The Survey Statistician, 2023, Vol. 87, 35-36

Book and Software Review

Multiple Imputation of Missing data in Practice Basic Theory and Analysis Strategies

Maria Michela Dickson

Department of Economics and management, University of Trento
mariamichela.dickson@unitn.it

Abstract

The book *Multiple Imputation of Missing Data in Practice - Basic Theory and Analysis Strategies* is a recent piece of literature on multiple imputation methods. It constitutes a very complete contribution in the field, providing a methodological and applied point of view on the topic and facing several real situations in which imputation is non-negligible. The authors deal with a variety of complex problems, such as, among others, how to apply multiple imputation with survival, longitudinal and survey data, without losing rigor and making this book useful both for scientists and practitioners.

Keywords: multiple imputation; missing data; complex surveys; imputed datasets.

Yulei He, Guangyu Zhang and Chiu-Hsieh Hsu released in 2022 the book *Multiple Imputation of Missing Data in Practice - Basic Theory and Analysis Strategies*, published by CRC Press. Since the beginning of its appearance in literature, multiple imputation techniques have been a topic of extreme interest and utility, experimenting in the last decades with a fast evolution due to the availability of new (and big) data sources that pose the attention to new challenges in missing data problems. This book fits into this context and covers a wide number of topics in the field of multiple imputation, maintaining a good balance between statistical methodology and its link with the real world. In fact, the approach of the book is really applied: every topic is presented, explained and contextualized through real examples and simulations, giving the chance, even for beginners, to approach the matter.

The book is designed in three parts, driving the reader through a path that includes the definition of the matter, the presentation of the main methods and their contextualization in several practical situations.

Following this road map, the first part is structured as follows: Chapter 1 consists of an introduction, Chapter 2 provides necessary statistical background and Chapter 3 introduces the basics in multiple imputation analysis.

The central part of the book discusses multiple imputation discriminating between techniques for univariate missing data and for multivariate missing data. In the first case, the authors define parametric methods (Chapter 4) and robust methods (Chapter 5), while in the second they present the joint modeling approach (Chapter 6) and the fully conditional specification approach (Chapter 7).

In the third and last part of the book He, Zhang and Hsu make a major contribution, proposing a very interesting discussion about the application of multiple imputation in real and complicated settings, such as for survival data analysis (Chapter 8), longitudinal data (Chapter 9), survey data (Chapter 10), measurement error (Chapter 11) and nonignorable missing data (Chapter 13). Chapter 12 and Chapter 14 deserve a special mention, due to the authors effort in offering insights on imputation diagnostics and on the combination of multiply imputed datasets. Along the entire book, examples are proposed in order to have a clear and immediate idea about the use of multiple imputation in practice.

As said before, a focus is proposed in Chapter 10 on multiple imputation analysis for complex data surveys. Since the 70s multiple imputation became a relevant topic in handle nonresponses in large surveys. Even if it has experienced advances in several fields of research, survey data remain the natural outlet for multiple imputation techniques. Herein, modeling and analysis techniques are discussed, together with the organizations point of view on data editing, processing and release. Usually data coming from complex surveys are managed by institutions for public use and their perspective is, even if too often ignored, undoubtedly relevant. The authors proposed instead an explanation of the different steps with which survey statisticians have to face in the pre-release phase of survey imputed data, such as for example the comparison with historical data, the consistency edits and the check with survey weights, in order to appropriately run the multiple imputation techniques. Interesting is also the paragraph devoted to the examples from previous literature related to imputed datasets for external users, since it helps the reader in understanding the problem in real situations.

Concluding, the book *Multiple Imputation of Missing Data in Practice - Basic Theory and Analysis Strategies* may be considered a quality reference book for those who are interested in, or have to deal with, multiple imputation in several applied contexts.