

Contents

1	Introduction	1
1.1	Building Valid Models.....	1
1.2	Motivating Examples.....	1
1.2.1	Assessing the Ability of NFL Kickers.....	1
1.2.2	Newspaper Circulation	4
1.2.3	Menu Pricing in a New Italian Restaurant in New York City.....	5
1.2.4	Effect of Wine Critics' Ratings on Prices of Bordeaux Wines.....	8
1.3	Level of Mathematics	13
2	Simple Linear Regression	15
2.1	Introduction and Least Squares Estimates.....	15
2.1.1	Simple Linear Regression Models.....	15
2.2	Inferences About the Slope and the Intercept.....	20
2.2.1	Assumptions Necessary in Order to Make Inferences About the Regression Model.....	21
2.2.2	Inferences About the Slope of the Regression Line	21
2.2.3	Inferences About the Intercept of the Regression Line	23
2.3	Confidence Intervals for the Population Regression Line.....	24
2.4	Prediction Intervals for the Actual Value of Y	25
2.5	Analysis of Variance.....	27
2.6	Dummy Variable Regression.....	30
2.7	Derivations of Results.....	33
2.7.1	Inferences about the Slope of the Regression Line.....	34
2.7.2	Inferences about the Intercept of the Regression Line	35
2.7.3	Confidence Intervals for the Population Regression Line	36
2.7.4	Prediction Intervals for the Actual Value of Y	37
2.8	Exercises.....	38

3 Diagnostics and Transformations for Simple Linear Regression 45

3.1 Valid and Invalid Regression Models:

 Anscombe’s Four Data Sets 45

 3.1.1 Residuals..... 48

 3.1.2 Using Plots of Residuals to Determine Whether
 the Proposed Regression Model Is a Valid Model 49

 3.1.3 Example of a Quadratic Model..... 50

3.2 Regression Diagnostics: Tools for Checking
 the Validity of a Model..... 50

 3.2.1 Leverage Points..... 51

 3.2.2 Standardized Residuals..... 59

 3.2.3 Recommendations for Handling Outliers
 and Leverage Points 66

 3.2.4 Assessing the Influence of Certain Cases..... 67

 3.2.5 Normality of the Errors..... 69

 3.2.6 Constant Variance 71

3.3 Transformations..... 76

 3.3.1 Using Transformations to Stabilize Variance 76

 3.3.2 Using Logarithms to Estimate Percentage Effects 79

 3.3.3 Using Transformations to Overcome Problems
 due to Nonlinearity..... 83

3.4 Exercises..... 103

4 Weighted Least Squares 115

4.1 Straight-Line Regression Based on Weighted Least Squares..... 115

 4.1.1 Prediction Intervals for Weighted Least Squares..... 118

 4.1.2 Leverage for Weighted Least Squares 118

 4.1.3 Using Least Squares to Calculate Weighted Least Squares.... 119

 4.1.4 Defining Residuals for Weighted Least Squares 121

 4.1.5 The Use of Weighted Least Squares 121

4.2 Exercises..... 122

5 Multiple Linear Regression..... 125

5.1 Polynomial Regression 125

5.2 Estimation and Inference in Multiple Linear Regression..... 130

5.3 Analysis of Covariance..... 140

5.4 Exercises..... 146

6 Diagnostics and Transformations for Multiple Linear Regression 151

6.1 Regression Diagnostics for Multiple Regression 151

 6.1.1 Leverage Points in Multiple Regression..... 152

 6.1.2 Properties of Residuals in Multiple Regression..... 154

 6.1.3 Added Variable Plots 162

6.2	Transformations.....	167
6.2.1	Using Transformations to Overcome Nonlinearity.....	167
6.2.2	Using Logarithms to Estimate Percentage Effects: Real Valued Predictor Variables.....	184
6.3	Graphical Assessment of the Mean Function Using Marginal Model Plots.....	189
6.4	Multicollinearity.....	195
6.4.1	Multicollinearity and Variance Inflation Factors.....	203
6.5	Case Study: Effect of Wine Critics' Ratings on Prices of Bordeaux Wines.....	203
6.6	Pitfalls of Observational Studies Due to Omitted Variables.....	210
6.6.1	Spurious Correlation Due to Omitted Variables.....	210
6.6.2	The Mathematics of Omitted Variables.....	213
6.6.3	Omitted Variables in Observational Studies.....	214
6.7	Exercises.....	215
7	Variable Selection.....	227
7.1	Evaluating Potential Subsets of Predictor Variables.....	228
7.1.1	Criterion 1: R^2 -Adjusted.....	228
7.1.2	Criterion 2: AIC, Akaike's Information Criterion.....	230
7.1.3	Criterion 3: AIC_c , Corrected AIC.....	231
7.1.4	Criterion 4: BIC, Bayesian Information Criterion.....	232
7.1.5	Comparison of AIC, AIC_c and BIC.....	232
7.2	Deciding on the Collection of Potential Subsets of Predictor Variables.....	233
7.2.1	All Possible Subsets.....	233
7.2.2	Stepwise Subsets.....	236
7.2.3	Inference After Variable Selection.....	238
7.3	Assessing the Predictive Ability of Regression Models.....	239
7.3.1	Stage 1: Model Building Using the Training Data Set.....	239
7.3.2	Stage 2: Model Comparison Using the Test Data Set.....	247
7.4	Recent Developments in Variable Selection – LASSO.....	250
7.5	Exercises.....	252
8	Logistic Regression.....	263
8.1	Logistic Regression Based on a Single Predictor.....	263
8.1.1	The Logistic Function and Odds.....	265
8.1.2	Likelihood for Logistic Regression with a Single Predictor.....	268
8.1.3	Explanation of Deviance.....	271
8.1.4	Using Differences in Deviance Values to Compare Models.....	272
8.1.5	R^2 for Logistic Regression.....	273
8.1.6	Residuals for Logistic Regression.....	274

- 8.2 Binary Logistic Regression..... 277
 - 8.2.1 Deviance for the Case of Binary Data..... 280
 - 8.2.2 Residuals for Binary Data 281
 - 8.2.3 Transforming Predictors in Logistic Regression
for Binary Data..... 282
 - 8.2.4 Marginal Model Plots for Binary Data..... 286
- 8.3 Exercises 294

- 9 Serially Correlated Errors 305**
 - 9.1 Autocorrelation 305
 - 9.2 Using Generalized Least Squares When the Errors Are AR(1)..... 310
 - 9.2.1 Generalized Least Squares Estimation 311
 - 9.2.2 Transforming a Model with AR(1) Errors into
a Model with iid Errors 315
 - 9.2.3 A General Approach to Transforming GLS into LS 316
 - 9.3 Case Study 319
 - 9.4 Exercises 325

- 10 Mixed Models 331**
 - 10.1 Random Effects..... 331
 - 10.1.1 Maximum Likelihood and Restricted
Maximum Likelihood..... 334
 - 10.1.2 Residuals in Mixed Models..... 345
 - 10.2 Models with Covariance Structures Which Vary Over Time..... 353
 - 10.2.1 Modeling the Conditional Mean 354
 - 10.3 Exercises 368

- Appendix: Nonparametric Smoothing 371**

- References 383**

- Index 387**