

Contents

1 • <u>An Introduction to Missing Data</u>	1
1.1 Introduction	1
1.2 Chapter Overview	2
1.3 Missing Data Patterns	2
1.4 A Conceptual Overview of Missing Data Theory	5
1.5 A More Formal Description of Missing Data Theory	9
1.6 Why Is the Missing Data Mechanism Important?	13
1.7 How Plausible Is the Missing at Random Mechanism?	14
1.8 An Inclusive Analysis Strategy	16
1.9 Testing the Missing Completely at Random Mechanism	17
1.10 Planned Missing Data Designs	21
1.11 The Three-Form Design	23
1.12 Planned Missing Data for Longitudinal Designs	28
1.13 Conducting Power Analyses for Planned Missing Data Designs	30
1.14 Data Analysis Example	32
1.15 Summary	35
1.16 Recommended Readings	36
2 • <u>Traditional Methods for Dealing with Missing Data</u>	37
2.1 Chapter Overview	37
2.2 An Overview of Deletion Methods	39
2.3 Listwise Deletion	39
2.4 Pairwise Deletion	40
2.5 An Overview of Single Imputation Methods	42
2.6 Arithmetic Mean Imputation	42
2.7 Regression Imputation	44
2.8 Stochastic Regression Imputation	46
2.9 Hot-Deck Imputation	49
2.10 Similar Response Pattern Imputation	49
2.11 Averaging the Available Items	50
2.12 Last Observation Carried Forward	51
2.13 An Illustrative Computer Simulation Study	52

2.14	Summary	54	
2.15	Recommended Readings	55	
3	• An Introduction to Maximum Likelihood Estimation		56
3.1	Chapter Overview	56	
3.2	The Univariate Normal Distribution	56	
3.3	The Sample Likelihood	59	
3.4	The Log-Likelihood	60	
3.5	Estimating Unknown Parameters	60	
3.6	The Role of First Derivatives	63	
3.7	Estimating Standard Errors	65	
3.8	Maximum Likelihood Estimation with Multivariate Normal Data	69	
3.9	A Bivariate Analysis Example	73	
3.10	Iterative Optimization Algorithms	75	
3.11	Significance Testing Using the Wald Statistic	77	
3.12	The Likelihood Ratio Test Statistic	78	
3.13	Should I Use the Wald Test or the Likelihood Ratio Statistic?	79	
3.14	Data Analysis Example 1	80	
3.15	Data Analysis Example 2	81	
3.16	Summary	83	
3.17	Recommended Readings	85	
4	• Maximum Likelihood Missing Data Handling		86
4.1	Chapter Overview	86	
4.2	The Missing Data Log-Likelihood	88	
4.3	How Do the Incomplete Data Records Improve Estimation?	92	
4.4	An Illustrative Computer Simulation Study	95	
4.5	Estimating Standard Errors with Missing Data	97	
4.6	Observed versus Expected Information	98	
4.7	A Bivariate Analysis Example	99	
4.8	An Illustrative Computer Simulation Study	102	
4.9	An Overview of the EM Algorithm	103	
4.10	A Detailed Description of the EM Algorithm	105	
4.11	A Bivariate Analysis Example	106	
4.12	Extending EM to Multivariate Data	110	
4.13	Maximum Likelihood Estimation Software Options	112	
4.14	Data Analysis Example 1	113	
4.15	Data Analysis Example 2	115	
4.16	Data Analysis Example 3	118	
4.17	Data Analysis Example 4	119	
4.18	Data Analysis Example 5	122	
4.19	Summary	125	
4.20	Recommended Readings	126	
5	• Improving the Accuracy of Maximum Likelihood Analyses		127
5.1	Chapter Overview	127	
5.2	The Rationale for an Inclusive Analysis Strategy	127	
5.3	An Illustrative Computer Simulation Study	129	
5.4	Identifying a Set of Auxiliary Variables	131	

5.5	Incorporating Auxiliary Variables into a Maximum Likelihood Analysis	133
5.6	The Saturated Correlates Model	134
5.7	The Impact of Non-Normal Data	140
5.8	Robust Standard Errors	141
5.9	Bootstrap Standard Errors	145
5.10	The Rescaled Likelihood Ratio Test	148
5.11	Bootstrapping the Likelihood Ratio Statistic	150
5.12	Data Analysis Example 1	154
5.13	Data Analysis Example 2	155
5.14	Data Analysis Example 3	157
5.15	Summary	161
5.16	Recommended Readings	163
6	<u>An Introduction to Bayesian Estimation</u>	164
6.1	Chapter Overview	164
6.2	What Makes Bayesian Statistics Different?	165
6.3	A Conceptual Overview of Bayesian Estimation	165
6.4	Bayes' Theorem	170
6.5	An Analysis Example	171
6.6	How Does Bayesian Estimation Apply to Multiple Imputation?	175
6.7	The Posterior Distribution of the Mean	176
6.8	The Posterior Distribution of the Variance	179
6.9	The Posterior Distribution of a Covariance Matrix	183
6.10	Summary	185
6.11	Recommended Readings	186
7	<u>The Imputation Phase of Multiple Imputation</u>	187
7.1	Chapter Overview	187
7.2	A Conceptual Description of the Imputation Phase	190
7.3	A Bayesian Description of the Imputation Phase	191
7.4	A Bivariate Analysis Example	194
7.5	Data Augmentation with Multivariate Data	199
7.6	Selecting Variables for Imputation	201
7.7	The Meaning of Convergence	202
7.8	Convergence Diagnostics	203
7.9	Time-Series Plots	204
7.10	Autocorrelation Function Plots	207
7.11	Assessing Convergence from Alternate Starting Values	209
7.12	Convergence Problems	210
7.13	Generating the Final Set of Imputations	211
7.14	How Many Data Sets Are Needed?	212
7.15	Summary	214
7.16	Recommended Readings	216
8	<u>The Analysis and Pooling Phases of Multiple Imputation</u>	217
8.1	Chapter Overview	217
8.2	The Analysis Phase	218
8.3	Combining Parameter Estimates in the Pooling Phase	219
8.4	Transforming Parameter Estimates Prior to Combining	220

8.5	Pooling Standard Errors	221
8.6	The Fraction of Missing Information and the Relative Increase in Variance	224
8.7	When Is Multiple Imputation Comparable to Maximum Likelihood?	227
8.8	An Illustrative Computer Simulation Study	229
8.9	Significance Testing Using the t Statistic	230
8.10	An Overview of Multiparameter Significance Tests	233
8.11	Testing Multiple Parameters Using the D_1 Statistic	233
8.12	Testing Multiple Parameters by Combining Wald Tests	239
8.13	Testing Multiple Parameters by Combining Likelihood Ratio Statistics	240
8.14	Data Analysis Example 1	242
8.15	Data Analysis Example 2	245
8.16	Data Analysis Example 3	247
8.17	Summary	252
8.18	Recommended Readings	252
9	• Practical Issues in Multiple Imputation	254
9.1	Chapter Overview	254
9.2	Dealing with Convergence Problems	254
9.3	Dealing with Non-Normal Data	259
9.4	To Round or Not to Round?	261
9.5	Preserving Interaction Effects	265
9.6	Imputing Multiple-Item Questionnaires	269
9.7	Alternate Imputation Algorithms	272
9.8	Multiple-Imputation Software Options	278
9.9	Data Analysis Example 1	279
9.10	Data Analysis Example 2	281
9.11	Summary	283
9.12	Recommended Readings	286
10	• Models for Missing Not at Random Data	287
10.1	Chapter Overview	287
10.2	An Ad Hoc Approach to Dealing with MNAR Data	289
10.3	The Theoretical Rationale for MNAR Models	290
10.4	The Classic Selection Model	291
10.5	Estimating the Selection Model	295
10.6	Limitations of the Selection Model	296
10.7	An Illustrative Analysis	297
10.8	The Pattern Mixture Model	298
10.9	Limitations of the Pattern Mixture Model	300
10.10	An Overview of the Longitudinal Growth Model	301
10.11	A Longitudinal Selection Model	303
10.12	Random Coefficient Selection Models	305
10.13	Pattern Mixture Models for Longitudinal Analyses	306
10.14	Identification Strategies for Longitudinal Pattern Mixture Models	307
10.15	Delta Method Standard Errors	309
10.16	Overview of the Data Analysis Examples	312
10.17	Data Analysis Example 1	314
10.18	Data Analysis Example 2	315
10.19	Data Analysis Example 3	317
10.20	Data Analysis Example 4	321
10.21	Summary	326
10.22	Recommended Readings	328

11 • Wrapping Things Up: Some Final Practical Considerations	329
11.1 Chapter Overview	329
11.2 Maximum Likelihood Software Options	329
11.3 Multiple-Imputation Software Options	333
11.4 Choosing between Maximum Likelihood and Multiple Imputation	336
11.5 Reporting the Results from a Missing Data Analysis	340
11.6 Final Thoughts	343
11.7 Recommended Readings	344
References	347
Author Index	359
Subject Index	365
About the Author	377

The companion website (www.appliedmissingdata.com) includes data files and syntax for the examples in the book, as well as up-to-date information on software.