Contents

Preface xiii

PART I OVERVIEW AND BASIC APPROACHES

1. Introduction 3
   1.1. The Problem of Missing Data, 3
   1.2. Missing-Data Patterns, 4
   1.3. Mechanisms That Lead to Missing Data, 11
   1.4. A Taxonomy of Missing-Data Methods, 19

2. Missing Data in Experiments 24
   2.1. Introduction, 24
   2.2. The Exact Least Squares Solution with Complete Data, 25
   2.3. The Correct Least Squares Analysis with Missing Data, 27
   2.4. Filling in Least Squares Estimates, 28
      2.4.1. Yates’s Method, 28
      2.4.2. Using a Formula for the Missing Values, 29
      2.4.3. Iterating to Find the Missing Values, 29
      2.4.4. ANCOVA with Missing-Value Covariates, 30
   2.5. Bartlett’s ANCOVA Method, 30
      2.5.1. Useful Properties of Bartlett’s Method, 30
      2.5.2. Notation, 30
      2.5.3. The ANCOVA Estimates of Parameters and Missing
             Y Values, 31
      2.5.4. ANCOVA Estimates of the Residual Sums of Squares
             and the Covariance Matrix of \( \hat{\beta} \), 31
2.6. Least Squares Estimates of Missing Values by ANCOVA Using Only Complete-Data Methods, 33
2.7. Correct Least Squares Estimates of Standard Errors and One Degree of Freedom Sums of Squares, 35
2.8. Correct Least Squares Sums of Squares with More Than One Degree of Freedom, 37

3. Complete-Case and Available-Case Analysis, Including Weighting Methods 41
   3.1. Introduction, 41
   3.2. Complete-Case Analysis, 41
   3.3. Weighted Complete-Case Analysis, 44
       3.3.1. Weighting Adjustments, 44
       3.3.2. Added Variance from Nonresponse Weighting, 50
       3.3.3. Post-Stratification and Raking To Known Margins, 51
       3.3.4. Inference from Weighted Data, 53
       3.3.5. Summary of Weighting Methods, 53
   3.4. Available-Case Analysis, 53

4. Single Imputation Methods 59
   4.1. Introduction, 59
   4.2. Imputing Means from a Predictive Distribution, 61
       4.2.1. Unconditional Mean Imputation, 61
       4.2.2. Conditional Mean Imputation, 62
   4.3. Imputing Draws from a Predictive Distribution, 64
       4.3.1. Draws Based on Explicit Models, 64
       4.3.2. Draws Based on Implicit Models, 66
   4.4. Conclusions, 72

5. Estimation of Imputation Uncertainty 75
   5.1. Introduction, 75
   5.2. Imputation Methods that Provide Valid Standard Errors from a Single Filled-in Data Set, 76
   5.3. Standard Errors for Imputed Data by Resampling, 79
       5.3.1. Bootstrap Standard Errors, 79
       5.3.2. Jackknife Standard Errors, 81
   5.4. Introduction to Multiple Imputation, 85
   5.5. Comparison of Resampling Methods and Multiple Imputation, 89
PART II  LIKELIHOOD-BASED APPROACHES TO THE
ANALYSIS OF MISSING DATA

6. Theory of Inference Based on the Likelihood Function  97
   6.1. Review of Likelihood-Based Estimation for Complete Data, 97
       6.1.1. Maximum Likelihood Estimation, 97
       6.1.2. Rudiments of Bayes Estimation, 104
       6.1.3. Large-Sample Maximum Likelihood and Bayes
               Inference, 105
       6.1.4. Bayes Inference Based on the Full Posterior
               Distribution, 112
       6.1.5. Simulating Draws from Posterior Distributions, 115
   6.2. Likelihood-Based Inference with Incomplete Data, 117
   6.3. A Generally Flawed Alternative to Maximum Likelihood:
        Maximizing Over the Parameters and the Missing Data, 124
       6.3.1. The Method, 124
       6.3.2. Background, 124
       6.3.3. Examples, 125
   6.4. Likelihood Theory for Coarsened Data, 127

7. Factored Likelihood Methods, Ignoring the Missing-Data
   Mechanism  133
   7.1. Introduction, 133
   7.2. Bivariate Normal Data with One Variable Subject to
        Nonresponse: ML Estimation, 133
       7.2.1. ML Estimates, 135
       7.2.2. Large-Sample Covariance Matrix, 139
   7.3. Bivariate Normal Monotone Data: Small-Sample Inference, 140
   7.4. Monotone Data With More Than Two Variables, 143
       7.4.1. Multivariate Data With One Normal Variable Subject
               to Nonresponse, 143
       7.4.2. Factorization of the Likelihood for a General Monotone
               Pattern, 144
       7.4.3. Computation for Monotone Normal Data via the Sweep
               Operator, 148
       7.4.4. Bayes Computation for Monotone Normal Data via
               the Sweep Operator, 155
   7.5. Factorizations for Special Nonmonotone Patterns, 156
8. Maximum Likelihood for General Patterns of Missing Data: Introduction and Theory with Ignorable Nonresponse

8.1. Alternative Computational Strategies, 164
8.2. Introduction to the EM Algorithm, 166
8.3. The E and M Steps of EM, 167
8.4. Theory of the EM Algorithm, 172
  8.4.1. Convergence Properties, 172
  8.4.2. EM for Exponential Families, 175
  8.4.3. Rate of Convergence of EM, 177
8.5. Extensions of EM, 179
  8.5.1. ECM Algorithm, 179
  8.5.2. ECME and AECM Algorithms, 183
  8.5.3. PX-EM Algorithm, 184
8.6. Hybrid Maximization Methods, 186

9. Large-Sample Inference Based on Maximum Likelihood Estimates

9.1. Standard Errors Based on the Information Matrix, 190
9.2. Standard Errors via Methods that do not Require Computing and Inverting an Estimate of the Observed Information Matrix, 191
  9.2.1. Supplemental EM Algorithm, 191
  9.2.2. Bootstrapping the Observed Data, 196
  9.2.3. Other Large Sample Methods, 197
  9.2.4. Posterior Standard Errors from Bayesian Methods, 198

10. Bayes and Multiple Imputation

10.1. Bayesian Iterative Simulation Methods, 200
  10.1.1. Data Augmentation, 200
  10.1.2. The Gibbs’ Sampler, 203
  10.1.3. Assessing Convergence of Iterative Simulations, 206
  10.1.4. Some Other Simulation Methods, 208
10.2. Multiple Imputation, 209
  10.2.1. Large-Sample Bayesian Approximation of the Posterior Mean and Variance Based on a Small Number of Draws, 209
  10.2.2. Approximations Using Test Statistics, 212
  10.2.3. Other Methods for Creating Multiple Imputations, 214
PART III LIKELIHOOD-BASED APPROACHES TO THE ANALYSIS OF INCOMPLETE DATA: SOME EXAMPLES

11. Multivariate Normal Examples, Ignoring the Missing-Data Mechanism

11.1. Introduction, 223
11.2. Inference for a Mean Vector and Covariance Matrix with Missing Data Under Normality, 223
   11.2.1. The EM Algorithm for Incomplete Multivariate Normal Samples, 226
   11.2.2. Estimated Asymptotic Covariance Matrix of \((\theta - \hat{\theta})\), 226
   11.2.3. Bayes Inference for the Normal Model via Data Augmentation, 227
11.3. Estimation with a Restricted Covariance Matrix, 231
11.4. Multiple Linear Regression, 237
   11.4.1. Linear Regression with Missing Values Confined to the Dependent Variable, 237
   11.4.2. More General Linear Regression Problems with Missing Data, 239
11.5. A General Repeated-Measures Model with Missing Data, 241
11.6. Time Series Models, 246
   11.6.1. Introduction, 246
   11.6.2. Autoregressive Models for Univariate Time Series with Missing Values, 246
   11.6.3. Kalman Filter Models, 248

12. Robust Estimation

12.1. Introduction, 253
12.2. Robust Estimation for a Univariate Sample, 253
12.3. Robust Estimation of the Mean and Covariance Matrix, 255
   12.3.1. Multivariate Complete Data, 255
   12.3.2. Robust Estimation of the Mean and Covariance Matrix from Data with Missing Values, 257
   12.3.3. Adaptive Robust Multivariate Estimation, 259
   12.3.4. Bayes Inferences for the \(t\) Model, 259
12.4. Further Extensions of the \(t\) Model, 260

13. Models for Partially Classified Contingency Tables, Ignoring the Missing-Data Mechanism

13.1. Introduction, 266
13.2. Factored Likelihoods for Monotone Multinomial Data, 267
  13.2.1. Introduction, 267
  13.2.2. ML Estimation for Monotone Patterns, 268
  13.2.3. Precision of Estimation, 275

13.3. ML and Bayes Estimation for Multinomial Samples with General Patterns of Missing Data, 278

13.4. Loglinear Models for Partially Classified Contingency Tables, 281
  13.4.1. The Complete-Data Case, 281
  13.4.2. Loglinear Models for Partially Classified Tables, 285
  13.4.3. Goodness-of-Fit Tests for Partially Classified Data, 289

14. Mixed Normal and Non-normal Data with Missing Values, Ignoring the Missing-Data Mechanism 292

  14.1. Introduction, 292
  14.2. The General Location Model, 292
    14.2.1. The Complete-Data Model and Parameter Estimates, 292
    14.2.2. ML Estimation with Missing Values, 294
    14.2.3. Details of the E Step Calculations, 296
    14.2.4. Bayes Computations for the Unrestricted General Location Model, 298

  14.3. The General Location Model with Parameter Constraints, 300
    14.3.1. Introduction, 300
    14.3.2. Restricted Models for the Cell Means, 300
    14.3.3. Loglinear Models for the Cell Probabilities, 303
    14.3.4. Modifications to the Algorithms of Sections 14.2.2 and 14.2.3 for Parameter Restrictions, 303
    14.3.5. Simplifications when the Categorical Variables are More Observed than the Continuous Variables, 305

  14.4. Regression Problems Involving Mixtures of Continuous and Categorical Variables, 306
    14.4.1. Normal Linear Regression with Missing Continuous or Categorical Covariates, 306
    14.4.2. Logistic Regression with Missing Continuous or Categorical Covariates, 308

  14.5. Further Extensions of the General Location Model, 309

15. Nonignorable Missing-Data Models 312

  15.1. Introduction, 312
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.2</td>
<td>Likelihood Theory for Nonignorable Models</td>
<td>315</td>
</tr>
<tr>
<td>15.3</td>
<td>Models with Known Nonignorable Missing-Data Mechanisms: Grouped and Rounded Data</td>
<td>316</td>
</tr>
<tr>
<td>15.4</td>
<td>Normal Selection Models</td>
<td>321</td>
</tr>
<tr>
<td>15.5</td>
<td>Normal Pattern-Mixture Models</td>
<td>327</td>
</tr>
<tr>
<td>15.5.1</td>
<td>Univariate Normal Pattern-Mixture Models</td>
<td>327</td>
</tr>
<tr>
<td>15.5.2</td>
<td>Bivariate Normal Pattern-Mixture Models Identified via Parameter Restrictions</td>
<td>331</td>
</tr>
<tr>
<td>15.6</td>
<td>Nonignorable Models for Normal Repeated-Measures Data</td>
<td>336</td>
</tr>
<tr>
<td>15.7</td>
<td>Nonignorable Models for Categorical Data</td>
<td>340</td>
</tr>
</tbody>
</table>

References 349

Author Index 365

Subject Index 371