Contents

1 Introduction and Basic Concepts 1
- 1.1 Overview 1
- 1.2 Introduction 1
- 1.3 Unit and Item Nonresponse 2
- 1.4 Patterns of Observed and Missing Values 3
- 1.5 Inferential Approaches and Concepts 4
  - 1.5.1 Design-Based Approach 5
  - 1.5.2 Model-Based Approaches 6
- 1.6 Statistical Inference and Missing Data 8
- 1.7 The Method of Multiple Imputation 9
- 1.8 Illustration of Concepts and Techniques 10
  - 1.8.1 Unit and Item Nonresponse, (Non)response Pattern 10
  - 1.8.2 Design-Based Approach 10
  - 1.8.3 Model-Based Approaches 14
- 1.9 Outline 20
- References 21

2 Missing Data Mechanism and Ignorability 23
- 2.1 Overview 23
- 2.2 Introduction 23
  - 2.2.1 Some Notation 24
- 2.3 The Missing Data Mechanism 25
  - 2.3.1 Missing Completely at Random (MCAR) 26
  - 2.3.2 Missing at Random (MAR) 28
  - 2.3.3 Missing Not at Random (MNAR) 30
- 2.4 Ignorability of the Missing Data Mechanism 31
  - 2.4.1 Frequentist Approach 33
  - 2.4.2 Likelihood Approach 35
  - 2.4.3 Bayesian Approach 40
- 2.5 Supplementary Notes 41

2.6 Diagnostic Tools 45
2.7 Summary 50
References 51

3 Missing Data Methods 53
3.1 Overview 53
3.2 Ad-hoc Methods 53
3.2.1 Complete Case and Available Case Analysis 54
3.2.2 Completing Data Sets 55
3.3 Maximum Likelihood Estimation 62
3.3.1 Introduction 62
3.3.2 The Ignorable Case 65
3.3.3 The Nonignorable Case 70
3.4 Weighting 75
3.5 Imputation Methods 78
3.5.1 Single Imputation 79
3.5.2 Multiple Imputation 80
3.6 Summary 81
References 82

4 Multiple Imputation: Theory 85
4.1 Overview 85
4.2 Multiple Imputation: Introduction 85
4.3 Theoretical Background 86
4.3.1 Valid Inferences 86
4.3.2 Bayesian Motivation 88
4.3.3 The Combining Rules 89
4.3.4 Frequentist Evaluation 90
4.3.5 Inferences Based on Finite M 98
4.4 How to Generate Multiple Imputations 100
4.4.1 General 100
4.4.2 Monotone-Distinct Structure 102
4.4.3 Beyond Monotone-Distinct Structures 106
4.4.4 Arbitrary Missing Data Patterns: A General Procedure 107
4.4.5 Generating Imputations by Joint Modeling 109
4.4.6 Generating Imputations by Fully Conditional Specification 110
4.5 Further Topics and Open Problems 111
4.5.1 Transforming, Rounding and Trimming 111
4.5.2 Imputation of Functions of Variables 115
4.5.3 Imputing Subgroups 118
4.5.4 The Imputation Models: (Non-)ignorability 122
4.5.5 Reproducible Results 122
4.5.6 Some Additional Notes 123

4.5.7 Choosing from a Large Number of Imputation Techniques ... 124
4.6 Summary ... 128
References ... 129

5 Multiple Imputation: Application ... 133
5.1 Overview ... 133
5.2 Data and Substantive Model ... 135
5.2.1 The CrimoC data ... 135
5.2.2 Growth Curve Models ... 137
5.3 Multiple Imputation of Missing Data ... 142
5.3.1 Reading in the CrimoC data ... 142
5.3.2 Missing Data Inspection ... 143
5.3.3 Choosing the Multiple Imputation Framework ... 149
5.3.4 Building the Imputation Model ... 149
5.4 Multiple Imputation with `mice` ... 150
5.4.1 Argument `method` ... 151
5.4.2 Argument `predictorMatrix` ... 151
5.4.3 Multiple Imputation by Predictive Mean Matching ... 152
5.4.4 Summary ... 157
5.5 Other R Packages and Functions for Multivariate Data Imputation ... 158
5.5.1 `norm2` ... 158
5.5.2 `Amelia` ... 165
5.5.3 `mi` ... 169
5.5.4 `aregImpute` ... 171
5.5.5 Comparison of Results ... 172
5.6 Multiple Imputation of Clustered or Panel Data Using Multilevel Models ... 174
5.6.1 A Brief Introduction to Multilevel Regression Models ... 175
5.6.2 Strategies to Impute Clustered or Panel Data ... 176
5.6.3 Application ... 177
5.7 Repeated Data Analysis and Pooling of Results in R ... 183
5.7.1 *t*-test ... 184
5.7.2 Linear Model ... 187
5.7.3 Growth Model ... 188
5.7.4 Generalized Linear Mixed-Effects Model ... 190
5.7.5 Two-Level Negative Binomial Hurdle Model ... 193
5.7.6 Correlation Coefficients ... 195
5.7.7 Further Useful Tools for Multiple Imputation Inference ... 196
5.8 Diagnostic Tools ... 204
5.8.1 Inspection of Imputed Values Using the `boxplot()` Function ... 204
5.8.2 Comparison of the Distributions of Observed and Imputed Values Using the `hist()` Function ... 206

5.8.3 Comparison of Observed and Imputed Values Using the `stripplot()` Function ... 208
5.8.4 Comparison of Observed and Imputed Values Using `compare.percent.count()` ... 209
5.8.5 Comparison of Observed and Imputed Values Using `compare.obs.imp()` ... 210
5.8.6 Assessing the Suitability of the Imputation Method by Means of Monte Carlo Simulation ... 211
References ... 214

6 Multiple Imputation: New Developments ... 219
6.1 Overview ... 219
6.2 Parametric Multiple Imputation of Incomplete Count Data ... 220
6.2.1 Count Data Modeling in a Nutshell ... 221
6.2.2 The `countimp` Package ... 222
6.2.3 Discussion ... 230
6.3 `ImputeRobust`—Imputation Based on Generalized Additive Models for Location, Scale and Shape ... 230
6.3.1 `ImputeRobust` in a Nutshell ... 231
6.3.2 Application ... 233
6.3.3 Discussion ... 235
6.4 Quantile Regression Based Multiple Imputation ... 235
6.4.1 Application ... 237
6.4.2 Discussion ... 238
6.5 Two-Level Predictive Mean Matching ... 239
6.5.1 The General Idea ... 239
6.5.2 Application ... 240
6.5.3 Discussion ... 240
6.6 The Growth Curve ZIP Model Revisited ... 241
6.7 Future Research—Multiple Imputation After More Than Three Decades ... 251
References ... 254

A Matrix Algebra, Random Variables and Some Technical Results ... 257
A.1 Matrix Algebra ... 257
A.1.1 Vectors and Matrices ... 257
A.1.2 Basic Operations ... 260
A.1.3 Properties of Matrices ... 262
A.2 Random Variables and Statistics ... 265
A.2.1 Sample Moments ... 266
A.2.2 Random Variables ... 267
A.2.3 Some Results on Expectations, Variances and Covariances ... 269
A.2.4 Normal Distribution ... 270
A.2.5 Conditional Normal Distribution ... 271

A.2.6 Truncated Normal Distribution and Normal Sample Selection Model 272
A.3 Estimation 276
A.3.1 Least Squares 276
A.3.2 Maximum Likelihood 278
A.3.3 Information Criteria 283
A.4 Multiple Imputation: Some Technical Results 283
A.4.1 Derivation of (4.6), (4.8) and (4.9) 284
A.4.2 Simplification of $f(\mathrm{u}_{(0)} \mid \mathrm{u}_{(1)}, \theta)$ and $p(\theta \mid \mathrm{u}_{(1)})$ 284
References 285

Glossary 287

Index 289