



TABLE OF CONTENTS

Preface xiii

Overview xix

Part I The Basics 1



1

The Analysis of Means: A Review of Basics and an Introduction to Linear Models 5

1.1 Introduction 5

1.2 Sampling Distributions 5

Sampling Distribution of the Mean 6

Sampling Distribution of the Variance 7

Sampling Distribution of the Ratio of Two Variances 8

Relationships among Distributions 9

1.3 Inferences on a Single Mean 9

Inferences Using the Sampling Distribution of the Mean 9

Inferences Using the Linear Model 11

Hypothesis Testing 12

1.4 Inferences for Two Means 17

Inferences Using the Sampling Distribution 17

Inferences Using the Linear Model 19

1.5 Inferences for Several Means, The Analysis of Variance 23

- 1.6 Summary** 28
- Exercises** 30

2**Simple Linear Regression: Linear Regression with One Independent Variable** 35

- 2.1 Introduction** 35
- 2.2 The Linear Regression Model** 37
- 2.3 Least Squares Estimation of Parameters β_0 and β_1** 40
 - Estimating the Parameters β_0 and β_1 40
 - Inferences on β_1 Using the Sampling Distribution 42
 - Inferences on β_1 Using the Linear Model 45
- 2.4 Inferences on the Response Variable** 49
- 2.5 Correlation and the Coefficient of Determination** 52
- 2.6 Regression through the Origin** 56
 - Regression through the Origin Using Sampling Distribution 56
 - Regression through the Origin Using the Linear Model 58
- 2.7 Assumptions on the Simple Linear Regression Model** 62
- 2.8 Uses and Misuses of Regression** 65
- 2.9 Inverse Prediction** 65
- 2.10 Summary** 67
- Exercises** 67

3**Multiple Regression** 75

- 3.1 Introduction** 75
- 3.2 The Multiple Linear Regression Model** 76
- 3.3 Estimation of Coefficients** 78

3.4	Interpreting the Partial Regression Coefficients	83
	Estimating Partial Coefficients with Residuals	85
3.5	Inferences on the Parameters	87
	Computing the Unrestricted SSE	90
	Computing the Restricted SSE	91
	The Hypothesis Test	91
	Commonly Used Tests	91
	The Test for the “Model”	92
	Tests for Individual Coefficients	93
	Simultaneous Inference	95
	The Test for a Coefficient Using Residuals	96
3.6	Testing a General Linear Hypothesis	99
3.7	Inferences on the Response Variable in Multiple Regression	102
3.8	Correlation and the Coefficient of Determination	104
	Multiple Correlation	104
	Partial Correlation	106
3.9	Summary and a Look Ahead	107
	Uses and Misuses of Regression Analysis	107
	Data Problems	109
	Model Problems	109
	Exercises	110

Part II Problems and Remedies 117

4

Problems with Observations 119

4.1	Introduction	119
4.2	Outliers and Influential Observations	120
	Statistics Based on Residuals	124
	Statistics Measuring Leverage	126
	Statistics Measuring Influence on Estimating the Response	127
	Using the DFBETAS Statistic	128
	Leverage Plots	129

- Statistics Measuring Influence on the Precision of Estimated Coefficients 130
- Comments 142
- Remedial Methods 143
- 4.3 Unequal Variances** 144
 - General Formulation 144
 - Weighted Regression by Estimating the Variances 150
 - Weights Based on Relationships 153
- 4.4 Robust Estimation** 157
- 4.5 Correlated Errors** 162
 - Autoregressive Models 162
 - Diagnostics for Autocorrelation 166
 - Remedial Methods 168
 - Alternative Estimation Technique 169
 - Model Modification 171
- 4.6 Summary** 175
- Exercises** 175

5**Multicollinearity** 181

- 5.1 Introduction** 181
- 5.2 The Effects of Multicollinearity** 182
- 5.3 Diagnosing Multicollinearity** 192
 - Variance Inflation Factors 192
 - Variance Proportions 195
 - Principal Components 195
- 5.4 Remedial Methods** 201
 - Redefining Variables 201
 - Methods Based on Knowledge of the Variables 202
 - Methods Based on Statistical Analysis 204
 - Principal Component Regression 207
 - Biased Estimation 216
 - Ridge Regression 218
 - Incomplete Principal Component Regression 219

- 5.5 Summary** 222
- Exercises** 223

6

Problems with the Model 229

- 6.1 Introduction** 229
- 6.2 Specification Error** 230
- 6.3 Lack of Fit Test** 234
- 6.4 Overspecification: Too Many Variables** 241
- 6.5 Variable Selection** 243
 - Size of Subset 244
 - The C_p Statistic 249
 - Other Selection Methods 250
- 6.6 Reliability of Variable Selection** 253
 - Cross Validation 254
 - Resampling 257
- 6.7 Usefulness of Variable Selection** 258
- 6.8 Variable Selection and Influential Observations** 263
 - Comments 266
- 6.9 Summary** 266
- Exercises** 267

Part III Additional Uses of Regression 271

7

Polynomial Models 273

- 7.1 Introduction** 273
- 7.2 Polynomial Models with One Independent Variable** 275

- 7.3 Segmented Polynomials with Known Knots** 283
 - Segmented Straight Lines 284
 - Segmented Polynomials 284
- 7.4 Polynomial Regression in Several Variables: Response Surfaces** 288
- 7.5 Summary** 298
 - Exercises** 298

8**Intrinsically Linear and Nonlinear Models** 303

- 8.1 Introduction** 303
- 8.2 Intrinsically Linear Models** 304
- 8.3 Intrinsically Nonlinear Models** 321
 - Growth Models 328
- 8.4 Summary** 334
 - Exercises** 335

9**The General Linear Model** 339

- 9.1 Introduction** 339
- 9.2 The Dummy Variables Model** 341
 - Mean and Variance of a Linear Function of Correlated Variables 347
- 9.3 Unequal Cell Frequencies** 348
- 9.4 Empty Cells** 354
- 9.5 Models with Dummy and Continuous Variables** 357
- 9.6 A Special Application: The Analysis of Covariance** 362
- 9.7 Heterogeneous Slopes in the Analysis of Covariance** 366
- 9.8 Summary** 371
 - Exercises** 371

Categorical Response Variable 375

- 10.1 Introduction** 375
- 10.2 Binary Response Variable** 376
 - The Linear Model with a Dichotomous Dependent Variable 377
- 10.3 Weighted Least Squares** 378
- 10.4 Simple Logistic Regression** 383
- 10.5 Multiple Logistic Regression** 390
- 10.6 Loglinear Model** 393
- 10.7 Summary** 401
- Exercises** 401

Appendix A: Statistical Tables 407

- A.1 The Standard Normal Distribution 408
- A.2 The t Distribution 413
- A.3 The χ^2 Distribution 414
- A.4 The F Distribution 416
- A.5 The Durbin-Watson Test Bounds 426

Appendix B: A Brief Introduction to Matrices 428

- B.1 Matrix Algebra 429
- B.2 Solving Linear Equations 432

Appendix C: Estimation Procedures 434

- C.1 Least Squares Estimation 434
- C.2 Maximum Likelihood Estimation 436

References 439