R-miss-tastic

A resource website on missing values – Methods and references for managing missing data

How to handle missing values in practice?

How to handle missing values if we want to do estimation, inference or prediction ? There doesn't exist a unique method to answer to this question. We propose several workflows in both R and Python to compare the most common methods, including in particular multiple imputation (missMDA in R, IterativeImputer in Python), low rank methods (softImpute) and some recent methods implemented in Python using optimal transport (Sinkhorn imputation) and autoencoders (MIWAE).

R - How to …

... generate missing values? (PDF, Rmd, Resource R)

... estimate some parameters with missing values? (PDF, Rmd)

... impute missing values? (PDF, Rmd, Resource R)

Python - How to …

... generate missing values? (Interactive notebook)

... impute missing values? (Interactive notebook)

... predict with missing values? (Interactive notebook)

External notebooks

The following notebooks have been developed by external contributors and reviewed by the R-miss-tastic committee. We recommend these notebooks as they provide useful and complementary guidance on how to predict with missing values, on how to impute incomplete data, and on how to handle incomplete data in treatment effect estimation.

About

This website is sponsored by R Consortium and maintained by Julie Josse, Imke Mayer, Aude Sportisse, Nicholas Tierney and Nathalie Vialaneix.

Article on arXiv →
Read more →
FAQ →

Follow us!

Events
GitHub
Twitter paper bot
MissCausal