Imputation Basics with R

Instructor:	Angelina Hammon
Date & Time:	Thursday, September 24, 2020, full-day, time tba
	Friday, September 25, 2020, half-day, time tba
Place:	Online workshop
Registration:	To register, please send an email to courses.bagss@uni-bamberg.de by September 1, 2020. Registration is mandatory.
	The number of participants is limited to 16.

Short Outline

The course will be held as a mixture of previously provided material (such as recorded lectures to specific topics covered by the course & exercises) and following live sessions via Zoom where we will be able to interact directly. These live sessions will give the opportunity to discuss open questions and to work on the more practical parts of the workshop. There will also be some kind of discussion board so that you are able to exchange with other participants of the course.

This course aims to introduce the theory and application of Multiple Imputation (MI), which has become a very popular way for handling missing data. With the advent of MI algorithms implemented in statistical standard software, the method has become more accessible to data analysts. For didactic purposes, the course starts by introducing some naive ways of handling missing data, and we use the examination of their weaknesses to create an understanding of the framework of Multiple Imputation. The first part of this course will be of somewhat theoretical nature and will introduce the basic idea and fundamentals of MI. The remainder of the course will be more practical by addressing frequent problems like regression with missing data and the actual application of MI in statistical software. Since R offers the widest range of MI algorithms, I will present how R can be used to handle missing values appropriately using the R package *mice* (some prior R knowledge might be helpful, but is not necessary!). There will also be some hands-on exercises so that you will have the opportunity to practice the covered material. At the end of this course, you should have a basic understanding of MI and be able to generate your first own imputations using R.

The workshop mainly aims at doctoral students, regular and associated members of BAGSS and members of LIfBi. In case of any vacant places, it is open to anyone interested.

Syllabus

Day 1

- Problems arising from missing data
- Introduction to missing-data mechanisms & patterns
- Introduction to other missing-data methods such as complete case analysis and different single imputation strategies
- Multiple imputation theory and fundamentals
- How to generate multiple imputations
- How to analyse multiply imputed data
- Multiple imputation by chained equations (MICE)

Day 2

- Different multiple imputation algorithms
- Multiple imputation in R using the R package mice
- Example R code for generating multiple imputations in R
- Hand-on exercises to apply MI and create own imputations in R
- Steps for choosing predictor variables
- Addressing frequent data issues and problems

About the Trainer

Angelina Hammon is a research associate at the Chair of Statistics and Econometrics, University of Bamberg and an associated member of BAGSS. She has worked as a research associate at LIfBi and currently at the SOEP. Her research interests include missing data, multiple imputation, modeling of complex survey designs, Bayesian inference.