Draft Thesis Proposal

Hanne Oberman

2019-09-03

Introduction

Research goals...

develop a valid method to investigate the plausibility of multiply imputed data based on:

- models (imputation and non-response)
- data features (cross-tabs, point estimates, aggregate statistics, etc.)
- assumptions
- algorithmic convergence

Literature review

Van Buuren (2018)

Abayomi, Gelman, and Levy (2008)

Bartlett et al. (2015)

Li et al. (1991)

Rubin (1987)

Rubin (1996)

Vink (n.d.)

Approach

R Shiny...

ShinyMICE

- single measure to assess whether algorithm converged (based on Gelman-Rubin statistic?)
- data visualizations pre and post imputation (scatterplots, densities, cross-tabs)
- statistical evaluation of relations between variables pre and post imputation (chi square or t-tests)

References

Abayomi, Kobi, Andrew Gelman, and Marc Levy. 2008. "Diagnostics for Multivariate Imputations." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 57 (3): 273–91. https://doi.org/10.1111/j.1467-9876.2007.00613.x.

Bartlett, Jonathan W, Shaun R Seaman, Ian R White, and James R Carpenter. 2015. "Multiple Imputation of Covariates by Fully Conditional Specification: Accommodating the Substantive Model." *Statistical Methods in Medical Research* 24 (4): 462–87. https://doi.org/10.1177/0962280214521348.

Li, Kim-Hung, Xiao-Li Meng, Trivellore E Raghunathan, and Donald B Rubin. 1991. "Significance Levels from Repeated P-Values with Multiply-Imputed Data." *Statistica Sinica*, 65–92.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. Wiley Series in Probability and Mathematical Statistics Applied Probability and Statistics. New York, NY: Wiley.

———. 1996. "Multiple Imputation After 18+ Years." *Journal of the American Statistical Association* 91 (434): 473–89. https://doi.org/10.2307/2291635.

Van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. Chapman and Hall/CRC.

Vink, Gerko. n.d. "Towards a Standardized Evaluation of Multiple Imputation Routines."