Contents

Preface

1. Introduction
  1.1 Regression and Model Building, 1
  1.2 Data Collection, 5
  1.3 Uses of Regression, 9
  1.4 Role of the Computer, 10

2. Simple Linear Regression
  2.1 Simple Linear Regression Model, 12
  2.2 Least-Squares Estimation of the Parameters, 13
     2.2.1 Estimation of $\beta_0$ and $\beta_1$, 13
     2.2.2 Properties of the Least-Squares Estimators and
          the Fitted Regression Model, 17
     2.2.3 Estimation of $\sigma^2$, 20
     2.2.4 Alternate Form of the Model, 21
  2.3 Hypothesis Testing on the Slope and Intercept, 22
     2.3.1 Use of $t$ Tests, 22
     2.3.2 Testing Significance of Regression, 23
     2.3.3 Analysis of Variance, 25
  2.4 Interval Estimation in Simple Linear Regression, 28
     2.4.1 Confidence Intervals on $\beta_0$, $\beta_1$, and $\sigma^2$, 28
     2.4.2 Interval Estimation of the Mean Response, 30
  2.5 Prediction of New Observations, 33
  2.6 Coefficient of Determination, 35
  2.7 Using SAS for Simple Linear Regression, 36
  2.8 Some Considerations in the Use of Regression, 37
  2.9 Regression Through the Origin, 41
  2.10 Estimation by Maximum Likelihood, 47
2.11 Case Where the Regressor $x$ is Random, 49
   2.11.1 $x$ and $y$ Jointly Distributed, 49
   2.11.2 $x$ and $y$ Jointly Normally Distributed:
            Correlation Model, 49

Problems, 54

3. Multiple Linear Regression

3.1 Multiple Regression Models, 63
3.2 Estimation of the Model Parameters, 66
   3.2.1 Least-Squares Estimation of the Regression
        Coefficients, 66
   3.2.2 Geometrical Interpretation of Least Squares, 74
   3.2.3 Properties of the Least-Squares Estimators, 75
   3.2.4 Estimation of $\sigma^2$, 76
   3.2.5 Inadequacy of Scatter Diagrams in Multiple
        Regression, 77
   3.2.6 Maximum-Likelihood Estimation, 79
3.3 Hypothesis Testing in Multiple Linear Regression, 80
   3.3.1 Test for Significance of Regression, 80
   3.3.2 Tests on Individual Regression Coefficients, 84
   3.3.3 Special Case of Orthogonal Columns in $X$, 89
   3.3.4 Testing the General Linear Hypothesis, 90
3.4 Confidence Intervals in Multiple Regression, 93
   3.4.1 Confidence Intervals on the Regression
        Coefficients, 93
   3.4.2 Confidence Interval Estimation of the Mean
        Response, 94
   3.4.3 Simultaneous Confidence Intervals on Regression
        Coefficients, 96
3.5 Prediction of New Observations, 99
3.6 Using SAS for Basic Multiple Linear Regression, 101
3.7 Hidden Extrapolation in Multiple Regression, 101
3.8 Standardized Regression Coefficients, 105
3.9 Multicollinearity, 109
3.10 Why Do Regression Coefficients Have the Wrong
        Sign?, 112
Problems, 114

4. Model Adequacy Checking

4.1 Introduction, 122
4.2 Residual Analysis, 123
4.2.1 Definition of Residuals, 123
4.2.2 Methods for Scaling Residuals, 123
4.2.3 Residual Plots, 129
4.2.4 Partial Regression and Partial Residual Plots, 134
4.2.5 Using MINITAB and SAS for Residual Analysis, 137
4.2.6 Other Residual Plotting and Analysis Methods, 138

4.3 PRESS Statistic, 141

4.4 Detection and Treatment of Outliers, 142

4.5 Lack of Fit of the Regression Model, 145
   4.5.1 Formal Test for Lack of Fit, 145
   4.5.2 Estimation of Pure Error from Near Neighbors, 149

Problems, 153

5. **Transformations and Weighting to Correct Model Inadequacies**

   5.1 Introduction, 160
   5.2 Variance-Stabilizing Transformations, 161
   5.3 Transformations to Linearize the Model, 164
   5.4 Analytical Methods for Selecting a Transformation, 171
       5.4.1 Transformations on y: The Box–Cox Method, 171
       5.4.2 Transformations on the Regressor Variables, 174
   5.5 Generalized and Weighted Least Squares, 176
       5.5.1 Generalized Least Squares, 177
       5.5.2 Weighted Least Squares, 179
       5.5.3 Some Practical Issues, 180

Problems, 183

6. **Diagnostics for Leverage and Influence**

   6.1 Importance of Detecting Influential Observations, 189
   6.2 Leverage, 190
   6.3 Measures of Influence: Cook's $D$, 193
   6.4 Measures of Influence: $DFITS$ and $DFBETAS$, 195
   6.5 A Measure of Model Performance, 197
   6.6 Detecting Groups of Influential Observations, 198
   6.7 Treatment of Influential Observations, 199

Problems, 199

7. **Polynomial Regression Models**

   7.1 Introduction, 201
   7.2 Polynomial Models in One Variable, 201
       7.2.1 Basic Principles, 201
7.2.2 Piecewise Polynomial Fitting (Splines), 207
7.2.3 Polynomial and Trigonometric Terms, 213

7.3 Nonparametric Regression, 214
7.3.1 Kernel Regression, 214
7.3.2 Locally Weighted Regression (Loess), 215
7.3.3 Final Cautions, 219

7.4 Polynomial Models in Two or More Variables, 220
7.5 Orthogonal Polynomials, 226
Problems, 231

8. Indicator Variables

8.1 General Concept of Indicator Variables, 237
8.2 Comments on the Use of Indicator Variables, 249
8.2.1 Indicator Variables versus Regression on Allocated Codes, 249
8.2.2 Indicator Variables as a Substitute for a Quantitative Regressor, 250

8.3 Regression Approach to Analysis of Variance, 251
Problems, 256

9. Variable Selection and Model Building

9.1 Introduction, 261
9.1.1 Model-Building Problem, 261
9.1.2 Consequences of Model Misspecification, 262
9.1.3 Criteria for Evaluating Subset Regression Models, 265

9.2 Computational Techniques for Variable Selection, 270
9.2.1 All Possible Regressions, 270
9.2.2 Stepwise Regression Methods, 277

9.3 Strategy for Variable Selection and Model Building, 283
9.4 Case Study: Gorman and Toman Asphalt Data Using SAS, 286
Problems, 300

10. Validation of Regression Models

10.1 Introduction, 305
10.2 Validation Techniques, 306
10.2.1 Analysis of Model Coefficients and Predicted Values, 306
10.2.2 Collecting Fresh Data—Confirmation Runs, 308
10.2.3 Data Splitting, 310
10.3 Data from Planned Experiments, 318
Problems, 319

11. Multicollinearity

11.1 Introduction, 323
11.2 Sources of Multicollinearity, 323
11.3 Effects of Multicollinearity, 326
11.4 Multicollinearity Diagnostics, 331
   11.4.1 Examination of the Correlation Matrix, 333
   11.4.2 Variance Inflation Factors, 334
   11.4.3 Eigensystem Analysis of $X'X$, 335
   11.4.4 Other Diagnostics, 340
   11.4.5 SAS Code for Generating Multicollinearity Diagnostics, 341
11.5 Methods for Dealing with Multicollinearity, 341
   11.5.1 Collecting Additional Data, 341
   11.5.2 Model Respecification, 342
   11.5.3 Ridge Regression, 344
   11.5.4 Principal-Component Regression, 355
   11.5.5 Comparison and Evaluation of Biased Estimators, 360
11.6 Using SAS to Perform Ridge and Principal-Component Regression, 363
Problems, 365

12. Robust Regression

12.1 Need for Robust Regression, 369
12.2 $M$-Estimators, 372
12.3 Properties of Robust Estimators, 384
   12.3.1 Breakdown Point, 385
   12.3.2 Efficiency, 385
12.4 Survey of Other Robust Regression Estimators, 386
   12.4.1 High-Breakdown-Point Estimators, 386
   12.4.2 Bounded Influence Estimators, 389
   12.4.3 Other Procedures, 391
   12.4.4 Computing Robust Regression Estimators, 392
Problems, 393

13. Introduction to Nonlinear Regression

13.1 Linear and Nonlinear Regression Models, 397
1.1.1 Linear Regression Models, 397
1.1.2 Nonlinear Regression Models, 398
1.2 Origins of Nonlinear Models, 399
1.3 Nonlinear Least Squares, 403
1.4 Transformation to a Linear Model, 405
1.5 Parameter Estimation in a Nonlinear System, 408
   1.5.1 Linearization, 408
   1.5.2 Other Parameter Estimation Methods, 414
   1.5.3 Starting Values, 415
   1.5.4 Computer Programs, 416
1.6 Statistical Inference in Nonlinear Regression, 417
1.7 Examples of Nonlinear Regression Models, 419
1.8 Using SAS PROC NLIN, 420
Problems, 423

14. Generalized Linear Models

14.1 Introduction, 427
14.2 Logistic Regression Models, 428
   14.2.1 Models with a Binary Response Variable, 428
   14.2.2 Estimating the Parameters in a Logistic Regression
       Model, 430
   14.2.3 Interpretation of the Parameters in a Logistic
       Regression Model, 433
   14.2.4 Statistical Inference on Model Parameters, 436
   14.2.5 Diagnostic Checking in Logistic Regression, 444
   14.2.6 Other Models for Binary Response Data, 446
   14.2.7 More Than Two Categorical Outcomes, 447
14.3 Poisson Regression, 449
14.4 The Generalized Linear Model, 454
   14.4.1 Link Functions and Linear Predictors, 455
   14.4.2 Parameter Estimation and Inference in the
       GLM, 456
   14.4.3 Prediction and Estimation with the GLM, 460
   14.4.4 Residual Analysis in the GLM, 461
   14.4.5 Overdispersion, 464
Problems, 465

15. Other Topics in the Use of Regression Analysis

15.1 Regression Models with Autocorrelated Errors, 475
   15.1.1 Source and Effects of Autocorrelation, 475
   15.1.2 Detecting the Presence of Autocorrelation, 476
   15.1.3 Parameter Estimation Methods, 479
15.2 Effect of Measurement Errors in the Regressors, 486
15.2.1 Simple Linear Regression, 486
15.2.2 Berkson Model, 488
15.3 Inverse Estimation—The Calibration Problem, 488
15.4 Bootstrapping in Regression, 493
15.4.1 Bootstrap Sampling in Regression, 494
15.4.2 Bootstrap Confidence Intervals, 494
15.5 Classification and Regression Trees (CART), 500
15.6 Neural Networks, 502
15.7 Designed Experiments for Regression, 505
Problems, 507

APPENDIX A. Statistical Tables

APPENDIX B. Data Sets For Exercises

APPENDIX C. Supplemental Technical Material

C.1 Background on Basic Test Statistics, 546
C.2 Background from the Theory of Linear Models, 548
C.3 Important Results on $SS_R$ and $SS_{Res}$, 552
C.4 Gauss–Markov Theorem, $\text{Var}(\varepsilon) = \sigma^2 I$, 558
C.5 Computational Aspects of Multiple Regression, 560
C.6 Result on the Inverse of a Matrix, 562
C.7 Development of the PRESS Statistic, 562
C.8 Development of $S_{(i)}^2$, 564
C.9 Outlier Test Based on $R$-Student, 565
C.10 Independence of Residuals and Fitted Values, 568
C.11 The Gauss–Markov Theorem, $\text{Var}(\varepsilon) = \mathbf{V}$, 569
C.12 Bias in $MS_{Res}$ When the Model Is Underspecified, 571
C.13 Computation of Influence Diagnostics, 572
C.14 Generalized Linear Models, 573

APPENDIX D. Introduction to SAS

D.1 Basic Data Entry, 584
D.2 Creating Permanent SAS Data Sets, 589
D.3 Importing Data from an EXCEL File, 590
D.4 Output Command, 591
D.5 Log File, 591
D.6 Adding Variables to an Existing SAS Data Set, 593

References

Index