

Contents

Section 1 Missing Data Theory

1 Missing Data Theory	3
Overview	3
Missing Data: What Is It?	4
Missing Data: History, Objectives, and Challenges	5
Terms	6
Model	6
Missingness	7
Distribution	8
Mechanisms of Missingness	8
Causes of Missingness	9
Mapping Causes of Missingness onto the Three Missingness Mechanisms	12
Missing Completely at Random (MCAR)	12
Missing at Random (MAR)	13
Not Missing at Random (NMAR; aka Missing Not at Random; MNAR)	15
MAR Versus NMAR Revisited	17
Can We Ever Know Whether MAR Holds?	17
Maybe We Can Know if MAR Holds	18
Measuring Estimation Bias	19
Factors Affecting Standardized Bias	20
Estimating Statistical and Practical Significance of Bias in Real Data	21
Percent Missing	22
Estimating r_{XY}	22
Estimating r_{ZY}	22
Estimating r_{ZR}	25

Sensitivity Analysis.....	26
Plausibility of MAR Given in Tables 1.4 and 1.5.....	27
Limitations (Nuisance Factors) Relating to Figures in Tables 1.4 and 1.5	28
Call for New Measures of the Practical Significance of Bias.....	29
Other Limitations of Figures Presented in Tables 1.4 and 1.5	32
Another Substantive Consideration: A Taxonomy of Attrition.....	33
The Value of Missing Data Diagnostics.....	34
Nonignorable Methods.....	36
Sensitivity Analysis.....	36
Design and Measurement Strategies for Assuring MAR.....	38
Measure the Possible Causes of Missingness	38
Measure and Include Relevant Auxiliary Variables in Missing Data Analysis Model.....	43
Track and Collect Data on a Random Sample of Those Initially Missing	43
References.....	44
2 Analysis of Missing Data	47
Goals of Analysis	47
Older Approaches to Handling Missing Data	47
Complete Cases Analysis (aka Listwise Deletion)	48
Pairwise Deletion	51
Mean Substitution	51
Averaging the Available Variables	51
Regression-Based Single Imputation	52
Basics of the Recommended Methods.....	53
Full Information Maximum Likelihood (FIML).....	53
Basics of the EM Algorithm for Covariance Matrices.....	54
Basics of Normal-Model Multiple Imputation	56
What Analyses Work with MI?.....	59
Normal-Model MI with Categorical Variables	60
Normal-Model MI with Longitudinal Data	60
Imputation for Statistical Interactions: The Imputation Model Must Be at Least as Complex as the Analysis Model	62
Normal-Model MI with ANOVA.....	63
Analyses for Which MI Is not Necessary	63
Missing Data Theory and the Comparison Between MI/ML and Other Methods.....	63
Estimation Bias with Multiple Regression Models	64
Missing Data Theory and the Comparison Between MI and ML.....	65
MI and ML and the Inclusion of Auxiliary Variables	65
MI Versus ML, the Importance of Number of Imputations	66
Computational Effort and Adjustments in Thinking About MI	67
References.....	68

Section 2 Multiple Imputation and Basic Analysis

3 Multiple Imputation with Norm 2.03	73
Step-by-Step Instructions for Multiple Imputation with NORM 2.03	73
Running NORM (Step 1): Getting NORM.....	73
Running NORM (Step 2): Preparing the Data Set.....	74
Writing Data Out of SPSS	75
Writing a Variable Names File from SPSS	76
Empirical Example Used Throughout this Chapter	76
Running NORM (Step 3): Variables	76
Running NORM (Step 4): The “Summarize” Tab	78
Running NORM (Step 5): EM Algorithm	81
Running NORM (Step 6): Impute from (EM) Parameters	86
Running NORM (Step 7): Data Augmentation (and Imputation).....	88
Running NORM (Step 8): Data Augmentation Diagnostics.....	89
References.....	94
4 Analysis with SPSS (Versions Without MI Module)	
Following Multiple Imputation with Norm 2.03	95
Analysis with the Single Data Imputed from EM Parameters.....	95
Before Running MIAutomate Utility.....	96
What the MIAutomate Utility Does.....	97
Files Expected to Make Automation Utility Work.....	97
Installing the MIAutomate Utility	98
Running the Automation Utility.....	98
Products of the Automation Utility	99
Analysis with the Single Data Set Imputed from EM Parameters	99
Analysis Following Multiple Imputation.....	100
Automation of SPSS Regression Analysis with Multiple Imputed Data Sets	101
Running the Automation Utility.....	101
Products of the Automation Utility	102
Rationale for Having Separate Input and Output Automation Utilities.....	103
Multiple Linear Regression in SPSS with Multiple Imputed Data Sets, Step 1	103
Multiple Linear Regression in SPSS with Multiple Imputed Data Sets, Step 2	104
Variability of Results with Multiple Imputation	106
A Note About Ethics in Multiple Imputation	106
Other SPSS Analyses with Multiple Imputed Data Sets.....	107

5 Multiple Imputation and Analysis with SPSS 17-20	111
Step-by-Step Instructions for Multiple Imputation with SPSS 17-20	111
Running SPSS 17/20 MI (Step 1): Getting SPSS MI	112
Running SPSS 17-20 MI (Step 2): Preparing the Data Set.....	112
Empirical Example Used Throughout this Chapter	113
Running SPSS 17-20 MI (Step 3): Variables	113
Running SPSS 17-20 MI (Step 4): Missingness Summary	113
Running SPSS 17-20 MI (Step 5): EM Algorithm	116
Running SPSS 17-20 MI (Step 6): Impute from (EM) Parameters	118
Running SPSS 17-20 MI (Step 7): MCMC (and Imputation)	118
Running SPSS 17-20 MI (Step 8): MCMC Diagnostics	119
Analysis of Multiple Data Sets Imputed with SPSS 17-20.....	120
Split File	120
Multiple Linear Regression in SPSS with Multiple Imputed Data Sets	120
Binary Logistic Regression in SPSS with Multiple Imputed Data Sets	121
SPSS 17-20 Analysis of Norm-Imputed Data: Analysis with the Single Data Imputed from EM Parameters.....	122
Before Running MIAutomate Utility	123
What the MIAutomate Utility Does.....	124
Files Expected to Make the MIAutomate Utility Work.....	124
Analysis with the Single Data Set Imputed from EM Parameters	126
SPSS 17-20 Analysis of Norm-Imputed Data: Analysis of Multiple Data Sets Imputed with Norm 2.03	126
Automation of SPSS Regression Analysis with Multiple Imputed Data Sets	128
Running the Automation Utility.....	128
Products of the Automation Utility	129
Setting Up Norm-Imputed Data for Analysis with SPSS 17-20	130
Multiple Linear Regression in SPSS with Norm-Imputed Data Sets	130
Binary Logistic Regression in SPSS with Norm-Imputed Data Sets	130
Other Analyses in SPSS with Norm-Imputed Data Sets.....	130
References.....	131
 6 Multiple Imputation and Analysis with Multilevel (Cluster) Data	133
Imputation for Multilevel Data Analysis	134
Taking Cluster Structure into Account (Random Intercepts Models).....	135

Limitations of the Random Intercepts, Hybrid Dummy Coding, Approach.....	137
Normal Model MI for Random Intercepts and Random Slopes Models.....	138
Limitations with the Impute-Within-Individual-Clusters Strategy.....	138
Multilevel Analysis of Norm-Imputed Data with SPSS/Mixed.....	139
Preparation of Data Imputed with Another Program	141
Multilevel Analysis of Norm-Imputed Data with SPSS 17-20/Mixed.....	141
Setting Up Norm-Imputed Data for Analysis with SPSS 17-20	142
Multiple Linear Mixed Regression in SPSS 17-20 with Norm-Imputed Data Sets	142
Multiple Linear Mixed Regression in SPSS 15/16 with Norm-Imputed Data Sets	143
Multilevel Analysis of Norm-Imputed Data with HLM 7	145
Step 1: Imputation with Norm 2.03	145
Step 2: Run MIAutomate Utility.....	145
Step 3: Enter HLM Information Window: Executable Information	146
Step 4: Enter HLM Information Window: HLM Model Information	146
Step 5: MI Inference	147
Limitations of the MIAutomate Utility for HLM	148
Special Issues Relating to Missing Data Imputation in Multilevel Data Situations	149
Number of Level-2 Units	149
Random Slopes Models	149
3-Level Models	149
Other MI Models.....	150
References.....	150
7 Multiple Imputation and Analysis with SAS	151
Step-by-Step Instructions for Multiple Imputation with PROC MI	152
Running PROC MI (Step 1): Getting SAS	152
Empirical Example Used Throughout This Chapter.....	152
Running PROC MI (Step 2): Preparing the Data Set	153
Running PROC MI (Step 3): Variables.....	155
Running PROC MI (Step 4): Summarizing the Missing Data.....	158
Running PROC MI (Step 5): EM Algorithm	162
Running PROC MI (Step 6): Impute From (EM) Parameters	168
Running PROC MI (Step 7): MCMC (and Imputation)	169
Running PROC MI (Step 8): MCMC Diagnostics.....	170

Direct Analysis of EM (MLE) Covariance Matrix with PROC FACTOR, PROC REG	174
PROC FACTOR with an EM Covariance Matrix as Input.....	174
PROC REG with an EM Covariance Matrix as Input	175
Analysis of Single Data Set Imputed from EM Parameters with PROC CORR ALPHA	176
Analysis of Multiple-Imputed Data with PROC REG, PROC LOGISTIC, PROC MIXED	177
Analysis of Multiple-Imputed Data with PROC REG.....	177
PROC MIANALYZE Output for PROC REG	178
Proc Reg with Multiple Dependent Variables.....	181
Analysis of Multiple-Imputed Data with PROC LOGISTIC	183
Analysis of Multiple-Imputed Data with PROC MIXED.....	185
References.....	190

Section 3 Practical Issues in Missing Data Analysis

8 Practical Issues Relating to Analysis with Missing Data:

Avoiding and Troubleshooting Problems	193
Strategies for Making It Work: Know Your Analysis	194
Strategies for Making It Work: Know Your Data	194
Causes of Missingness	194
Auxiliary Variables	196
Bottom Line: Think FIML	197
Troubleshooting Problems	197
Disclaimer.....	197
Underlying Problem 1	198
Solution 1	198
Underlying Problem 2.....	198
Solution 2a	198
Solution 2b	198
Underlying Problem: Redundancies in Variable List (Matrix Not Positive Definite)	200
Solution1	203
Solution1b.....	204
Underlying Problem.....	205
First Conceptual Basis for This Missingness Pattern.....	208
Solution.....	208
Second Conceptual Basis for This Missingness Pattern	208
Solutions	208
Summary of Troubleshooting Symptoms, Causes, and Solutions	210
References.....	212

9 Dealing with the Problem of Having Too Many Variables in the Imputation Model	213
Think FIML	213
Imputing Whole Scales	214
Determining Whether a Scale Is Homogeneous or Heterogeneous	215
Decision Rules for These Scenarios	219
Decisions About Throwing Away Partial Data Versus Imputing at the Item Level	220
Issues Regarding Decision Rules	222
Splitting Variable Set for Multiple-Pass Multiple Imputation	223
A Solution That Makes Sense	224
Comments	228
References	228
10 Simulations with Missing Data	229
Who Should Read This Chapter?	229
Background	229
General Issues to Consider with Simulations	230
What Are the Goals of Your Simulation?	230
What Other Approaches Are Available to Achieve Your Goals?	230
What Should the Simulation Parameters Be?	231
What Should the Range of Parameter Values Be?	231
Monte Carlo Simulations	232
Start with a Population and Generate Samples	232
Degrading the Sample	233
Automation Strategies	237
Technical Issues to Consider in Monte Carlo Simulations	238
Non-Monte Carlo Simulation with the MGSEM Procedure	239
The Multiple Group SEM Procedure for MCAR Missingness: Overview	240
Examples of Good Uses of the MGSEM Procedure for MCAR Missingness	241
Overview of MGSEM Procedure for MAR/NMAR Missingness	241
Examples of Good Uses of the MGSEM Procedure for MAR/NMAR Missingness	244
What Simulations Cannot Be Addressed with the MGSEM Procedures?	248
Other Considerations with the MGSEM Procedures for MAR/NMAR Missingness	250
References	251

11 Using Modern Missing Data Methods with Auxiliary Variables to Mitigate the Effects of Attrition on Statistical Power.....	253
Effective Sample Size	254
An Artificial Data Demonstration of Improving Power	
Using a Missing Data Model with Auxiliary Variables	256
Details of MGSEM Procedure	256
Artificial Data Example for One Auxiliary Variable, 100 % of Eligible Subjects with Data	261
Artificial Data Demonstrations with More Realistic Attrition Patterns	262
Less than 100 % of Eligible Subjects with Data for the One Auxiliary Variable	262
Two Auxiliary Variables.....	264
Two Auxiliary Variables with Different Values for r_{YZ1} , r_{YZ2} , $\%Z_1$, and $\%Z_2$	264
Estimating N_{EFF} with One or Two Auxiliary Variables: The General Case	267
Automation Utility for Estimating N_{EFF} in All One Auxiliary Variable Scenarios.....	267
Automation Utility for Estimating N_{EFF} in All Two Auxiliary Variable Scenarios.....	269
Implications of N_{EFF} for Statistical Power Calculations.....	269
Loose Ends.....	271
What Happens When Pretest Covariates Are Included in the Model?	271
Multilevel Models.....	271
“Highly Inclusive” Versus “Selectively Inclusive” Models.....	272
What Other Factors May Affect the True N_{EFF} Benefit?	275
References.....	275

Section 4 Planned Missing Data Design

12 Planned Missing Data Designs I: The 3-Form Design	279
Who Should Read This Chapter?.....	279
Reasons for Not Using the 3-Form Design.....	279
Reasons for Using the 3-Form Design.....	280
The 3-Form Design: History, Layout, Design Advantages.....	280
Matrix Sampling: Early Designs.....	281
History of the 3-Form Design	281
Basic Layout of the 3-Form Design.....	282
Advantages of the 3-Form Design over Other Designs.....	282
Disadvantages of the 3-Form Design Compared to the 1-Form Design	283

The Disadvantage of the 3-Form Design Is Not Really a Disadvantage.....	289
3-Form Design: Other Design Elements and Issues	290
Item Order.....	290
The X Set	290
Variations of the 3-Form Design: A Family of Designs.....	291
Keeping Scale Items Within One Item Set Versus Splitting Them Across Item Sets	293
References.....	294
13 Planned Missing Data Design 2: Two-Method Measurement.....	295
Definition of Response Bias.....	296
The Bias-Correction Model	297
Benefits of the Bias-Correction Model	298
The Idea of the Benefit.....	298
How the Sample Size Benefit Works in Bias-Correction Model.....	301
Factors Affecting the N_{EFF} Benefit.....	302
Real Effects on Statistical Power	303
Potential Applications of Two-Method Measurement	307
Cigarette Smoking Research.....	307
Alcohol Research	308
Blood-Vessel Health.....	308
Measurement of Hypertension.....	308
Nutrition Research	309
Measuring Body Composition/Adiposity	309
Assessment of Physical Conditioning and Physical Activity	309
Survey Research.....	310
Retrospective Reports	310
Cost Ratio Issues.....	310
Calculating Cost Ratio and Estimating Benefits in Studies with Narrow Focus.....	311
Calculating Cost Ratio and Estimating Benefits in Studies with Broad Focus	312
The Full Bias-Correction Model.....	314
A Note on Estimation Bias.....	317
Assumptions.....	317
Assumption 1: The Expensive Measure is More Valid than the Cheap Measure.....	318
Assumption 2: The Model Will “Work” Once You Have Collected the Data.....	319
Individual Versus Group Level Focus of the Research.....	320
Alternative Model: The Auxiliary Variable Model	321
References.....	322