Assignment for
EP16: Missing Values in Clinical Research

Multiple Imputation

14 – 18 May, 2018

Data

The **MIdat12** data comprise data of 796 children and their mothers on vitamin D exposure of the mother during pregnancy and child bone health, measured by DXA scan, at 6 years of age. Maternal serum samples were taken in the third trimester of pregnancy.

The dataset contains the following variables:

variable	explanation
`ID`	subject identifier
`gender`	gender of the child
`season`	season of blood sampling
`gravidity`	number of times the mother has been pregnant (primigravida: this was the first pregnancy, multigravida: the mother had previous pregnancies)
`bdate`	child'd birth date
`weight`	child's total weight at DXA scan
`sun`	average sun light duration in minutes/day in the month before blood sampling
`leanfrac`	proportion of child's lean mass (lean mass/total mass; lean mass = total mass - fat mass)
`sun_birth`	average sun light duration in the month before birth in hours/day
`birthwgt`	birthweight in kg
`length`	child's length at time of DXA scan in meters
`vitD`	mother's serum vitamin D concentrations in 10 nmol/L
`BMC`	bone mineral content of the child in grams, determined by DXA scan
`sports`	does the child do sports regularily?
`ancestry`	child's ancestry

Analysis model of interest

The analysis model of interest is a linear regression with outcome `BMC` and covariates `vitD`, `ancestry`, `gender`, `leanfrac`, `sports`, `sun`, `season`, `length` and `weight`.

We assume that `vitD` has a non-linear (quadratic) effect.