

2 Multiple Regression **28**

2.1 Introduction 28
2.2 Regression Model in Matrix Notation 28
2.3 Least Squares Estimates 30
2.4 Examples 31
2.5 Gauss-Markov Conditions 35
2.6 Mean and Variance of Estimates Under G-M Conditions . . 35
2.7 Estimation of σ^2 37
2.8 Measures of Fit 39
2.9 The Gauss-Markov Theorem 41
2.10 The Centered Model 42
2.11 Centering and Scaling 44
2.12 *Constrained Least Squares 44
Appendix to Chapter 2 46
Problems 49

3 Tests and Confidence Regions **60**

3.1 Introduction 60
3.2 Linear Hypothesis 60

3.3 *Likelihood Ratio Test . . . 62
3.4 *Distribution of Test Statistic . . . 64
3.5 Two Special Cases . . . 65
3.6 Examples . . . 66
3.7 Comparison of Regression Equations . . . 67
3.8 Confidence Intervals and Regions . . . 71
3.8.1 C.I. for the Expectation of a Predicted Value . . . 71
3.8.2 C.I. for a Future Observation . . . 71
3.8.3 *Confidence Region for Regression Parameters . . . 72
3.8.4 *C.I.'s for Linear Combinations of Coefficients . . . 73
Problems . . . 74

4 Indicator Variables **83**
4.1 Introduction . . . 83
4.2 A Simple Application . . . 83
4.3 Polychotomous Variables . . . 84
4.4 Continuous and Indicator Variables . . . 88
4.5 Broken Line Regression . . . 89
4.6 Indicators as Dependent Variables . . . 92
Problems . . . 95

5 The Normality Assumption **100**
5.1 Introduction . . . 100
5.2 Checking for Normality . . . 101
5.2.1 Probability Plots . . . 101
5.2.2 Tests for Normality . . . 105
5.3 Invoking Large Sample Theory . . . 106
5.4 *Bootstrapping . . . 107
5.5 *Asymptotic Theory . . . 108
Problems . . . 110

6 Unequal Variances **111**
6.1 Introduction . . . 111
6.2 Detecting Heteroscedasticity . . . 111
6.2.1 Formal Tests . . . 114
6.3 Variance Stabilizing Transformations . . . 115
6.4 Weighting . . . 118
Problems . . . 128

7 *Correlated Errors **132**
7.1 Introduction . . . 132
7.2 Generalized Least Squares: Case When Ω Is Known . . . 133
7.3 Estimated Generalized Least Squares . . . 134
7.3.1 Error Variances Unequal and Unknown . . . 134
7.4 Nested Errors . . . 136

7.5 The Growth Curve Model . . . 138
7.6 Serial Correlation . . . 140
7.6.1 The Durbin-Watson Test . . . 142
7.7 Spatial Correlation . . . 143
7.7.1 Testing for Spatial Correlation . . . 143
7.7.2 Estimation of Parameters . . . 144
Problems . . . 146

8 Outliers and Influential Observations 154
8.1 Introduction . . . 154
8.2 The Leverage . . . 155
8.2.1 *Leverage as Description of Remoteness . . . 156
8.3 The Residuals . . . 156
8.4 Detecting Outliers and Points That Do Not Belong to the Model . . . 157
8.5 Influential Observations . . . 158
8.5.1 Other Measures of Influence . . . 160
8.6 Examples . . . 161
Appendix to Chapter 8 . . . 173
Problems . . . 176

9 Transformations 180
9.1 Introduction . . . 180
9.1.1 An Important Word of Warning . . . 180
9.2 Some Common Transformations . . . 181
9.2.1 Polynomial Regression . . . 181
9.2.2 Splines . . . 182
9.2.3 Multiplicative Models . . . 182
9.2.4 The Logit Model for Proportions . . . 186
9.3 Deciding on the Need for Transformations . . . 188
9.3.1 Examining Residual Plots . . . 189
9.3.2 Use of Additional Terms . . . 190
9.3.3 Use of Repeat Measurements . . . 192
9.3.4 Daniel and Wood Near-Neighbor Approach . . . 194
9.3.5 Another Method Based on Near Neighbors . . . 195
9.4 Choosing Transformations . . . 197
9.4.1 Graphical Method: One Independent Variable . . . 197
9.4.2 Graphical Method: Many Independent Variables . . . 200
9.4.3 Analytic Methods: Transforming the Response . . . 204
9.4.4 Analytic Methods: Transforming the Predictors . . . 209
9.4.5 Simultaneous Power Transformations for Predictors and Response . . . 209
Appendix to Chapter 9 . . . 211
Problems . . . 213

10 Multicollinearity **218**
10.1 Introduction 218
10.2 Multicollinearity and Its Effects 218
10.3 Detecting Multicollinearity 222
10.3.1 Tolerances and Variance Inflation Factors 222
10.3.2 Eigenvalues and Condition Numbers 223
10.3.3 Variance Components 224
10.4 Examples 225
Problems 231

11 Variable Selection **233**
11.1 Introduction 233
11.2 Some Effects of Dropping Variables 234
11.2.1 Effects on Estimates of β_j 235
11.2.2 *Effect on Estimation of Error Variance 236
11.2.3 *Effect on Covariance Matrix of Estimates 236
11.2.4 *Effect on Predicted Values: Mallows' C_p 237
11.3 Variable Selection Procedures 238
11.3.1 Search Over All Possible Subsets 239
11.3.2 Stepwise Procedures 240
11.3.3 Stagewise and Modified Stagewise Procedures 243
11.4 Examples 243
Problems 251

12 *Biased Estimation **253**
12.1 Introduction 253
12.2 Principal Component Regression 253
12.2.1 Bias and Variance of Estimates 255
12.3 Ridge Regression 256
12.3.1 Physical Interpretations of Ridge Regression 257
12.3.2 Bias and Variance of Estimates 258
12.4 Shrinkage Estimator 261
Problems 263

A Matrices **267**
A.1 Addition and Multiplication 267
A.2 The Transpose of a Matrix 268
A.3 Null and Identity Matrices 268
A.4 Vectors 269
A.5 Rank of a Matrix 270
A.6 Trace of a Matrix 271
A.7 Partitioned Matrices 271
A.8 Determinants 272
A.9 Inverses 273
A.10 Characteristic Roots and Vectors 276

A.11 Idempotent Matrices 277
A.12 The Generalized Inverse 278
A.13 Quadratic Forms 279
A.14 Vector Spaces 280
Problems 282

B Random Variables and Random Vectors **284**
B.1 Random Variables 284
B.1.1 Independent Random Variables 284
B.1.2 Correlated Random Variables 285
B.1.3 Sample Statistics 285
B.1.4 Linear Combinations of Random Variables 285
B.2 Random Vectors 286
B.3 The Multivariate Normal Distribution 288
B.4 The Chi-Square Distributions 290
B.5 The F and t Distributions 292
B.6 Jacobian of Transformations 293
B.7 Multiple Correlation 295
Problems 297

C Nonlinear Least Squares **298**
C.1 Gauss-Newton Type Algorithms 299
C.1.1 The Gauss-Newton Procedure 299
C.1.2 Step Halving 300
C.1.3 Starting Values and Derivatives 301
C.1.4 Marquardt Procedure 302
C.2 Some Other Algorithms 302
C.2.1 Steepest Descent Method 303
C.2.2 Quasi-Newton Algorithms 303
C.2.3 The Simplex Method 305
C.2.4 Weighting 305
C.3 Pitfalls 306
C.4 Bias, Confidence Regions and Measures of Fit 308
C.5 Examples 310
Problems 316

Tables **319**

References **327**

Index **336**

Author Index **345**