

WILEY HANDBOOKS IN  
SURVEY METHODOLOGY



HANDBOOK OF  
STATISTICAL  
DATA EDITING  
— AND —  
IMPUTATION

Ton de Waal  
Jeroen Pannekoek  
Sander Scholtus

 WILEY

# Contents

## PREFACE

IX



## 1 INTRODUCTION TO STATISTICAL DATA EDITING AND IMPUTATION

1

- 1.1 Introduction, 1
- 1.2 Statistical Data Editing and Imputation in the Statistical Process, 4
- 1.3 Data, Errors, Missing Data, and Edits, 6
- 1.4 Basic Methods for Statistical Data Editing and Imputation, 13
- 1.5 An Edit and Imputation Strategy, 17  
References, 21



## 2 METHODS FOR DEDUCTIVE CORRECTION

23

- 2.1 Introduction, 23
- 2.2 Theory and Applications, 24
- 2.3 Examples, 27
- 2.4 Summary, 55  
References, 55



## 3 AUTOMATIC EDITING OF CONTINUOUS DATA

57

- 3.1 Introduction, 57
- 3.2 Automatic Error Localization of Random Errors, 59
- 3.3 Aspects of the Fellegi–Holt Paradigm, 63
- 3.4 Algorithms Based on the Fellegi–Holt Paradigm, 65
- 3.5 Summary, 101
- 3.A Appendix: Chernikova’s Algorithm, 103  
References, 104

**4** **AUTOMATIC EDITING: EXTENSIONS TO CATEGORICAL DATA** **111**

- 4.1 Introduction, 111
- 4.2 The Error Localization Problem for Mixed Data, 112
- 4.3 The Fellegi–Holt Approach, 115
- 4.4 A Branch-and-Bound Algorithm for Automatic Editing of Mixed Data, 129
- 4.5 The Nearest-Neighbor Imputation Methodology, 140  
References, 158

**5** **AUTOMATIC EDITING: EXTENSIONS TO INTEGER DATA** **161**

- 5.1 Introduction, 161
- 5.2 An Illustration of the Error Localization Problem for Integer Data, 162
- 5.3 Fourier–Motzkin Elimination in Integer Data, 163
- 5.4 Error Localization in Categorical, Continuous, and Integer Data, 172
- 5.5 A Heuristic Procedure, 182
- 5.6 Computational Results, 183
- 5.7 Discussion, 187  
References, 189

**6** **SELECTIVE EDITING** **191**

- 6.1 Introduction, 191
- 6.2 Historical Notes, 193
- 6.3 Micro-selection: The Score Function Approach, 195
- 6.4 Selection at the Macro-level, 208
- 6.5 Interactive Editing, 212
- 6.6 Summary and Conclusions, 217  
References, 219

**7** **IMPUTATION** **223**

- 7.1 Introduction, 223
- 7.2 General Issues in Applying Imputation Methods, 226
- 7.3 Regression Imputation, 230
- 7.4 Ratio Imputation, 244
- 7.5 (Group) Mean Imputation, 246
- 7.6 Hot Deck Donor Imputation, 249

- 7.7 A General Imputation Model, 255
- 7.8 Imputation of Longitudinal Data, 261
- 7.9 Approaches to Variance Estimation with Imputed Data, 264
- 7.10 Fractional Imputation, 271
  - References, 272

## **8 MULTIVARIATE IMPUTATION** **277**

- 8.1 Introduction, 277
- 8.2 Multivariate Imputation Models, 280
- 8.3 Maximum Likelihood Estimation in the Presence of Missing Data, 285
- 8.4 Example: The Public Libraries, 295
  - References, 297

## **9 IMPUTATION UNDER EDIT CONSTRAINTS** **299**

- 9.1 Introduction, 299
- 9.2 Deductive Imputation, 301
- 9.3 The Ratio Hot Deck Method, 311
- 9.4 Imputing from a Dirichlet Distribution, 313
- 9.5 Imputing from a Singular Normal Distribution, 318
- 9.6 An Imputation Approach Based on Fourier–Motzkin Elimination, 334
- 9.7 A Sequential Regression Approach, 338
- 9.8 Calibrated Imputation of Numerical Data Under Linear Edit Restrictions, 343
- 9.9 Calibrated Hot Deck Imputation Subject to Edit Restrictions, 349
  - References, 358

## **10 ADJUSTMENT OF IMPUTED DATA** **361**

- 10.1 Introduction, 361
- 10.2 Adjustment of Numerical Variables, 362
- 10.3 Adjustment of Mixed Continuous and Categorical Data, 377
  - References, 389

**11 PRACTICAL APPLICATIONS 391**

11.1 Introduction, 391

11.2 Automatic Editing of Environmental Costs, 391

11.3 The EUREDIT Project: An Evaluation Study, 400

11.4 Selective Editing in the Dutch Agricultural Census, 420  
References, 426

**INDEX 429**