

Multiple imputation with Supervised Principal Component Regression

Using **supervised principal component regression** as an univariate **imputation** model in **MICE** is a great way to solve the **many-variables** imputation problem.

Large data with missing data

Respondent	X_1	X_2	X_3	X_{157}	$X_{(p-1)}$	X_p
Mihai	9	10	1	8	7	5
Anton	1	3	4	7	2	8
Leonie	5	3	5	5	6	3
Joran	8	1	3	7	8	7
Esther	9	4	8	2	8	8

Expert imputation model specification

- Remove constants and collinear variables.
- Evaluate connection between variables in the data.
- Apply a correlation-thresholding strategy.
- Extra: use total scores for item scales
- Extra: use single measurement in longitudinal data

Automatic imputation model specification

- MICE with Principal component regression
- MICE with Association-threshold Supervised Principal Component regression
- MICE with Principal Covariates regression
- MICE with Partial least square

Percent relative bias



Figure: The percent relative bias for the four PCR-based imputation methods is reported (Y-axis) as a function of the number of components used (X-axis).

Confidence interval coverage



Figure: The percent relative bias for the four PCR-based imputation methods is reported (Y-axis) as a function of the number of components used (X-axis).

Project summary and code



Play with the Shiny app



More research like this

