

# Contents

## PART ONE

### SIMPLE LINEAR REGRESSION 1

#### Chapter 1

#### Linear Regression with One Predictor Variable 2

- 1.1 Relations between Variables 2
  - Functional Relation between Two Variables* 2
  - Statistical Relation between Two Variables* 3
- 1.2 Regression Models and Their Uses 5
  - Historical Origins* 5
  - Basic Concepts* 5
  - Construction of Regression Models* 7
  - Uses of Regression Analysis* 8
  - Regression and Causality* 8
  - Use of Computers* 9
- 1.3 Simple Linear Regression Model with Distribution of Error Terms Unspecified 9
  - Formal Statement of Model* 9
  - Important Features of Model* 9
  - Meaning of Regression Parameters* 11
  - Alternative Versions of Regression Model* 12
- 1.4 Data for Regression Analysis 12
  - Observational Data* 12
  - Experimental Data* 13
  - Completely Randomized Design* 13
- 1.5 Overview of Steps in Regression Analysis 13
- 1.6 Estimation of Regression Function 15
  - Method of Least Squares* 15
  - Point Estimation of Mean Response* 21
  - Residuals* 22
  - Properties of Fitted Regression Line* 23
- 1.7 Estimation of Error Terms Variance  $\sigma^2$  24
  - Point Estimator of  $\sigma^2$*  24
- 1.8 Normal Error Regression Model 26
  - Model* 26
  - Estimation of Parameters by Method of Maximum Likelihood* 27

- Cited References 33
- Problems 33
- Exercises 37
- Projects 38

#### Chapter 2

#### Inferences in Regression and Correlation Analysis 40

- 2.1 Inferences Concerning  $\beta_1$  40
  - Sampling Distribution of  $b_1$*  41
  - Sampling Distribution of  $(b_1 - \beta_1)/s\{b_1\}$*  44
  - Confidence Interval for  $\beta_1$*  45
  - Tests Concerning  $\beta_1$*  47
- 2.2 Inferences Concerning  $\beta_0$  48
  - Sampling Distribution of  $b_0$*  48
  - Sampling Distribution of  $(b_0 - \beta_0)/s\{b_0\}$*  49
  - Confidence Interval for  $\beta_0$*  49
- 2.3 Some Considerations on Making Inferences Concerning  $\beta_0$  and  $\beta_1$  50
  - Effects of Departures from Normality* 50
  - Interpretation of Confidence Coefficient and Risks of Errors* 50
  - Spacing of the X Levels* 50
  - Power of Tests* 50
- 2.4 Interval Estimation of  $E\{Y_h\}$  52
  - Sampling Distribution of  $\hat{Y}_h$*  52
  - Sampling Distribution of  $(\hat{Y}_h - E\{Y_h\})/s\{\hat{Y}_h\}$*  54
  - Confidence Interval for  $E\{Y_h\}$*  54
- 2.5 Prediction of New Observation 55
  - Prediction Interval for  $Y_{h(\text{new})}$  when Parameters Known* 56
  - Prediction Interval for  $Y_{h(\text{new})}$  when Parameters Unknown* 57
  - Prediction of Mean of m New Observations for Given  $X_h$*  60
- 2.6 Confidence Band for Regression Line 61
- 2.7 Analysis of Variance Approach to Regression Analysis 63
  - Partitioning of Total Sum of Squares* 63
  - Breakdown of Degrees of Freedom* 66

- Mean Squares* 66
  - Analysis of Variance Table* 67
  - Expected Mean Squares* 68
  - F Test of  $\beta_1 = 0$  versus  $\beta_1 \neq 0$*  69
  - 2.8 General Linear Test Approach** 72
    - Full Model* 72
    - Reduced Model* 72
    - Test Statistic* 73
    - Summary* 73
  - 2.9 Descriptive Measures of Linear Association between  $X$  and  $Y$**  74
    - Coefficient of Determination* 74
    - Limitations of  $R^2$*  75
    - Coefficient of Correlation* 76
  - 2.10 Considerations in Applying Regression Analysis** 77
  - 2.11 Normal Correlation Models** 78
    - Distinction between Regression and Correlation Model* 78
    - Bivariate Normal Distribution* 78
    - Conditional Inferences* 80
    - Inferences on Correlation Coefficients* 83
    - Spearman Rank Correlation Coefficient* 87
    - Cited References* 89
    - Problems* 89
    - Exercises* 97
    - Projects* 98
- Chapter 3**
- Diagnostics and Remedial Measures 100**
- 3.1 Diagnostics for Predictor Variable** 100
  - 3.2 Residuals** 102
    - Properties of Residuals* 102
    - Semistudentized Residuals* 103
    - Departures from Model to Be Studied by Residuals* 103
  - 3.3 Diagnostics for Residuals** 103
    - Nonlinearity of Regression Function* 104
    - Nonconstancy of Error Variance* 107
    - Presence of Outliers* 108
    - Nonindependence of Error Terms* 108
    - Nonnormality of Error Terms* 110
    - Omission of Important Predictor Variables* 112
    - Some Final Comments* 114
  - 3.4 Overview of Tests Involving Residuals** 114
    - Tests for Randomness* 114
    - Tests for Constancy of Variance* 115
    - Tests for Outliers* 115
    - Tests for Normality* 115
  - 3.5 Correlation Test for Normality** 115
  - 3.6 Tests for Constancy of Error Variance** 116
    - Brown-Forsythe Test* 116
    - Breusch-Pagan Test* 118
  - 3.7 F Test for Lack of Fit** 119
    - Assumptions* 119
    - Notation* 121
    - Full Model* 121
    - Reduced Model* 123
    - Test Statistic* 123
    - ANOVA Table* 124
  - 3.8 Overview of Remedial Measures** 127
    - Nonlinearity of Regression Function* 128
    - Nonconstancy of Error Variance* 128
    - Nonindependence of Error Terms* 128
    - Nonnormality of Error Terms* 128
    - Omission of Important Predictor Variables* 129
    - Outlying Observations* 129
  - 3.9 Transformations** 129
    - Transformations for Nonlinear Relation Only* 129
    - Transformations for Nonnormality and Unequal Error Variances* 132
    - Box-Cox Transformations* 134
  - 3.10 Exploration of Shape of Regression Function** 137
    - Lowess Method* 138
    - Use of Smoothed Curves to Confirm Fitted Regression Function* 139
  - 3.11 Case Example—Plutonium Measurement** 141
    - Cited References* 146
    - Problems* 146
    - Exercises* 151
    - Projects* 152
    - Case Studies* 153

## Chapter 4

### Simultaneous Inferences and Other Topics in Regression Analysis 154

- 4.1 Joint Estimation of  $\beta_0$  and  $\beta_1$  154  
*Need for Joint Estimation* 154  
*Bonferroni Joint Confidence Intervals* 155
- 4.2 Simultaneous Estimation of Mean Responses 157  
*Working-Hotelling Procedure* 158  
*Bonferroni Procedure* 159
- 4.3 Simultaneous Prediction Intervals for New Observations 160
- 4.4 Regression through Origin 161  
*Model* 161  
*Inferences* 161  
*Important Cautions for Using Regression through Origin* 164
- 4.5 Effects of Measurement Errors 165  
*Measurement Errors in Y* 165  
*Measurement Errors in X* 165  
*Berkson Model* 167
- 4.6 Inverse Predictions 168
- 4.7 Choice of X Levels 170  
 Cited References 172  
 Problems 172  
 Exercises 175  
 Projects 175

## Chapter 5

### Matrix Approach to Simple Linear Regression Analysis 176

- 5.1 Matrices 176  
*Definition of Matrix* 176  
*Square Matrix* 178  
*Vector* 178  
*Transpose* 178  
*Equality of Matrices* 179
- 5.2 Matrix Addition and Subtraction 180
- 5.3 Matrix Multiplication 182  
*Multiplication of a Matrix by a Scalar* 182  
*Multiplication of a Matrix by a Matrix* 182
- 5.4 Special Types of Matrices 185  
*Symmetric Matrix* 185  
*Diagonal Matrix* 185

- Vector and Matrix with All Elements Unity* 187  
*Zero Vector* 187
- 5.5 Linear Dependence and Rank of Matrix 188  
*Linear Dependence* 188  
*Rank of Matrix* 188
- 5.6 Inverse of a Matrix 189  
*Finding the Inverse* 190  
*Uses of Inverse Matrix* 192
- 5.7 Some Basic Results for Matrices 193
- 5.8 Random Vectors and Matrices 193  
*Expectation of Random Vector or Matrix* 193  
*Variance-Covariance Matrix of Random Vector* 194  
*Some Basic Results* 196  
*Multivariate Normal Distribution* 196
- 5.9 Simple Linear Regression Model in Matrix Terms 197
- 5.10 Least Squares Estimation of Regression Parameters 199  
*Normal Equations* 199  
*Estimated Regression Coefficients* 200
- 5.11 Fitted Values and Residuals 202  
*Fitted Values* 202  
*Residuals* 203
- 5.12 Analysis of Variance Results 204  
*Sums of Squares* 204  
*Sums of Squares as Quadratic Forms* 205
- 5.13 Inferences in Regression Analysis 206  
*Regression Coefficients* 207  
*Mean Response* 208  
*Prediction of New Observation* 209  
 Cited Reference 209  
 Problems 209  
 Exercises 212

## PART TWO

### MULTIPLE LINEAR REGRESSION 213

## Chapter 6

### Multiple Regression I 214

- 6.1 Multiple Regression Models 214

- Need for Several Predictor Variables* 214
    - First-Order Model with Two Predictor Variables* 215
    - First-Order Model with More than Two Predictor Variables* 217
    - General Linear Regression Model* 217
  - 6.2** **General Linear Regression Model in Matrix Terms** 222
  - 6.3** **Estimation of Regression Coefficients** 223
  - 6.4** **Fitted Values and Residuals** 224
  - 6.5** **Analysis of Variance Results** 225
    - Sums of Squares and Mean Squares* 225
    - F Test for Regression Relation* 226
    - Coefficient of Multiple Determination* 226
    - Coefficient of Multiple Correlation* 227
  - 6.6** **Inferences about Regression Parameters** 227
    - Interval Estimation of  $\beta_k$*  228
    - Tests for  $\beta_k$*  228
    - Joint Inferences* 228
  - 6.7** **Estimation of Mean Response and Prediction of New Observation** 229
    - Interval Estimation of  $E\{Y_h\}$*  229
    - Confidence Region for Regression Surface* 229
    - Simultaneous Confidence Intervals for Several Mean Responses* 230
    - Prediction of New Observation  $Y_{h(\text{new})}$*  230
    - Prediction of Mean of  $m$  New Observations at  $X_h$*  230
    - Predictions of  $g$  New Observations* 231
    - Caution about Hidden Extrapolations* 231
  - 6.8** **Diagnostics and Remedial Measures** 232
    - Scatter Plot Matrix* 232
    - Three-Dimensional Scatter Plots* 233
    - Residual Plots* 233
    - Correlation Test for Normality* 234
    - Brown-Forsythe Test for Constancy of Error Variance* 234
    - Breusch-Pagan Test for Constancy of Error Variance* 234
    - F Test for Lack of Fit* 235
    - Remedial Measures* 236
  - 6.9** **An Example—Multiple Regression with Two Predictor Variables** 236
    - Setting* 236
    - Basic Calculations* 237
    - Estimated Regression Function* 240
    - Fitted Values and Residuals* 241
    - Analysis of Appropriateness of Model* 241
    - Analysis of Variance* 243
    - Estimation of Regression Parameters* 245
    - Estimation of Mean Response* 245
    - Prediction Limits for New Observations* 247
  - Cited Reference 248
  - Problems 248
  - Exercises 253
  - Projects 254
- Chapter 7**
- Multiple Regression II 256**
- 7.1** **Extra Sums of Squares** 256
    - Basic Ideas* 256
    - Definitions* 259
    - Decomposition of SSR into Extra Sums of Squares* 260
    - ANOVA Table Containing Decomposition of SSR* 261
  - 7.2** **Uses of Extra Sums of Squares in Tests for Regression Coefficients** 263
    - Test whether a Single  $\beta_k = 0$*  263
    - Test whether Several  $\beta_k = 0$*  264
  - 7.3** **Summary of Tests Concerning Regression Coefficients** 266
    - Test whether All  $\beta_k = 0$*  266
    - Test whether a Single  $\beta_k = 0$*  267
    - Test whether Some  $\beta_k = 0$*  267
    - Other Tests* 268
  - 7.4** **Coefficients of Partial Determination** 268
    - Two Predictor Variables* 269
    - General Case* 269
    - Coefficients of Partial Correlation* 270
  - 7.5** **Standardized Multiple Regression Model** 271
    - Roundoff Errors in Normal Equations Calculations* 271
    - Lack of Comparability in Regression Coefficients* 272
    - Correlation Transformation* 272
    - Standardized Regression Model* 273
    - X'X Matrix for Transformed Variables* 274

*Estimated Standardized Regression**Coefficients* 275**7.6 Multicollinearity and Its Effects** 278*Uncorrelated Predictor Variables* 279*Nature of Problem when Predictor Variables**Are Perfectly Correlated* 281*Effects of Multicollinearity* 283*Need for More Powerful Diagnostics for*  
*Multicollinearity* 289

Cited Reference 289

Problems 289

Exercise 292

Projects 293

**Chapter 8**  
**Regression Models for Quantitative**  
**and Qualitative Predictors** 294**8.1 Polynomial Regression Models** 294*Uses of Polynomial Models* 294*One Predictor Variable—Second Order* 295*One Predictor Variable—Third Order* 296*One Predictor Variable—Higher Orders* 296*Two Predictor Variables—Second Order* 297*Three Predictor Variables—Second*  
*Order* 298*Implementation of Polynomial Regression*  
*Models* 298*Case Example* 300*Some Further Comments on Polynomial*  
*Regression* 305**8.2 Interaction Regression Models** 306*Interaction Effects* 306*Interpretation of Interaction Regression**Models with Linear Effects* 306*Interpretation of Interaction Regression**Models with Curvilinear Effects* 309*Implementation of Interaction Regression*  
*Models* 311**8.3 Qualitative Predictors** 313*Qualitative Predictor with Two**Classes* 314*Interpretation of Regression Coefficients* 315*Qualitative Predictor with More than Two**Classes* 318*Time Series Applications* 319**8.4 Some Considerations in Using Indicator**  
**Variables** 321*Indicator Variables versus Allocated*  
*Codes* 321*Indicator Variables versus Quantitative*  
*Variables* 322*Other Codings for Indicator Variables* 323**8.5 Modeling Interactions between Quantitative**  
**and Qualitative Predictors** 324*Meaning of Regression Coefficients* 324**8.6 More Complex Models** 327*More than One Qualitative Predictor*  
*Variable* 328*Qualitative Predictor Variables Only* 329**8.7 Comparison of Two or More Regression**  
**Functions** 329*Soap Production Lines Example* 330*Instrument Calibration Study Example* 334

Cited Reference 335

Problems 335

Exercises 340

Projects 341

Case Study 342

**Chapter 9**  
**Building the Regression Model I:**  
**Model Selection and Validation** 343**9.1 Overview of Model-Building Process** 343*Data Collection* 343*Data Preparation* 346*Preliminary Model Investigation* 346*Reduction of Explanatory Variables* 347*Model Refinement and Selection* 349*Model Validation* 350**9.2 Surgical Unit Example** 350**9.3 Criteria for Model Selection** 353 $R_p^2$  or  $SSE_p$  Criterion 354 $R_{a,p}^2$  or  $MSE_p$  Criterion 355Mallows'  $C_p$  Criterion 357 $AIC_p$  and  $SBC_p$  Criteria 359PRESS $_p$  Criterion 360**9.4 Automatic Search Procedures for Model**  
**Selection** 361*"Best" Subsets Algorithm* 361*Stepwise Regression Methods* 364

- Forward Stepwise Regression* 364
- Other Stepwise Procedures* 367
- 9.5** Some Final Comments on Automatic Model Selection Procedures 368
- 9.6** Model Validation 369
  - Collection of New Data to Check Model* 370
  - Comparison with Theory, Empirical Evidence, or Simulation Results* 371
  - Data Splitting* 372
  - Cited References 375
  - Problems 376
  - Exercise 380
  - Projects 381
  - Case Studies 382

## Chapter 10

### Building the Regression Model II: Diagnostics 384

- 10.1** Model Adequacy for a Predictor Variable—Added-Variable Plots 384
- 10.2** Identifying Outlying  $Y$  Observations—Studentized Deleted Residuals 390
  - Outlying Cases* 390
  - Residuals and Semistudentized Residuals* 392
  - Hat Matrix* 392
  - Studentized Residuals* 394
  - Deleted Residuals* 395
  - Studentized Deleted Residuals* 396
- 10.3** Identifying Outlying  $X$  Observations—Hat Matrix Leverage Values 398
  - Use of Hat Matrix for Identifying Outlying  $X$  Observations* 398
  - Use of Hat Matrix to Identify Hidden Extrapolation* 400
- 10.4** Identifying Influential Cases—*DFFITs*, Cook's Distance, and *DFBETAS* Measures 400
  - Influence on Single Fitted Value—DFFITs* 401
  - Influence on All Fitted Values—Cook's Distance* 402
  - Influence on the Regression Coefficients—DFBETAS* 404

- Influence on Inferences* 405
- Some Final Comments* 406
- 10.5** Multicollinearity Diagnostics—Variance Inflation Factor 406
  - Informal Diagnostics* 407
  - Variance Inflation Factor* 408
- 10.6** Surgical Unit Example—Continued 410
  - Cited References 414
  - Problems 414
  - Exercises 419
  - Projects 419
  - Case Studies 420

## Chapter 11

### Building the Regression Model III: Remedial Measures 421

- 11.1** Unequal Error Variances Remedial Measures—Weighted Least Squares 421
  - Error Variances Known* 422
  - Error Variances Known up to Proportionality Constant* 424
  - Error Variances Unknown* 424
- 11.2** Multicollinearity Remedial Measures—Ridge Regression 431
  - Some Remedial Measures* 431
  - Ridge Regression* 432
- 11.3** Remedial Measures for Influential Cases—Robust Regression 437
  - Robust Regression* 438
  - IRLS Robust Regression* 439
- 11.4** Nonparametric Regression: Lowess Method and Regression Trees 449
  - Lowess Method* 449
  - Regression Trees* 453
- 11.5** Remedial Measures for Evaluating Precision in Nonstandard Situations—Bootstrapping 458
  - General Procedure* 459
  - Bootstrap Sampling* 459
  - Bootstrap Confidence Intervals* 460
- 11.6** Case Example—MNDOT Traffic Estimation 464
  - The AADT Database* 464
  - Model Development* 465
  - Weighted Least Squares Estimation* 468

- Cited References 471
- Problems 472
- Exercises 476
- Projects 476
- Case Studies 480

## Chapter 12

### Autocorrelation in Time Series Data 481

- 12.1 Problems of Autocorrelation 481
- 12.2 First-Order Autoregressive Error Model 484
  - Simple Linear Regression* 484
  - Multiple Regression* 484
  - Properties of Error Terms* 485
- 12.3 Durbin-Watson Test for Autocorrelation 487
- 12.4 Remedial Measures for Autocorrelation 490
  - Addition of Predictor Variables* 490
  - Use of Transformed Variables* 490
  - Cochrane-Orcutt Procedure* 492
  - Hildreth-Lu Procedure* 495
  - First Differences Procedure* 496
  - Comparison of Three Methods* 498
- 12.5 Forecasting with Autocorrelated Error Terms 499
  - Cited References 502
  - Problems 502
  - Exercises 507
  - Projects 508
  - Case Studies 508

## PART THREE

### NONLINEAR REGRESSION 509

## Chapter 13

### Introduction to Nonlinear Regression and Neural Networks 510

- 13.1 Linear and Nonlinear Regression Models 510
  - Linear Regression Models* 510
  - Nonlinear Regression Models* 511
  - Estimation of Regression Parameters* 514

### 13.2 Least Squares Estimation in Nonlinear Regression 515

- Solution of Normal Equations* 517
- Direct Numerical Search—Gauss-Newton Method* 518
- Other Direct Search Procedures* 525
- 13.3 Model Building and Diagnostics 526
- 13.4 Inferences about Nonlinear Regression Parameters 527
  - Estimate of Error Term Variance* 527
  - Large-Sample Theory* 528
  - When Is Large-Sample Theory Applicable?* 528
  - Interval Estimation of a Single  $\gamma_k$*  531
  - Simultaneous Interval Estimation of Several  $\gamma_k$*  532
  - Test Concerning a Single  $\gamma_k$*  532
  - Test Concerning Several  $\gamma_k$*  533

### 13.5 Learning Curve Example 533

### 13.6 Introduction to Neural Network Modeling 537

- Neural Network Model* 537
- Network Representation* 540
- Neural Network as Generalization of Linear Regression* 541
- Parameter Estimation: Penalized Least Squares* 542
- Example: Ischemic Heart Disease* 543
- Model Interpretation and Prediction* 546
- Some Final Comments on Neural Network Modeling* 547
- Cited References 547
- Problems 548
- Exercises 552
- Projects 552
- Case Studies 554

## Chapter 14

### Logistic Regression, Poisson Regression, and Generalized Linear Models 555

- 14.1 Regression Models with Binary Response Variable 555
  - Meaning of Response Function when Outcome Variable Is Binary* 556

- Special Problems when Response Variable Is Binary* 557
- 14.2** Sigmoidal Response Functions for Binary Responses 559
- Probit Mean Response Function* 559
  - Logistic Mean Response Function* 560
  - Complementary Log-Log Response Function* 562
- 14.3** Simple Logistic Regression 563
- Simple Logistic Regression Model* 563
  - Likelihood Function* 564
  - Maximum Likelihood Estimation* 564
  - Interpretation of  $b_1$*  567
  - Use of Probit and Complementary Log-Log Response Functions* 568
  - Repeat Observations—Binomial Outcomes* 568
- 14.4** Multiple Logistic Regression 570
- Multiple Logistic Regression Model* 570
  - Fitting of Model* 571
  - Polynomial Logistic Regression* 575
- 14.5** Inferences about Regression Parameters 577
- Test Concerning a Single  $\beta_k$ : Wald Test* 578
  - Interval Estimation of a Single  $\beta_k$*  579
  - Test whether Several  $\beta_k = 0$ : Likelihood Ratio Test* 580
- 14.6** Automatic Model Selection Methods 582
- Model Selection Criteria* 582
  - Best Subsets Procedures* 583
  - Stepwise Model Selection* 583
- 14.7** Tests for Goodness of Fit 586
- Pearson Chi-Square Goodness of Fit Test* 586
  - Deviance Goodness of Fit Test* 588
  - Hosmer-Lemeshow Goodness of Fit Test* 589
- 14.8** Logistic Regression Diagnostics 591
- Logistic Regression Residuals* 591
  - Diagnostic Residual Plots* 594
  - Detection of Influential Observations* 598
- 14.9** Inferences about Mean Response
- Point Estimator* 602
  - Interval Estimation* 602
  - Simultaneous Confidence Intervals for Several Mean Responses* 603
- 14.10** Prediction of a New Observation 604
- Choice of Prediction Rule* 604
  - Validation of Prediction Error Rate* 607
- 14.11** Polytomous Logistic Regression for Nominal Response 608
- Pregnancy Duration Data with Polytomous Response* 609
  - $J - 1$  Baseline-Category Logits for Nominal Response* 610
  - Maximum Likelihood Estimation* 612
- 14.12** Polytomous Logistic Regression for Ordinal Response 614
- 14.13** Poisson Regression 618
- Poisson Distribution* 618
  - Poisson Regression Model* 619
  - Maximum Likelihood Estimation* 620
  - Model Development* 620
  - Inferences* 621
- 14.14** Generalized Linear Models 623
- Cited References* 624
  - Problems* 625
  - Exercises* 634
  - Projects* 635
  - Case Studies* 640
- PART FOUR**
- DESIGN AND ANALYSIS OF SINGLE-FACTOR STUDIES 641**
- Chapter 15**
- Introduction to the Design of Experimental and Observational Studies 642**
- 15.1** Experimental Studies, Observational Studies, and Causation 643
- Experimental Studies* 643
  - Observational Studies* 644
  - Mixed Experimental and Observational Studies* 646
- 15.2** Experimental Studies: Basic Concepts 647



- Factors 647
  - Crossed and Nested Factors 648
  - Treatments 649
  - Choice of Treatments 649
  - Experimental Units 652
  - Sample Size and Replication 652
  - Randomization 653
  - Constrained Randomization:
    - Blocking 655
    - Measurements 658
  - 15.3 An Overview of Standard Experimental Designs 658**
    - Completely Randomized Design 659
    - Factorial Experiments 660
    - Randomized Complete Block Designs 661
    - Nested Designs 662
    - Repeated Measures Designs 663
    - Incomplete Block Designs 664
    - Two-Level Factorial and Fractional Factorial Experiments 665
    - Response Surface Experiments 666
  - 15.4 Design of Observational Studies 666**
    - Cross-Sectional Studies 666
    - Prospective Studies 667
    - Retrospective Studies 667
    - Matching 668
  - 15.5 Case Study: Paired-Comparison Experiment 669**
  - 15.6 Concluding Remarks 672**
    - Cited References 672
    - Problems 672
    - Exercise 676
- Chapter 16**  
**Single-Factor Studies 677**
- 16.1 Single-Factor Experimental and Observational Studies 677**
  - 16.2 Relation between Regression and Analysis of Variance 679**
    - Illustrations 679
    - Choice between Two Types of Models 680
  - 16.3 Single-Factor ANOVA Model 681**
    - Basic Ideas 681
    - Cell Means Model 681
    - Important Features of Model 682
    - The ANOVA Model Is a Linear Model 683
    - Interpretation of Factor Level Means 684
    - Distinction between ANOVA Models I and II 685
  - 16.4 Fitting of ANOVA Model 685**
    - Notation 686
    - Least Squares and Maximum Likelihood Estimators 687
    - Residuals 689
  - 16.5 Analysis of Variance 690**
    - Partitioning of SSTO 690
    - Breakdown of Degrees of Freedom 693
    - Mean Squares 693
    - Analysis of Variance Table 694
    - Expected Mean Squares 694
  - 16.6 F Test for Equality of Factor Level Means 698**
    - Test Statistic 698
    - Distribution of  $F^*$  699
    - Construction of Decision Rule 699
  - 16.7 Alternative Formulation of Model 701**
    - Factor Effects Model 701
    - Definition of  $\mu$  702
    - Test for Equality of Factor Level Means 704
  - 16.8 Regression Approach to Single-Factor Analysis of Variance 704**
    - Factor Effects Model with Unweighted Mean 705
    - Factor Effects Model with Weighted Mean 709
    - Cell Means Model 710
  - 16.9 Randomization Tests 712**
  - 16.10 Planning of Sample Sizes with Power Approach 716**
    - Power of F Test 716
    - Use of Table B.12 for Single-Factor Studies 718
    - Some Further Observations on Use of Table B.12 720
  - 16.11 Planning of Sample Sizes to Find "Best" Treatment 721**
    - Cited Reference 722

- Problems 722
- Exercises 730
- Projects 730
- Case Studies 732

## Chapter 17

### Analysis of Factor Level Means 733

- 17.1 Introduction 733
- 17.2 Plots of Estimated Factor Level Means 735
  - Line Plot 735
  - Bar Graph and Main Effects Plot 736
- 17.3 Estimation and Testing of Factor Level Means 737
  - Inferences for Single Factor Level Mean 737
  - Inferences for Difference between Two Factor Level Means 739
  - Inferences for Contrast of Factor Level Means 741
  - Inferences for Linear Combination of Factor Level Means 743
- 17.4 Need for Simultaneous Inference Procedures 744
- 17.5 Tukey Multiple Comparison Procedure 746
  - Studentized Range Distribution 746
  - Simultaneous Estimation 747
  - Simultaneous Testing 747
  - Example 1—Equal Sample Sizes 748
  - Example 2—Unequal Sample Sizes 750
- 17.6 Scheffé Multiple Comparison Procedure 753
  - Simultaneous Estimation 753
  - Simultaneous Testing 754
  - Comparison of Scheffé and Tukey Procedures 755
- 17.7 Bonferroni Multiple Comparison Procedure 756
  - Simultaneous Estimation 756
  - Simultaneous Testing 756
  - Comparison of Bonferroni Procedure with Scheffé and Tukey Procedures 757
  - Analysis of Means 758

### 17.8 Planning of Sample Sizes with Estimation Approach 759

- Example 1—Equal Sample Sizes 759
- Example 2—Unequal Sample Sizes 761

### 17.9 Analysis of Factor Effects when Factor Is Quantitative 762

- Cited References 766
- Problems 767
- Exercises 773
- Projects 774
- Case Studies 774

## Chapter 18

### ANOVA Diagnostics and Remedial Measures 775

- 18.1 Residual Analysis 775
  - Residuals 776
  - Residual Plots 776
  - Diagnosis of Departures from ANOVA Model 778
- 18.2 Tests for Constancy of Error Variance 781
  - Hartley Test 782
  - Brown-Forsythe Test 784
- 18.3 Overview of Remedial Measures 786
- 18.4 Weighted Least Squares 786
- 18.5 Transformations of Response Variable 789
  - Simple Guides to Finding a Transformation 789
  - Box-Cox Procedure 791
- 18.6 Effects of Departures from Model 793
  - Nonnormality 793
  - Unequal Error Variances 794
  - Nonindependence of Error Terms 794
- 18.7 Nonparametric Rank  $F$  Test 795
  - Test Procedure 795
  - Multiple Pairwise Testing Procedure 797
- 18.8 Case Example—Heart Transplant 798
  - Cited References 801
  - Problems 801
  - Exercises 807
  - Projects 807
  - Case Studies 809

## PART FIVE MULTI-FACTOR STUDIES 811

### Chapter 19

#### Two-Factor Studies with Equal Sample Sizes 812

- 19.1** Two-Factor Observational and Experimental Studies 812  
*Examples of Two-Factor Experiments and Observational Studies* 812  
*The One-Factor-at-a-Time (OFAAT) Approach to Experimentation* 815  
*Advantages of Crossed, Multi-Factor Designs* 816
- 19.2** Meaning of ANOVA Model Elements 817  
*Illustration* 817  
*Treatment Means* 817  
*Factor Level Means* 818  
*Main Effects* 818  
*Additive Factor Effects* 819  
*Interacting Factor Effects* 822  
*Important and Unimportant Interactions* 824  
*Transformable and Nontransformable Interactions* 826  
*Interpretation of Interactions* 827
- 19.3** Model I (Fixed Factor Levels) for Two-Factor Studies 829  
*Cell Means Model* 830  
*Factor Effects Model* 831
- 19.4** Analysis of Variance 833  
*Illustration* 833  
*Notation* 834  
*Fitting of ANOVA Model* 834  
*Partitioning of Total Sum of Squares* 836  
*Partitioning of Degrees of Freedom* 839  
*Mean Squares* 839  
*Expected Mean Squares* 840  
*Analysis of Variance Table* 840
- 19.5** Evaluation of Appropriateness of ANOVA Model 842
- 19.6** *F* Tests 843  
*Test for Interactions* 844  
*Test for Factor A Main Effects* 844  
*Test for Factor B Main Effects* 845  
*Kimball Inequality* 846
- 19.7** Strategy for Analysis 847
- 19.8** Analysis of Factor Effects when Factors Do Not Interact 848  
*Estimation of Factor Level Mean* 848  
*Estimation of Contrast of Factor Level Means* 849  
*Estimation of Linear Combination of Factor Level Means* 850  
*Multiple Pairwise Comparisons of Factor Level Means* 850  
*Multiple Contrasts of Factor Level Means* 852  
*Estimates Based on Treatment Means* 853  
*Example 1—Pairwise Comparisons of Factor Level Means* 853  
*Example 2—Estimation of Treatment Means* 855
- 19.9** Analysis of Factor Effects when Interactions Are Important 856  
*Multiple Pairwise Comparisons of Treatment Means* 856  
*Multiple Contrasts of Treatment Means* 857  
*Example 1—Pairwise Comparisons of Treatment Means* 857  
*Example 2—Contrasts of Treatment Means* 860
- 19.10** Pooling Sums of Squares in Two-Factor Analysis of Variance 861
- 19.11** Planning of Sample Sizes for Two-Factor Studies 862  
*Power Approach* 862  
*Estimation Approach* 863  
*Finding the “Best” Treatment* 864  
 Problems 864  
 Exercises 876  
 Projects 876  
 Case Studies 879

## Chapter 20

### Two-Factor Studies—One Case per Treatment 880

- 20.1** No-Interaction Model 880  
     *Model* 881  
     *Analysis of Variance* 881  
     *Inference Procedures* 881  
     *Estimation of Treatment Mean* 884
- 20.2** Tukey Test for Additivity 886  
     *Development of Test Statistic* 886  
     *Remedial Actions if Interaction Effects Are Present* 888
- Cited Reference 889  
 Problems 889  
 Exercises 891  
 Case Study 891

## Chapter 21

### Randomized Complete Block Designs 892

- 21.1** Elements of Randomized Complete Block Designs 892  
     *Description of Designs* 892  
     *Criteria for Blocking* 893  
     *Advantages and Disadvantages* 894  
     *How to Randomize* 895  
     *Illustration* 895
- 21.2** Model for Randomized Complete Block Designs 897
- 21.3** Analysis of Variance and Tests 898  
     *Fitting of Randomized Complete Block Model* 898  
     *Analysis of Variance* 898
- 21.4** Evaluation of Appropriateness of Randomized Complete Block Model 901  
     *Diagnostic Plots* 901  
     *Tukey Test for Additivity* 903
- 21.5** Analysis of Treatment Effects 904
- 21.6** Use of More than One Blocking Variable 905
- 21.7** Use of More than One Replicate in Each Block 906

- 21.8** Factorial Treatments 908
- 21.9** Planning Randomized Complete Block Experiments 909  
     *Power Approach* 909  
     *Estimation Approach* 910  
     *Efficiency of Blocking Variable* 911  
 Problems 912  
 Exercises 916

## Chapter 22

### Analysis of Covariance 917

- 22.1** Basic Ideas 917  
     *How Covariance Analysis Reduces Error Variability* 917  
     *Concomitant Variables* 919
- 22.2** Single-Factor Covariance Model 920  
     *Notation* 921  
     *Development of Covariance Model* 921  
     *Properties of Covariance Model* 922  
     *Generalizations of Covariance Model* 923  
     *Regression Formula of Covariance Model* 924  
     *Appropriateness of Covariance Model* 925  
     *Inferences of Interest* 925
- 22.3** Example of Single-Factor Covariance Analysis 926  
     *Development of Model* 926  
     *Test for Treatment Effects* 928  
     *Estimation of Treatment Effects* 930  
     *Test for Parallel Slopes* 932
- 22.4** Two-Factor Covariance Analysis 933  
     *Covariance Model for Two-Factor Studies* 933  
     *Regression Approach* 934  
     *Covariance Analysis for Randomized Complete Block Designs* 937
- 22.5** Additional Considerations for the Use of Covariance Analysis 939  
     *Covariance Analysis as Alternative to Blocking* 939  
     *Use of Differences* 939  
     *Correction for Bias* 940

*Interest in Nature of Treatment**Effects* 940

Problems 941

Exercise 947

Projects 947

Case Studies 950

**Chapter 23****Two-Factor Studies with Unequal Sample Sizes 951****23.1 Unequal Sample Sizes 951***Notation* 952**23.2 Use of Regression Approach for Testing Factor Effects when Sample Sizes Are Unequal 953***Regression Approach to Two-Factor Analysis of Variance* 953**23.3 Inferences about Factor Effects when Sample Sizes Are Unequal 959***Example 1—Pairwise Comparisons of Factor Level Means* 962*Example 2—Single-Degree-of-Freedom Test* 964**23.4 Empty Cells in Two-Factor Studies 964***Partial Analysis of Factor Effects* 965*Analysis if Model with No Interactions Can Be Employed* 966*Missing Observations in Randomized Complete Block Designs* 967**23.5 ANOVA Inferences when Treatment Means Are of Unequal Importance 970***Estimation of Treatment Means and Factor Effects* 971*Test for Interactions* 972*Tests for Factor Main Effects by Use of Equivalent Regression Models* 972*Tests for Factor Main Effects by Use of Matrix Formulation* 975*Tests for Factor Effects when Weights Are Proportional to Sample Sizes* 977**23.6 Statistical Computing Packages 980**

Problems 981

Exercises 988

Projects 988

Case Studies 990

**Chapter 24****Multi-Factor Studies 992****24.1 ANOVA Model for Three-Factor Studies 992***Notation* 992*Illustration* 993*Main Effects* 993*Two-Factor Interactions* 995*Three-Factor Interactions* 996*Cell Means Model* 996*Factor Effects Model* 997**24.2 Interpretation of Interactions in Three-Factor Studies 998***Learning Time Example 1: Interpretation of Three-Factor Interactions* 998*Learning Time Example 2: Interpretation of Multiple Two-Factor Interactions* 999*Learning Time Example 3: Interpretation of a Single Two-Factor Interaction* 1000**24.3 Fitting of ANOVA Model 1003***Notation* 1003*Fitting of ANOVA Model* 1003*Evaluation of Appropriateness of ANOVA Model* 1005**24.4 Analysis of Variance 1008***Partitioning of Total Sum of Squares* 1008*Degrees of Freedom and Mean Squares* 1009*Tests for Factor Effects* 1009**24.5 Analysis of Factor Effects 1013***Strategy for Analysis* 1013*Analysis of Factor Effects when Factors Do Not Interact* 1014*Analysis of Factor Effects with Multiple Two-Factor Interactions or Three-Factor Interaction* 1016*Analysis of Factor Effects with Single Two-Factor Interaction* 1016*Example—Estimation of Contrasts of Treatment Means* 1018**24.6 Unequal Sample Sizes in Multi-Factor Studies 1019***Tests for Factor Effects* 1019*Inferences for Contrasts of Factor Level Means* 1020

- 24.7 Planning of Sample Sizes 1021  
*Power of F Test for Multi-Factor Studies* 1021  
*Use of Table B.12 for Multi-Factor Studies* 1021  
 Cited Reference 1022  
 Problems 1022  
 Exercises 1027  
 Projects 1027  
 Case Studies 1028

## Chapter 25 Random and Mixed Effects Models 1030

- 25.1 Single-Factor Studies—ANOVA Model II 1031  
*Random Cell Means Model* 1031  
*Questions of Interest* 1034  
*Test whether  $\sigma_{\mu}^2 = 0$*  1035  
*Estimation of  $\mu$*  1038  
*Estimation of  $\sigma_{\mu}^2 / (\sigma_{\mu}^2 + \sigma^2)$*  1040  
*Estimation of  $\sigma^2$*  1041  
*Point Estimation of  $\sigma_{\mu}^2$*  1042  
*Interval Estimation of  $\sigma_{\mu}^2$*  1042  
*Random Factor Effects Model* 1047
- 25.2 Two-Factor Studies—ANOVA Models II and III 1047  
*ANOVA Model II—Random Factor Effects* 1047  
*ANOVA Model III—Mixed Factor Effects* 1049
- 25.3 Two-Factor Studies—ANOVA Tests for Models II and III 1052  
*Expected Mean Squares* 1052  
*Construction of Test Statistics* 1053
- 25.4 Two-Factor Studies—Estimation of Factor Effects for Models II and III 1055  
*Estimation of Variance Components* 1055  
*Estimation of Fixed Effects in Mixed Model* 1056
- 25.5 Randomized Complete Block Design: Random Block Effects 1060  
*Additive Model* 1061  
*Interaction Model* 1064

- 25.6 Three-Factor Studies—ANOVA Models II and III 1066  
*ANOVA Model II—Random Factor Effects* 1066  
*ANOVA Model III—Mixed Factor Effects* 1066  
*Appropriate Test Statistics* 1067  
*Estimation of Effects* 1069
- 25.7 ANOVA Models II and III with Unequal Sample Sizes 1070  
*Maximum Likelihood Approach* 1072  
 Cited References 1077  
 Problems 1077  
 Exercises 1085  
 Projects 1085

## PART SIX SPECIALIZED STUDY DESIGNS 1087

### Chapter 26 Nested Designs, Subsampling, and Partially Nested Designs 1088

- 26.1 Distinction between Nested and Crossed Factors 1088
- 26.2 Two-Factor Nested Designs 1091  
*Development of Model Elements* 1091  
*Nested Design Model* 1092  
*Random Factor Effects* 1093
- 26.3 Analysis of Variance for Two-Factor Nested Designs 1093  
*Fitting of Model* 1093  
*Sums of Squares* 1094  
*Degrees of Freedom* 1095  
*Tests for Factor Effects* 1097  
*Random Factor Effects* 1099
- 26.4 Evaluation of Appropriateness of Nested Design Model 1099
- 26.5 Analysis of Factor Effects in Two-Factor Nested Designs 1100  
*Estimation of Factor Level Means  $\mu_i$*  1100  
*Estimation of Treatment Means  $\mu_{ij}$*  1102  
*Estimation of Overall Mean  $\mu_{..}$*  1103  
*Estimation of Variance Components* 1103

- 26.6** Unbalanced Nested Two-Factor Designs 1104
- 26.7** Subsampling in Single-Factor Study with Completely Randomized Design 1106  
*Model* 1107  
*Analysis of Variance and Tests of Effects* 1108  
*Estimation of Treatment Effects* 1110  
*Estimation of Variances* 1111
- 26.8** Pure Subsampling in Three Stages 1113  
*Model* 1113  
*Analysis of Variance* 1113  
*Estimation of  $\mu_{..}$*  1113
- 26.9** Three-Factor Partially Nested Designs 1114  
*Development of Model* 1114  
*Analysis of Variance* 1115  
 Cited Reference 1119  
 Problems 1119  
 Exercises 1125  
 Projects 1125
- Chapter 27**  
**Repeated Measures and Related Designs** 1127
- 27.1** Elements of Repeated Measures Designs 1127  
*Description of Designs* 1127  
*Advantages and Disadvantages* 1128  
*How to Randomize* 1128
- 27.2** Single-Factor Experiments with Repeated Measures on All Treatments 1129  
*Model* 1129  
*Analysis of Variance and Tests* 1130  
*Evaluation of Appropriateness of Repeated Measures Model* 1134  
*Analysis of Treatment Effects* 1137  
*Ranked Data* 1138  
*Multiple Pairwise Testing Procedure* 1138
- 27.3** Two-Factor Experiments with Repeated Measures on One Factor 1140  
*Description of Design* 1140  
*Model* 1141  
*Analysis of Variance and Tests* 1142  
*Evaluation of Appropriateness of Repeated Measures Model* 1144  
*Analysis of Factor Effects: Without Interaction* 1145  
*Analysis of Factor Effects: With Interaction* 1148  
*Blocking of Subjects in Repeated Measures Designs* 1153
- 27.4** Two-Factor Experiments with Repeated Measures on Both Factors 1153  
*Model* 1154  
*Analysis of Variance and Tests* 1155  
*Evaluation of Appropriateness of Repeated Measures Model* 1157  
*Analysis of Factor Effects* 1157
- 27.5** Regression Approach to Repeated Measures Designs 1161
- 27.6** Split-Plot Designs 1162  
 Cited References 1164  
 Problems 1164  
 Exercise 1171  
 Projects 1171
- Chapter 28**  
**Balanced Incomplete Block, Latin Square, and Related Designs** 1173
- 28.1** Balanced Incomplete Block Designs 1173  
*Advantages and Disadvantages of BIBDs* 1175
- 28.2** Analysis of Balanced Incomplete Block Designs 1177  
*BIBD Model* 1177  
*Regression Approach to Analysis of Balanced Incomplete Block Designs* 1177  
*Analysis of Treatment Effects* 1180  
*Planning of Sample Sizes with Estimation Approach* 1182
- 28.3** Latin Square Designs 1183  
*Basic Ideas* 1183  
*Description of Latin Square Designs* 1184  
*Advantages and Disadvantages of Latin Square Designs* 1185

*Randomization of Latin Square*

*Design* 1185

**28.4** Latin Square Model 1187

**28.5** Analysis of Latin Square

Experiments 1188

*Notation* 1188

*Fitting of Model* 1188

*Analysis of Variance* 1188

*Test for Treatment Effects* 1190

*Analysis of Treatment Effects* 1190

*Residual Analysis* 1191

*Factorial Treatments* 1192

*Random Blocking Variable Effects* 1193

*Missing Observations* 1193

**28.6** Planning Latin Square

Experiments 1193

*Power of F Test* 1193

*Necessary Number of Replications* 1193

*Efficiency of Blocking Variables* 1193

**28.7** Additional Replications with Latin

Square Designs 1195

*Replications within Cells* 1195

*Additional Latin Squares* 1196

**28.8** Replications in Repeated Measures

Studies 1198

*Latin Square Crossover Designs* 1198

*Use of Independent Latin Squares* 1200

*Carryover Effects* 1201

Cited References 1202

Problems 1202

## Chapter 29

### Exploratory Experiments: Two-Level Factorial and Fractional Factorial Designs 1209

**29.1** Two-Level Full Factorial Experiments 1210

*Design of Two-Level Studies* 1210

*Notation* 1210

*Estimation of Factor Effects* 1212

*Inferences about Factor Effects* 1214

**29.2** Analysis of Unreplicated Two-Level Studies 1216

*Pooling of Interactions* 1218

*Pareto Plot* 1219

*Dot Plot* 1220

*Normal Probability Plot* 1221

*Center Point Replications* 1222

**29.3** Two-Level Fractional Factorial Designs 1223

*Confounding* 1224

*Defining Relation* 1227

*Half-Fraction Designs* 1228

*Quarter-Fraction and Smaller-Fraction*

*Designs* 1229

*Resolution* 1231

*Selecting a Fraction of Highest*

*Resolution* 1232

**29.4** Screening Experiments 1239

*$2_{III}^{k-f}$  Fractional Factorial Designs* 1239

*Plackett-Burman Designs* 1240

**29.5** Incomplete Block Designs for Two-Level Factorial Experiments 1240

*Assignment of Treatments to Blocks* 1241

*Use of Center Point Replications* 1243

**29.6** Robust Product and Process Design 1244

*Location and Dispersion Modeling* 1246

*Incorporating Noise Factors* 1250

*Case Study—Clutch Slave Cylinder*

*Experiment* 1252

Cited References 1256

Problems 1256

Exercises 1266

## Chapter 30

### Response Surface Methodology 1267

**30.1** Response Surface Experiments 1267

**30.2** Central Composite Response Surface Designs 1268

*Structure of Central Composite Designs* 1268

*Commonly Used Central Composite Designs* 1270

*Rotatable Central Composite Designs* 1271

*Other Criteria for Choosing a Central Composite Design* 1273

*Blocking Central Composite Designs* 1275



*Additional General-Purpose Response Surface Designs* 1276

**30.3 Optimal Response Surface Designs** 1276

*Purpose of Optimal Designs* 1276

*Optimal Design Approach* 1278

*Design Criteria for Optimal Design Selection* 1279

*Construction of Optimal Response Surface Designs* 1282

*Some Final Cautions* 1283

**30.4 Analysis of Response Surface Experiments** 1284

*Model Interpretation and Visualization* 1284

*Response Surface Optimum Conditions* 1286

**30.5 Sequential Search for Optimum Conditions—Method of Steepest Ascent** 1290

    Cited References 1292

    Problems 1292

    Projects 1295

**Chapter 30 Response Surface Methodology** 1297

    30.1 Response Surface Experiments 1297

    30.2 Central Composite Response Surface Designs 1298

    30.3 Optimal Response Surface Designs 1276

    30.4 Analysis of Response Surface Experiments 1284

    30.5 Sequential Search for Optimum Conditions—Method of Steepest Ascent 1290

**Chapter 29 Factorial and Fractional Factorial Designs** 1209

    29.1 Two-Level Full Factorial Experiments 1210

    29.2 Analysis of Unreplicated Two-Level Studies 1210

    29.3 Design of Two-Level Studies 1210

    29.4 Model Interpretation and Visualization 1212

    29.5 Estimation of Factor Effects 1214

    29.6 Influence of Factor Effects 1214

    29.7 Additional Replications with Latin Square Designs 1215

    29.8 Replications in Repeated Measures Designs 1215

    29.9 Latin Square Designs 1215

    29.10 Use of Independent Latin Squares 1216

    29.11 Carover Effect 1216

    29.12 Cited References 1216

    29.13 Problems 1216

**Chapter 28 Additional Replications with Latin Square Designs** 1197

    28.1 Power of F Test 1197

    28.2 Experiments 1197

    28.3 Analysis of Variance 1198

    28.4 Fitting of Model 1198

    28.5 Notation 1198

    28.6 Randomization of Latin Square Experiments 1198

**Chapter 27 Multiple Factorial Designs** 1171

    27.1 Design of Multiple Factorial Experiments 1171

    27.2 Community-Centered Central Composite Designs 1171

    27.3 Rotatable Central Composite Designs 1171

    27.4 Other Central Composite Designs 1171

    27.5 Blocking Central Composite Designs 1171

    27.6 Other Central Composite Designs 1171

    27.7 Cited References 1171

    27.8 Problems 1171

    27.9 Projects 1171

**Appendix A Some Basic Results in Probability and Statistics** 1297

**Appendix B Tables** 1315

**Appendix C Data Sets** 1348

**Appendix D Rules for Developing ANOVA Models and Tables for Balanced Designs** 1358

**Appendix E Selected Bibliography** 1374

**Index** 1385