

Contents

PART I OVERVIEW AND HISTORICAL APPROACHES

1. Introduction	3
1.1. Missing Data	3
1.2. A Broad Taxonomy of Methods with Partially Missing Data	6
1.3. Missing-Data Patterns	7
1.4. Mechanisms That Lead to Missing Data	8
1.5. Univariate Samples with Missing Values	9
1.6. More Than One Variable, but Only One Subject to Nonresponse	13
1.7. Multivariate Missing Data	17
References	18
Problems	19
2. Missing Data in Experiments	21
2.1. Introduction	21
2.2. The Exact Least Squares Solution with Complete Data	22
2.3. The Correct Least Squares Analysis with Missing Data	24
2.4. Filling in Least Squares Estimates	25
2.4.1. Yates's Method	25
2.4.2. Using a Formula for the Missing Values	26
2.4.3. Iterating to Find the Missing Values	26
2.4.4. ANCOVA with Missing-Value Covariates	27
2.5. Bartlett's ANCOVA Method	27
2.5.1. Useful Properties of Bartlett's Method	27
2.5.2. Notation	28

2.5.3. The ANCOVA Estimates of Parameters and Missing Y Values	28
2.5.4. ANCOVA Estimates of the Residual Sums of Squares and the Variance–Covariance Matrix of $\hat{\beta}$	29
2.6. Least Squares Estimates of Missing Values by ANCOVA Using Only Complete-Data Methods	30
2.7. Correct Least Squares Estimates of Standard Errors and One Degree of Freedom Sums of Squares	32
2.8. Correct Least Squares Sums of Squares with More Than One Degree of Freedom	34
References	36
Problems	37
3. Quick Methods for Multivariate Data with Missing Values	39
3.1. Introduction	39
3.2. Complete-Case Analysis	40
3.3. Available-Case Methods	41
3.4. Filling in the Missing Values	43
3.4.1. Introduction	43
3.4.2. Imputing Unconditional Means	44
3.4.3. Imputing Conditional Means: Buck's Method	44
3.4.4. Other Approaches	47
References	48
Problems	48
4. Nonresponse in Sample Surveys	50
4.1. Introduction	50
4.2. Randomization Inference with Complete Data	50
4.3. Quasi-Randomization Inference for Data with Missing Values	53
4.4. Weighting Methods	55
4.4.1. Weighting Cell Estimators	55
4.4.2. Choice of Adjustment Cells	56
4.4.3. Other Weighting Adjustments	58
4.5. Imputation Procedures	60
4.5.1. Introduction	60
4.5.2. Mean Imputation	61
4.5.3. Hot Deck Imputation	62

4.6. Estimation of Sampling Variance in the Presence of Nonresponse	67
References	71
Problems	72

PART II LIKELIHOOD-BASED APPROACHES TO THE ANALYSIS OF MISSING DATA

5. Theory of Inference Based on the Likelihood Function	79
5.1. The Complete-Data Case	79
5.2. Inference Based on ML Estimates	84
5.2.1. Interval Estimation	84
5.2.2. Significance Levels for Null Values of θ	87
5.3. Likelihood-Based Estimation for Incomplete Data	88
5.4. Maximizing over the Parameters and the Missing Data	92
5.4.1. The Method	92
5.4.2. Background	93
5.4.3. Examples	93
References	95
Problems	95
6. Methods Based on Factoring the Likelihood, Ignoring the Missing-Data Mechanism	97
6.1. Introduction	97
6.2. Bivariate Normal Data with One Variable Subject to Nonresponse: ML Estimation	98
6.3. Bivariate Normal Data with One Variable Subject to Nonresponse: Precision of Estimation	102
6.3.1. Large-Sample Covariance Matrix	102
6.3.2. Small-Sample Inference for the Parameters	104
6.3.3. Numerical Illustration	105
6.4. Monotone Data with More Than Two Variables	107
6.5. Computation for Monotone Normal Data via the Sweep Operator	112
6.6. Factorizations for Special Nonmonotone Patterns	119
References	124
Problems	125

7. Maximum Likelihood for General Patterns of Missing Data: Introduction and Theory with Ignorable Nonresponse	127
7.1. Alternative Computational Strategies	127
7.2. The EM Algorithm: Background Material	129
7.3. The E Step and the M Step of EM	130
7.4. Theory of the EM Algorithm	134
7.5. The Missing Information	137
7.6. EM Theory for Exponential Families	138
References	139
Problems	140
8. Maximum Likelihood Estimation for Multivariate Normal Examples, Ignoring the Missing-Data Mechanism	142
8.1. Introduction	142
8.2. Estimating a Mean Vector and Covariance Matrix	142
8.2.1. The EM Algorithm for Incomplete Multivariate Normal Samples	142
8.2.2. Estimated Asymptotic Covariance Matrix of $(\theta - \hat{\theta})$ Based on the Information Matrix	145
8.3. Estimation with a Restricted Covariance Matrix	146
8.4. Multiple Linear Regression	152
8.4.1. Linear Regression with Missing Values Confined to the Dependent Variable	152
8.4.2. Linear Regression with Missing Values in the Predictor Variables	153
8.5. A General Repeated Measures Model with Missing Data	157
8.6. Time Series Models	162
8.6.1. Introduction	162
8.6.2. Autoregressive Models for Univariate Time Series with Missing Values	162
8.6.3. Kalman Filter Models	165
References	168
Problems	169
9. Models for Partially Classified Contingency Tables, Ignoring the Missing-Data Mechanism	171
9.1. Introduction	171

9.2. Factored Likelihoods for Monotone Multinomial Data	172
9.2.1. Introduction	172
9.2.2. ML Estimation for Monotone Patterns	173
9.2.3. Estimating the Precision of the ML Estimates	180
9.3. ML Estimation for Multinomial Samples with General Patterns of Missing Data	181
9.4. Loglinear Models for Partially Classified Contingency Tables	185
9.4.1. The Complete-Data Case	185
9.4.2. Loglinear Models for Partially Classified Tables	187
9.4.3. Goodness-of-Fit Tests for Partially Classified Data	192
References	193
Problems	193
10. Mixed Normal and Nonnormal Data with Missing Values, Ignoring the Missing-Data Mechanism	195
10.1. Introduction	195
10.2. The General Location Model	196
10.2.1. The Complete-Data Model and Parameter Estimates	196
10.2.2. ML Estimation with Missing Values	197
10.2.3. Details of the E-Step Calculations	200
10.3. Extensions of the General Location Model That Place Constraints on the Parameters	203
10.3.1. Introduction	203
10.3.2. Restricted Models for the Cell Means	203
10.3.3. Loglinear Models for the Cell Probabilities	203
10.3.4. Modifications to the Algorithm of Section 10.2.2	204
10.3.5. Restricted Models for the St. Louis Data	205
10.4. Relationships with Other EM Algorithms for Special Missing-Data Patterns	206
10.5. ML Algorithms for Robust Estimation	209
10.5.1. Introduction	209
10.5.2. Robust Estimation from a Univariate Sample	209
10.5.3. Robust Estimation of the Mean and Covariance Matrix: Complete Data	211

10.5.4. Robust Estimation of the Mean and Covariance Matrix from Data with Missing Values	212
10.5.5. Extensions of the Model	215
References	216
Problems	216
11. Nonignorable Missing-Data Models	218
11.1. Introduction	218
11.2. Likelihood Theory for Nonignorable Models	220
11.3. Models with Known Nonignorable Missing-Data Mechanisms: Grouped and Rounded Data	221
11.4. Stochastic Censoring Models	223
11.4.1. Maximum Likelihood Estimation for Stochastic Censoring Models	223
11.4.2. Sensitivity of ML Estimation to Normality	225
11.4.3. Heckman's Two-Step Fitting Method	229
11.5. A Predictive Bayesian Approach to Nonresponse Bias	230
11.6. Nonignorable Models for Categorical Data	235
References	241
Problems	242
12. The Model-Based Approach to Survey Nonresponse	244
12.1. Bayesian Theory with Complete Response	244
12.2. Bayesian Models for Survey Data with Nonresponse	247
12.3. Methods Based on Ignorable Nonresponse Models	250
12.4. Multiple Imputation	255
12.5. Nonignorable Nonresponse	259
12.6. Nonignorable Nonresponse with Follow-ups	262
References	264
Problems	265
Author Index	267
Subject Index	271