Institut zur Qualitätsentwicklung im Bildungswesen
Forschungsdatenzentrum
WISSENSCHAFTLICHE EINRICHTUNG DER LÄNDER
AN DER HUMBOLDT-UNIVERSITÄT ZU BERLIN E.V.

Workshop: Statistical methods for handling missing data in R

Lecturer: Dr. Takuya Yanagida

Date: Part 1: Thursday, 05.09.2024, 14:00 – 18:00

Part 2: Friday, 06.09.2024, 14:00 – 18:00

Abstract

The presence of missing data is a common issue in statistical data analysis that can result in biased parameter estimates and reduced statistical power. The amount of bias and reduction in power depends on the percentage of missing values in the dataset, the missing data mechanism, and the missing data handling method that should take into account correlates of incomplete variables (i.e., auxiliary variables).

Content

The goal of the workshop is to provide an overview of three modern methods for handling missing data: (1) full information maximum likelihood using the R package lavaan and semTools, (2) Bayesian estimation using the latent variable modeling program Blimp, and (3) multiple imputation using the R package mice.

The following contents are covered in the course:

- Missing Data Mechanism and Auxiliary Variables
  - MCAR, MAR, and MNAR mechanism
  - Diagnosing Missing Data Mechanisms
  - Identifying Correlates of Incomplete Variables and Correlates of Missingness
- Maximum Likelihood Estimation
  - Saturated Correlates Model
  - Two-Stage Maximum Likelihood Approach
- Bayesian estimation
  - Markov Chain Monte Carlo (MCMC) Estimation
  - Bayesian Approach for Missing Data Handling
- Multiple Imputation
  - Joint Modeling vs. Fully Conditional Specification
  - Incompatible Imputation Model

Prerequisite

Basic knowledge of regression analysis and practical experience with the statistics program R are required for participation.

Literature

Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.). The Guilford Press.

Keller, B. T., & Enders, C. K. (2023). *Blimp user's guide* (Version 3). Retrieved from www.appliedmissingdata.com/blimp

Lüdtke, O., Robitzsch, A., & West, S. G. (2020). Analysis of interactions and nonlinear effects with missing data: A factored regression modeling approach using maximum likelihood estimation. *Multivariate Behavioral Research*, *55*(3), 361–381. https://doi.org/10.1080/00273171.2019.1640104

Van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman & Hall/CRC.

Software

Blimp 3 (freely available at https://www.appliedmissingdata.com/blimp), R version 4.3.0 or later, RStudio version 2023.03.0 or later, and the latest versions of the following R packages are required to work through the examples and exercises: lavaan, mdmb, mice, misty, and semTools.