Assignment for
EP16: Missing Values in Clinical Research

Multiple Imputation

14 – 18 May, 2018

Data

The **MIdat13** data comprise data of 792 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`	child's gender
`birth`	date of birth
`weight`	child's total weight in kg at DXA scan
`sports`	does the child play do sports regularily?
`BMC`	bone mineral content of the child in grams, determined by DXA scan
`sun_birth`	average sun light duration in the month before birth in hours/day
`season`	season of blood sampling
`vitD`	mother's serum vitamin D concentrations in 10 nmol/L
`sun`	average sun light duration in hours/day in the month before blood sampling
`ethn`	child's ethnicity
`length`	child's length in cm at time of DXA scan
`parity`	number of pregnancies of more than 20 weeks the mother had (nulliparity: this was the first pregnancy, multiparity: mother had previous pregnancies)
`birthwgt`	birthweight (standard deviation score)
`leanfrac`	proportion of lean mass (lean mass/total mass; lean mass = total mass - fat mass)

Analysis model of interest

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

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