Assignment for
EP16: Missing Values in Clinical Research

Multiple Imputation

14 – 18 May, 2018

Deadline & submission:

Assignments need to be sumbitted via Canvas no later than 11 june 2018.

Task:

Impute and analyze the dataset you have been assigned and provide documentation that explains all steps performed and allows to reproduce them.

You may either provide

- R syntax, output and textual explanations **combined in one document**, for example using rmarkdown or knitr (submit the `.Rnw` or `.Rmd` syntax file and the resulting `.pdf` or `.html` file), or
- **two separate documents**, where
    1. one document is a report explaining all steps you have performed, supported by relevant plots and output (which could be a standard microsoft word document or a PDF), and
    2. one R syntax file (with ending `.R`; do NOT copy R syntax into a word document!)

Report:

The report needs to demonstrate

- that you are aware of all the characteristics of your data
- that you know what those characteristics imply for the imputation
- that you chose an appropriate imputation approach: you may use the packages we have used for imputation in complex settings (**JointAI**, **smcfcs** or **jomo**), but using **mice** for imputation is sufficient for this assignment
- that you are aware of pro's and con's of the imputation appoach you chose (mention limitations and, where possible, name approaches that could overcome these limitations could be overcome)
- that you did not just automatically/blindly accept the default settings
- that you properly checked that the imputation was successfull
- which adjustments you made to the imputation procedure to correct problems that have occured, and that these adjustments fixed the problem

Moreover

- Use plots and other R output (this you may also copy into the microsoft word document, if necessary) to support your decisions.
- If you run into issues that you cannot fix, document that as well. (e.g., if R throws any warning messages include them in the report and, if possible, explain why they can be ignored or that you could not finde a way to prevent them)
- Textual explanations can be brief, but should written in complete sentences to be readable.

  For example:
  "Histograms of the continuous variables (see figure below) show that all continuous covariates, except for `xxx`, are approximately normally distributed, hence, we chose `method1` to impute `xxx` and `method2` to impute the other continuous variables."
- The analysis model of interest is given for each dataset, but you need to decide yourself
  - if additional variables should be used as auxiliary variables
  - if variables need to be re-coded or re-scaled
- You do not need to check standard model assumptions in the final analysis model.

Syntax file

The syntax file needs to follow the same structure as the report, so that the steps explained in the report can be followed in the R syntax in the same order.
Use sections in the report and provide the same sections in comments in the R syntax file (if you use a separate `.R` syntax file).

For example, if you have a structure

1. first section
2. second section
   2.1. first subsection
   2.2. second subsection

your syntax file you could look like this:

```
################################################################################
# 1. first section                                                             #
################################################################################
...

################################################################################
# 2. second section                                                            #
################################################################################

# 2.1 first subsection ---------------------------------------------------------
3 + 3

# transform with log
log(3) + log(3)

# 2.2 second subsection --------------------------------------------------------
...
```