

Contents

1. Introduction	1
1.1 Regression and Model Building,	1
1.2 Uses of Regression,	5
1.3 Role of the Computer,	6
2. Simple Linear Regression and Correlation	7
2.1 Simple Linear Regression Model,	7
2.2 Least Squares Estimation of the Parameters,	8
2.2.1 Estimation of p_0 and β_1 ,	8
2.2.2 Properties of the Least Squares Estimators and the Fitted Regression Model,	13
2.2.3 Estimation of σ^2 ,	15
2.2.4 An Alternate Form of the Model,	17
2.3 Hypothesis Testing on the Slope and Intercept,	18
2.4 Interval Estimation in Simple Linear Regression,	24
2.4.1 Confidence Intervals on β_0 , β_1 , and r^2 ,	24
2.4.2 Interval Estimation of the Mean Response,	26
2.5 Prediction of New Observations,	30
2.6 Simultaneous Inference in Simple Linear Regression,	32
2.6.1 Simultaneous Inference on Model Parameters,	32
2.6.2 Simultaneous Estimation of Mean Response,	38
2.6.3 Prediction of m New Observations,	39
2.7 Coefficient of Determination,	40
2.8 Hazards in the Use of Regression,	42
2.9 Regression through the Origin,	45
2.10 Estimation by Maximum Likelihood,	51
2.11 Correlation,	52
2.12 Sample Computer Output, 57 Problems,	60

- 3.1 Introduction, 67
- 3.2 Residual Analysis, 68
 - 3.2.1 Definition of Residuals, 68
 - 3.2.2 Normal Probability Plot, 69
 - 3.2.3 Plot of Residuals Against \hat{y}_i , 74
 - 3.2.4 Plot of Residuals Against $^*_{i}$, 75
 - 3.2.5 Other Residual Plots, 77
 - 3.2.6 Statistical Tests on Residuals, 79
- 3.3 Detection and Treatment of Outliers, 80
- 3.4 A Test for Lack of Fit, 85
- 3.5 Transformations to a Straight Line, 89
- 3.6 Variance-Stabilizing Transformations, 98
- 3.7 Analytical Methods for Selecting a Transformation, 103
 - 3.7.1 Transformations on y , 103
 - 3.7.2 Transformations on x , 105
- 3.8 Weighted Least Squares, 108
 - Problems, 113

4. Multiple Linear Regression

- 4.1 Multiple Regression Models, 118
- 4.2 Estimation of the Model Parameters, 120
 - 4.2.1 Least Squares Estimation of the Regression Coefficients, 120
 - 4.2.2 A Geometrical Interpretation of Least Squares, 127
 - 4.2.3 Properties of the Least Squares Estimators, 128
 - 4.2.4 Estimation of a^{-2} , 129
 - 4.2.5 Inadequacy of Scatter Diagrams in Multiple Regression, 130
- 4.3 Confidence Intervals in Multiple Regression, 132
 - 4.3.1 Confidence Intervals on the Regression Coefficients, 132
 - 4.3.2 Confidence Interval Estimation of the Mean Response, 133
- 4.4 Hypothesis Testing in Multiple Linear Regression, 135
 - 4.4.1 Test for Significance of Regression, 135
 - 4.4.2 Tests on Individual Regression Coefficients, 138
 - 4.4.3 Special Case of Orthogonal columns in X , 143
 - 4.4.4 Testing the General Linear Hypothesis $T/3 = 0$, 145

- 4.5 Prediction of New Observations, 148
- 4.6 Hidden Extrapolation, 148
- 4.7 Simultaneous Inference in Multiple Regression, 152
 - 4.7.1 Simultaneous Confidence Region on Coefficients, 152
 - 4.7.2 Bonferroni, Scheffé, and Maximum Modulus t Intervals, 152
 - 4.7.3 Simultaneous Estimation of Mean Response, 153
 - 4.7.4 Prediction of m New Observations, 154
- 4.8 Standardized Regression Coefficients, 155
- 4.9 Regression Diagnostics and Measures of Model Adequacy, 159
 - 4.9.1 Coefficient of Multiple Determination, 159
 - 4.9.2 Residual Plots, 161
 - 4.9.3 Methods for Scaling Residuals, 169
 - 4.9.4 Estimation of Pure Error from Near-neighbors, 177
 - 4.9.5 Influence Diagnostics, 179
 - 4.9.6 Multicollinearity, 189
- 4.10 Sample Computer Output, 192
 - Problems, 195

5. Polynomial Regression Models

202

- 5.1 Introduction, 202
- 5.2 Polynomial Models in One Variable, 202
 - 5.2.1 Basic Principles, 202
 - 5.2.2 Piecewise Polynomial Fitting (Splines), 210
- 5.3 Polynomial Models in Two or More Variables, 218
- 5.4 Orthogonal Polynomials, 226
 - Problems, 231

6. Indicator Variables

237

- 6.1 The General Concept of Indicator Variables, 237
- 6.2 Comments on the Use of Indicator Variables, 252
 - 6.2.1 Indicator Variables versus Regression on Allocated Codes, 252
 - 6.2.2 Indicator Variables as a Substitute for a Quantitative Regressor, 253
 - 6.2.3 Models with only Indicator Variables, 253

- 6.3 Regression Models with an Indicator Response Variable, 254
 - 6.3.1 A Linear Model, 255
 - 6.3.2 A Nonlinear Model, 258
- Problems, 262

7. Variable Selection and Model Building

265

- 7.1 Introduction, 265
 - 7.1.1 Model-Building Problem, 265
 - 7.1.2 Consequences of Model Misspecification, 266
 - 7.1.3 Criteria for Evaluating Subset Regression Models, 270
- 7.2 Computational Techniques for Variable Selection, 275
 - 7.2.1 All Possible Regressions, 276
 - 7.2.2 Directed Search on t , 286
 - 7.2.3 Stepwise Regression Methods, 290
 - 7.2.4 Other Procedures, 298
- 7.3 Some Final Considerations, 300
 - Problems, 302

8. Multicollinearity

305

- 8.1 Introduction, 305
- 8.2 Sources of Multicollinearity, 305
- 8.3 Effects of Multicollinearity, 308
- 8.4 Multicollinearity Diagnostics, 314
 - 8.4.1 Examination of the Correlation Matrix, 314
 - 8.4.2 Variance Inflation Factors, 317
 - 8.4.3 Eigensystem Analysis of $X'X$, 318
 - 8.4.4 Other Diagnostics, 323
- 8.5 Methods for Dealing with Multicollinearity, 325
 - 8.5.1 Collecting Additional Data, 325
 - 8.5.2 Model Respecification, 326
 - 8.5.3 Ridge Regression, 329
 - 8.5.4 Generalized Ridge Regression, 344
 - 8.5.5 Principal Components Regression, 353
 - 8.5.6 Latent Root Regression Analysis, 357
 - 8.5.7 Comparison and Evaluation of Biased Estimators, 358
- Problems, 361

- 9.1 Autocorrelation, 366
 - 9.1.1 Source and Effects of Autocorrelation, 366
 - 9.1.2 Detecting the Presence of Autocorrelation, 367
 - 9.1.3 Parameter Estimation Methods, 372
- 9.2 Generalized and Weighted Least Squares, 378
- 9.3 Robust Regression, 382
 - 9.3.1 The Need for Robust Estimation, 382
 - 9.3.2 Af-Estimators, 384
 - 9.3.3 *R*- and *L*-Estimation, 397
 - 9.3.4 Robust Ridge Regression, 398
- 9.4 Why Do Regression Coefficients Have the "Wrong" Sign?, 399
- 9.5 Effect of Measurement Errors in the *X*'s, 402
- 9.6 Inverse Estimation (Calibration or Discrimination), 404
- 9.7 Designed Experiments for Regression, 409
- 9.8 Relationship between Regression and Analysis of Variance, 411
- 9.9 Generalized Linear Models, 418
- 9.10 An Introduction to Nonlinear Regression, 421
 - 9.10.1 Nonlinear Regression Model, 421
 - 9.10.2 Nonlinear Least Squares, 422
 - 9.10.3 Transforming to a Linear Model, 423
 - 9.10.4 Parameter Estimation in a Nonlinear System, 426
 - 9.10.5 Starting Values, 431
 - 9.10.6 Examples of Nonlinear Regression Models, 432

10. Validation of Regression Models

- 10.1 Introduction, 443
- 10.2 Validation Techniques, 444
 - 10.2.1 Analysis of Model Coefficients and Predicted Values, 445
 - 10.2.2 Collecting Fresh Data, 446
 - 10.2.3 Data Splitting, 448
- 10.3 Data from Planned Experiments, 458
 - Problems, 459

Appendix C	Supplemental Technical Material	498
	C.1 Computational Aspects of Multiple Regression, 498	
	C.2 A Result on the Inverse of a Matrix, 500	
	C.3 Development of the PRESS statistic, 501	
	C.4 Development of Sf_{iy} , 503	
	C.5 Computation of Influence Diagnostics, 504	
References		507
Index		523