

Contents

ONE	Review of Simple Regression	1
	1.1. The Linear Model and Assumptions	2
	1.2. Least Squares Estimation	3
	1.3. Predicted Values and Residuals	5
	1.4. Analysis of Variation in the Dependent Variable	7
	1.5. Precision of Estimates	10
	1.6. Tests of Significance and Confidence Intervals	15
	1.7. Regression Through the Origin	19
	1.8. The Linear Model with Several Independent Variables	24
	1.9. Summary	25
	Exercises	25
TWO	Introduction to Matrices	31
	2.1. Basic Definitions	31
	2.2. Special Types of Matrices	32
	2.3. Matrix Operations	33
	2.4. Geometric Interpretation of Vectors	41
	2.5. Linear Equations and Solutions	44
	2.6. Orthogonal Transformations and Projections	48
	2.7. Eigenvalues and Eigenvectors	50
	2.8. Singular Value Decomposition of a Rectangular Matrix	54
	2.9. Summary	60
	Exercises	60

THREE	Multiple Regression in Matrix Notation	66
	3.1. The Model	66
	3.2. The Normal Equations and Their Solution	69
	3.3. The \hat{Y} Vector and the Residuals Vector	70
	3.4. Precision of the Estimates	72
	3.5. Distribution of Linear Functions of Normal Variables	78
	3.6. Summary of Matrix Formulas	79
	Exercises	79
FOUR	Analysis of Variance and Quadratic Forms	85
	4.1. Introduction to Quadratic Forms	85
	4.2. Analysis of Variance	91
	4.3. Expectations of Quadratic Forms	96
	4.4. Distribution of Quadratic Forms	99
	4.5. General Form for Hypothesis Testing	101
	4.5.1. The General Linear Hypothesis	101
	4.5.2. Some Special Cases of the General Form	103
	4.5.3. A Numerical Example	105
	4.5.4. Computing Q from Differences in Sums of Squares	108
	4.5.5. Sequential Sums of Squares	111
	4.5.6. The R -Notation to Label Sums of Squares	112
	4.5.7. Example: Sequential and Partial Sums of Squares	114
	4.6. Confidence Intervals and Joint Confidence Regions	116
	4.6.1. Univariate Confidence Intervals	116
	4.6.2. Simultaneous Confidence Statements	117
	4.6.3. Joint Confidence Regions	119
	4.7. Estimation of Pure Error	122
	Exercises	126
FIVE	Case Study: Five Independent Variables	134
	5.1. Spartina Biomass Production in the Cape Fear Estuary	134
	5.2. Regression Analysis for the Full Model	135
	5.2.1. The Correlation Matrix	137
	5.2.2. Multiple Regression Results: The Full Model	138
	5.3. Simplifying the Model	139
	5.4. Results of the Final Model	142

5.5.	General Comments	148	
	Exercises	150	
SIX	Geometric Interpretation of Least Squares		153
6.1.	The Linear Model and the Solution	153	
6.2.	Sums of Squares and Degrees of Freedom	157	
6.3.	Correction for the Mean and Sequential Regressions	161	
6.4.	The Collinearity Problem	162	
6.5.	Summary	165	
	Exercises	165	
SEVEN	Model Development: Selection of Variables		168
7.1.	Uses of the Regression Equation	169	
7.2.	Effects of Variable Selection on Least Squares	171	
7.3.	All Possible Regressions	172	
7.4.	Stepwise Regression Methods	175	
7.5.	Criteria for Choice of Subset Size	180	
	7.5.1. Coefficient of Determination	181	
	7.5.2. Residual Mean Square	182	
	7.5.3. Adjusted Coefficient of Determination	183	
	7.5.4. Mallows' C_p Statistic	183	
	7.5.5. "Significance Levels" for Choice of Subset Size	185	
7.6.	Model Validation	186	
	Exercises	189	
EIGHT	Class Variables in Regression		192
8.1.	Description of Class Variables	193	
8.2.	The Model for One-Way Structured Data	193	
8.3.	Reparameterizing to Remove Singularities	196	
	8.3.1. Reparameterizing with the Means Model	196	
	8.3.2. Reparameterizing Using the Constraint $\sum \tau_i = 0$	199	
	8.3.3. Reparameterizing Using the Constraint $\tau_j = 0$	201	
	8.3.4. Reparameterizing: A Numerical Example	202	
8.4.	Generalized Inverse Approach to Singular Models	204	
8.5.	The Model for Two-Way Classified Data	206	

8.6.	Class Variables to Test Homogeneity of Regressions	209
8.7.	Analysis of Covariance	213
8.8.	Numerical Examples	218
8.8.1.	Analysis of Variance	220
8.8.2.	Test of Homogeneity of Regression Coefficients	223
8.8.3.	Analysis of Covariance	224
	Exercises	231
NINE	Problem Areas in Least Squares	237
9.1.	Nonnormality	238
9.2.	Heterogeneous Variances	239
9.3.	Correlated Errors	240
9.4.	Influential Data Points and Outliers	241
9.5.	Model Inadequacies	243
9.6.	The Collinearity Problem	244
9.7.	Errors in the Independent Variables	245
9.8.	Summary	246
	Exercises	247
TEN	Regression Diagnostics	249
10.1.	Residuals Analysis	249
10.1.1.	Plots of e Versus \hat{Y}	253
10.1.2.	Plots of e Versus X_i	257
10.1.3.	Plots of e Versus Time	258
10.1.4.	Plots of e_i Versus e_{i-1}	260
10.1.5.	Normal Probability Plots	262
10.1.6.	Partial Regression Leverage Plots	265
10.2.	Influence Statistics	267
10.2.1.	Cook's D	268
10.2.2.	DFFITS	269
10.2.3.	DFBETAS	270
10.2.4.	COVRATIO	270
10.3.	Collinearity Diagnostics	273
10.3.1.	Condition Number and Condition Index	275
10.3.2.	Variance Inflation Factor	277
10.3.3.	Variance Decomposition Proportions	278
10.4.	Regression Diagnostics on the Linthurst Data	281
10.4.1.	Plots of Residuals	284

	10.4.2. Influence Statistics	290	
	10.4.3. Collinearity Diagnostics	292	
10.A.	Appendix: Uniform Identically Distributed Residuals		293
	Exercises	295	
ELEVEN	Transformation of Variables		299
	11.1. Reasons for Making Transformations	299	
	11.2. Transformations to Simplify Relationships	301	
	11.3. Transformations to Stabilize Variances	308	
	11.4. Transformations to Improve Normality	310	
	11.5. Generalized Least Squares	313	
	11.5.1. Weighted Least Squares	315	
	11.5.2. Generalized Least Squares	318	
	11.6. Summary	323	
	Exercises	323	
TWELVE	Collinearity		327
	12.1. Understanding the Structure of the X -Space	328	
	12.2. Biased Regression Methods	336	
	12.2.1. Ridge Regression	338	
	12.2.2. Principal Component Regression	344	
	12.3. General Comments	352	
	12.4. Summary	353	
	Exercises	354	
THIRTEEN	Case Study: Collinearity Problems		356
	13.1. The Problem	356	
	13.2. Multiple Regression: Ordinary Least Squares	360	
	13.3. Analysis of the Correlational Structure	362	
	13.4. Biased Regression	369	
	13.4.1. Ridge Regression	370	
	13.4.2. Principal Component Regression	371	
	13.5. Summary	374	
	Exercises	375	

FOURTEEN	Response Curve Modeling	376
	14.1. Considerations in Specifying the Functional Form	377
	14.2. Polynomial Response Surface Models	378
	14.3. Models Nonlinear in the Parameters	388
	14.4. Fitting Models Nonlinear in the Parameters	397
	Exercises	408
FIFTEEN	Case Study: Response Surface Modeling	415
	15.1. The Ozone–Sulfur Dioxide Response Surface (1981)	417
	15.1.1. Polynomial Response Model	419
	15.1.2. Nonlinear Weibull Response Model	423
	15.2. Analysis of the Combined Soybean Data	429
	Exercises	439
SIXTEEN	Analysis of Unbalanced Data	441
	16.1. Sources of Imbalance	442
	16.2. Effects of Imbalance	443
	16.3. Unweighted Analysis of Cell Means	445
	16.4. Weighted Analysis of Cell Means	448
	16.5. General Linear Models Analysis of Unbalanced Data	448
	16.5.1. Estimable Functions with Balanced Data	449
	16.5.2. Estimable Functions with Unbalanced Data	453
	16.5.3. Unbalanced Data: An Example	457
	16.5.4. Least Squares Means	460
	16.6. Models with More Than One Random Effect	464
	Exercises	470
SEVENTEEN	Case Study: Analysis of Unbalanced Data	474
	17.1. The Analysis of Variance	476
	17.2. Mean Square Expectations and Choice of Errors	486
	17.3. Least Squares Means and Standard Errors	489
	17.4. MIXMOD Analysis	493
	Exercises	495

Contents	xvii
Appendix Tables	497
Table A. Distribution of t	498
Table B. Percentage Points for the F -Distribution (Variance Ratio)	499
Table C. Significance Points of d_L and d_U for the Durbin–Watson Test for Correlation	504
Table D. Empirical Percentage Points of the Approximate W' test	507
Answers to Selected Exercises	509
Bibliography	535
Index	545