

MISSING DATA

A GENTLE INTRODUCTION

Patrick E. McKnight
Katherine M. McKnight
Souraya Sidani
Aurelio José Figueredo

Series Editor's Note by David A. Kenny



THE GUILFORD PRESS
New York London

Contents

1 • A Gentle Introduction to Missing Data	1
The Concept of Missing Data	2
The Prevalence of Missing Data	2
Why Data Might Be Missing	5
The Impact of Missing Data	6
A Cost–Benefit Approach to Missing Data	10
Missing Data—Not Just for Statisticians Anymore	13
Purpose of This Book	14
2 • Consequences of Missing Data	17
Three General Consequences of Missing Data	18
Consequences of Missing Data on Construct Validity	20
Consequences of Missing Data on Internal Validity	25
Consequences for Causal Generalization	35
Summary	39
3 • Classifying Missing Data	40
“The Silence That Betokens”	41
The Current Classification System:	
Mechanisms of Missing Data	42
Expanding the Classification System	52
Summary	62
4 • Preventing Missing Data by Design	65
Overall Study Design	66
Characteristics of the Target Population and the Sample	72
Data Collection Methods	75
Treatment Implementation	84

Data Entry Process	85
Summary	85

5 • Diagnostic Procedures 88

Traditional Diagnostics	89
Dummy Coding Missing Data	90
Numerical Diagnostic Procedures	92
Diagnostic Procedures Using Graphing	110
Summary	114

6 • The Selection of Data Analytic Procedures 116

Preliminary Steps	117
Decision Making	127
Summary	138

7 • Data Deletion Methods for Handling Missing Data 140

Data Sets	140
Complete Case Method	141
Available Case Method	147
Available Item Method	148
Individual Growth Curve Analysis	152
Multisample Analyses	156
Summary	158

8 • Data Augmentation Procedures 160

Model-Based Procedures	161
Markov Chain Monte Carlo	166
Adjustment Methods	168
Summary	172

9 • Single Imputation Procedures 173

Constant Replacement Methods	175
Random Value Imputation	181
Nonrandom Value Imputation: Single Condition	185
Nonrandom Value Imputation: Multiple Conditions	190
Summary	194

10 • Multiple Imputation 196

The MI Process	199
Summary	210

11 • Reporting Missing Data and Results	213
APA Task Force Recommendations	213
Missing Data and Study Stages	214
TFSI Recommendations and Missing Data	218
Reporting Format	221
Summary	223
12 • Epilogue	225
References	229
Author Index	237
Subject Index	241

Those missing data points are a random subset of the data. There is nothing systematic going on that makes some data more likely to be missing than others. Missing at Random, MAR, means there is a systematic relationship between the propensity of missing values and the observed data, but not the missing data. Whether an observation is missing has nothing to do with the missing values, but it does have to do with the values of an individual's observed variables. Missing Not at Random: There is a pattern in the missing data that affect your primary dependent variables. For example, lower-income participants are less likely to respond and thus affect your conclusions about income and likelihood to recommend. Missing not at random is your worst-case scenario. Proceed with caution. And here are seven things you can do about that missing data: Listwise Deletion: Delete all data from any participant with missing values. Missing Data Part 1: Overview, Traditional Methods. Richard Williams, University of Notre Dame, <https://www3.nd.edu/~rwilliam/>. Last revised January 17, 2015. Longitudinal Modeling with Randomly and Systematically Missing Data: A Simulation of Ad Hoc, Maximum Likelihood, and Multiple Imputation Techniques. Organizational Research Methods, Vol. 6 No. 3, July 2003 pp. 328-362. Patrick Royston's series of articles in volumes 4 and 5 of The Stata Journal on multiple imputation. Missing Data & How to Deal: An overview of missing data. Melissa Humphries Population Research Center. Goals. } Discuss ways to evaluate and understand missing data } Discuss common missing data methods } Know the advantages and disadvantages of common. General Steps for Analysis with Missing Data. } 1. Identify patterns/reasons for missing and recode correctly. } 2. Understand distribution of missing data } 3. Decide on best method of analysis. Step One: Understand your data.