Applied Data Science Thesis
Missing Data Reading List

Compiled by Kyle M. Lang

2022-04-26

The sources listed below offer an overall introduction to missing data analysis via multiple imputation (MI) with additional focus on prediction problems.

Books

van Buuren, S. (2012). *Flexible imputation of missing data* (2nd ed.). Boca Raton, FL: CRC Press. Retrieved from https://stefvanbuuren.name/fimd/

Introductions/Reviews

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*(1), 5–37.

Enders, C. K. (2013). Dealing with missing data in developmental research. *Child Development Perspectives*, *7*(1), 27–31.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, *60*, 549–576.

Lang, K. M., & Little, T. D. (2018). Principled missing data treatments. *Prevention Science*, *19*(3), 284–294.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2013). On the joys of missing data. *Journal of Pediatric Psychology*, 1–12. doi: 10.1093/jpepsy/jsto48

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473–489.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of state of the art. *Psychological Methods*, *7*(2), 147–177. doi: 10.1037//1082-989X.7.2.147

The FCS/MICE Algorithm

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*(1), 85–96.

van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivarite imputation. *Journal of Statistical Computation and Simulation*, *76*(12), 1049–1064.

Model-Specification Issues

Collins, L. M., Schafer, J. L., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Pscyhological Methods*, *6*(4), 330–351. doi: 10.1037//1082-989X.6.4.330

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399.

Prediction-Focused

García-Laencina, P. J., Sancho-Gómez, J.-L., & Figueiras-Vidal, A. R. (2010). Pattern classification with missing data: A review. *Neural Computing & Applications*, *19*(2), 263–282. doi: 10.1007/s00521-009-0295-6

Saar-Tsechansky, M., & Provost, F. (2007). Handling missing values when applying classification models. *Journal of Machine Learning Research*, *8*, 1217–1250.