

DUXBURY APPLIED SERIES

INTRODUCTION TO REGRESSION MODELING



BOVAS ABRAHAM - JOHANNES LEDOLTER

DATA DISK ENCLOSED



Contents

CHAPTER 1

Introduction to Regression Models 1

- 1.1 Introduction 1
- 1.2 Examples 2
 - 1.2.1 Payout of an Investment 2
 - 1.2.2 Period of Oscillation of a Pendulum 3
 - 1.2.3 Salary of College Teachers 3
 - 1.2.4 Hardness Data 5
 - 1.2.5 Urea Formaldehyde Foam Insulation 7
 - 1.2.6 Oral Contraceptive Data 9
 - 1.2.7 Gas Consumption Data 11
- 1.3 A General Model 15
- 1.4 Important Reasons for Modeling 16
- 1.5 Data Plots and Empirical Modeling 17
- 1.6 An Iterative Model Building Approach 21
- 1.7 Some Comments on Data 22
- Exercises 23

CHAPTER 2

Simple Linear Regression 26

- 2.1 The Model 26
 - 2.1.1 Important Assumptions 26
 - 2.1.2 Objectives of the Analysis 27
- 2.2 Estimation of Parameters 27
 - 2.2.1 Maximum Likelihood Estimation 27
 - 2.2.2 Least Squares Estimation 28
- 2.3 Fitted Values, Residuals, and the Estimate of σ^2 29
 - 2.3.1 Consequences of the Least Squares Fit 30
 - 2.3.2 Estimation of σ^2 30
 - 2.3.3 Least Squares Calculations for the Hardness Example 31
- 2.4 Properties of Least Squares Estimates 31
 - 2.4.1 Expected Values of Least Squares Estimates 32

- 2.4.2 Variances of Least Squares Estimates 32
- 2.5 Inferences about the Regression Parameters 33
 - 2.5.1 Inference about β_1 33
 - 2.5.2 Inference about $\mu_0 = \beta_0 + \beta_1 x_0$ 36
 - 2.5.3 Hardness Example Continued 37
- 2.6 Prediction 38
 - 2.6.1 Hardness Example Continued 40
- 2.7 Analysis of Variance Approach to Regression 41
 - 2.7.1 Coefficient of Determination: R^2 44
 - 2.7.2 Hardness Example Continued 45
- 2.8 Another Example 46
- 2.9 Related Models 51
 - 2.9.1 Regression through the Origin 51
 - 2.9.2 The Case of Random x 's 52
- Appendix: Univariate Distributions 53
- Exercises 56

CHAPTER 3

A Review of Matrix Algebra and Important Results on Random Vectors 67

- 3.1 Review of Matrix Algebra 67
- 3.2 Matrix Approach to Simple Linear Regression 74
 - 3.2.1 Least Squares 74
 - 3.2.2 Hardness Example 76
- 3.3 Vectors of Random Variables 76
- 3.4 The Multivariate Normal Distribution 80

- 3.5 Important Results on Quadratic Forms 82
- Exercises 83

CHAPTER 4

Multiple Linear Regression Model 87

- 4.1 Introduction 87
 - 4.1.1 Two Examples 87
 - 4.1.2 The General Linear Model 89
- 4.2 Estimation of the Model 90
 - 4.2.1 A Geometric Interpretation of Least Squares 91
 - 4.2.2 Useful Properties of Estimates and Other Related Vectors 95
 - 4.2.3 Preliminary Discussion of Residuals 99
- 4.3 Statistical Inference 102
 - 4.3.1 Confidence Intervals and Tests of Hypotheses for a Single Parameter 102
 - 4.3.2 Prediction of a New Observation 108
- 4.4 The Additional Sum of Squares Principle 112
 - 4.4.1 Introduction 112
 - 4.4.2 Test of a Set of Linear Hypotheses 114
 - 4.4.3 Joint Confidence Regions for Several Parameters 120
- 4.5 The Analysis of Variance and the Coefficient of Determination, R^2 121
 - 4.5.1 Coefficient of Determination, R^2 124
- 4.6 Generalized Least Squares 125
 - 4.6.1 Introduction 125
 - 4.6.2 Generalized Least Squares Estimation 128

4.6.3 Weighted Least Squares 129

Appendix: Proofs of Results 130

Exercises 134

CHAPTER 5**Specification Issues in Regression Models 144****5.1 Elementary Special Cases 144****5.1.1 One-Sample Problem 144****5.1.2 Two-Sample Problem 145****5.1.3 Polynomial Models 146****5.2 Systems of Straight Lines 147****5.3 Comparison of Several
“Treatments” 151****5.4 X Matrices with Nearly
Linear-Dependent Columns 154****5.4.1 Detection of Multicollinearity
158****5.4.2 Guarding against
Multicollinearity 159****5.5 X Matrices with Orthogonal
Columns 160**

Exercises 163

CHAPTER 6**Model Checking 169****6.1 Introduction 169****6.2 Residual Analysis 170****6.2.1 Residuals and Residual
Plots 170****6.2.2 Added Variable Plots 173****6.2.3 Checking the Normality
Assumption 177****6.2.4 Serial Correlation among the
Errors 177****6.2.5 Example: Residual Plots for
the UFFI Data 182****6.3 The Effect of Individual Cases 182****6.3.1 Outliers 182****6.3.2 Leverage and Influence
Measures 186****6.3.3 Example: Forestry Data 192****6.4 Assessing the Adequacy of the
Functional Form: Testing
for Lack of Fit 199****6.4.1 Lack-of-Fit Test with More
Than One Independent
Variable 204****6.5 Variance-Stabilizing Transformations
205****6.5.1 Box–Cox Transformations
206****6.5.2 Several Useful Results for
Logarithmic Transformations
207**

Appendix 208

Exercises 210

CHAPTER 7**Model Selection 222****7.1 Introduction 222****7.2 All Possible Regressions 227****7.2.1 R^2 , R^2_{adj} , s^2 , and Akaike’s
Information Criterion 227****7.2.2 C_p Statistic 231****7.2.3 PRESS Statistic 233****7.3 Automatic Methods 234****7.3.1 Forward Selection 234****7.3.2 Backward Elimination 236****7.3.3 Stepwise Regression 237**

Exercises 238

CHAPTER 8**Case Studies in Linear
Regression 245****8.1 Educational Achievement of Iowa
Students 245****8.1.1 Analysis of Achievement
Scores 246**

- 8.1.2 Analysis of Teacher Salaries 249
- 8.1.3 Concluding Comments 252
- 8.2 Predicting the Price of Bordeaux Wine 253
- 8.3 Factors Influencing the Auction Price of Iowa Cows 258
- 8.4 Predicting U.S. Presidential Elections 266
 - 8.4.1 A Purely Economics-Based Model Proposed by Ray Fair 266
 - 8.4.2 Prediction Models Proposed by Political Scientists 274
 - 8.4.3 Concluding Comments 278
- 8.5 Student Projects Leading to Additional Case Studies 280
 - Exercises 283

CHAPTER 9

Nonlinear Regression Models 288

- 9.1 Introduction 288
- 9.2 Overview of Useful Deterministic Models, With Emphasis on Nonlinear Growth Curve Models 289
- 9.3 Nonlinear Regression Models 293
- 9.4 Inference in the Nonlinear Regression Model 294
 - 9.4.1 The Newton–Raphson Method of Determining the Minimum of a Function 294
 - 9.4.2 Application to Nonlinear Regression: Newton–Raphson and Gauss–Newton Methods 296
 - 9.4.3 Standard Errors of the Maximum Likelihood Estimates 297
 - 9.4.4 Implementation Issues 299
- 9.5 Examples 299
 - 9.5.1 Example 1: Loss in Chlorine Concentration 299

- 9.5.2 Example 2: Shoot Length of Plants Exposed to Gamma Irradiation 302

Exercises 305

CHAPTER 10

Regression Models for Time Series Situations 307

- 10.1 A Brief Introduction to Time Series Models 307
 - 10.1.1 First-Order Autoregressive Model 307
 - 10.1.2 Random Walk Model 310
 - 10.1.3 Second-Order Autoregressive Model 311
 - 10.1.4 Noisy Random Walk Model 311
 - 10.1.5 Summary of Time Series Models 312
 - 10.1.6 Remark 313
- 10.2 The Effects of Ignoring the Autocorrelation in the Errors 313
 - 10.2.1 Inefficiency of Least Squares Estimation 313
 - 10.2.2 Spurious Regression Results When Nonstationary Errors Are Involved 315
- 10.3 The Estimation of Combined Regression Time Series Models 316
 - 10.3.1 A Regression Model with First-Order Autoregressive Errors 316
 - 10.3.2 Regression Model with Noisy Random Walk Errors 319
- 10.4 Forecasting With Combined Regression Time Series Models 320
 - 10.4.1 Forecasts from the Regression Model with First-Order Autoregressive Errors 321

10.4.2	Forecasts from the Regression Model with Errors Following a Noisy Random Walk	322
10.4.3	Forecasts When the Explanatory Variable Must Be Forecast	323
10.5	Model-Building Strategy and Example	324
10.6	Cointegration and Regression with Time Series Data: An Example	327
	Exercises	334
CHAPTER 11		
Logistic Regression 343		
11.1	The Model	343
11.2	Interpretation of the Parameters	346
11.2.1	Relationship between Logistic Regression and the Analysis of 2×2 Contingency Tables	347
11.2.2	Another Advantage of Modeling Odds Ratios by Logistic Regression	349
11.3	Estimation of the Parameters	350
11.4	Inference	353
11.4.1	Likelihood Ratio Tests	353
11.4.2	Deviance	354
11.4.3	Standard Errors of the Maximum Likelihood Estimates	355
11.5	Model-Building Approach in the Context of Logistic Regression Models	356
11.5.1	Preliminary Model Specification	356
11.5.2	Assessing Model Fit	357
11.5.3	Model Diagnostics	359
11.6	Examples	360
11.6.1	Example 1: Death Penalty and Race of the Victim	360
11.6.2	Example 2: Hand Washing in Bathrooms at the University of Iowa	364

11.6.3	Example 3: Survival of the Donner Party	366
11.7	Overdispersion	372
11.8	Other Models for Binary Responses	373
11.9	Modeling Responses with More Than Two Categorical Outcomes	374
	Exercises	377

CHAPTER 12

Generalized Linear Models and Poisson Regression 382

12.1	The Model	382
12.2	Estimation of the Parameters in the Poisson Regression Model	383
12.3	Inference in the Poisson Regression Model	385
12.3.1	Likelihood Ratio Tests	385
12.3.2	Standard Errors of the Maximum Likelihood Estimates and Wald Tests	386
12.4	Overdispersion	387
12.5	Examples	389
12.5.1	Example 1: Mating Success of African Elephants	389
12.5.2	Example 2: Reproduction of Ceriodaphnia Organisms	395
	Exercises	399

ANSWERS TO SELECTED EXERCISES 403

STATISTICAL TABLES 414

REFERENCES 421

LIST OF DATA FILES 425

INDEX 427