# Read me

This is a list of the files provided to show the proposal in the paper: "*Variable selection for hidden Markov models with continuous variables and missing data*" written by


- F. Pennoni (University of Milano-Bicocca, IT)
- F. Bartolucci (University of Perugia, IT)
- S. Pandolfi (University of Perugia, IT)


The code was written/evaluated in R with the following software versions: R version 4.2.2 (2022-10-31) Platform: a2.7 GHz Intel Core i7 quad-core Running under: macOS Monterey 12.6.6

This folder contains the following data and files that can be used to provide an example of data and results proposed in the paper.

#——

**Implemented functions**

The following functions have been implemented to estimate the hidden Markov model with missing data and perform model selection:

- lmbasic.cont.MISS.R —> estimate the basic HM model for continuous outcomes with intermittent missingness using the extended EM algorithm
- lmestContMISS.R —> Estimate the HM model for continuous outcomes with intermittent missingness
- regress_miss.R —> Fit a multiple linear regression model under the MAR assumption on the responses
- item_selection.R –> Perform item selection with the steps described in the article (tol by default is 10^-10)
- complk_cont_miss.R —> Compute complete log-likelihood of the basic HM model for continuous outcomes (internal use)
- forward_regress_miss.R —> Stepwise regression for Y with covariates X where ind is the index of the response variable on which it makes model selection
- drawHMBasicCont.R —> Draw samples of size n from a basic latent Markov model for continuous data
- compute_BIC.R —> Compute Bayesian information criterion (internal use)
- count_eq.R —> Required internal function
- functions.R —> Required internal function
- bootstrapMISS.R —> Perform non-parametric bootstrap procedure in order to compute standard errors of model parameters
- lmestDecoding.R —> Perform local decoding

#—

**Application**

- dt1.Rdata —> Data for example_trial.R

- example_trial.R —> Example file that loads the data "dt.RData" and applies the proposed procedure for model and variables selection; also apply the non-parametric bootstrap once the model is estimated. It produces the output file example_trial.Rdata
- ResultsApplication.R —> Print the results of the estimated model and some descriptive plots of the data

#——

**SimulationStudy**

- SimulatedScenario1.R —> Example file that reproduces one of the simulated scenarios presented in the paper with a proportion of missing values of 0.05
- TakeResSim.R —> Read the results of the simulations from file SimulatedScenario1.R