Assignment for
EP16: Missing Values in Clinical Research

Multiple Imputation

14 – 18 May, 2018

Data

The **MIdat33** data comprise 1950 observations of 542 mothers measured repeatedly before and during pregnancy. All women were scheduled to have their weight measured once each trimester and were asked for their pre-pregnancy weight and BMI.

The dataset contains the following variables:

variable	explanation
id	subject identifier
gage	gestational age at measurement (`gage = 0` refers to a measurement before pregnancy)
weight	maternal weight
preterm	was the baby born before 37 weeks of gestation (preterm) or later?
gender	child gender
visit_center	was intake performed at the study center? (0: no, 1: yes)
parity	number of pregnancies of more than 20 weeks the mother had (nulliparity: this was the first pregnancy, ¿=1 child: mother had previous pregnancies)
alc	alcohol consumption behaviour of the mother during pregnancy
educ	educational level of the mother
income	household income
gestbir	gestational age at birth
kcal	average daily kcal intake (calculated from food frequency questionnaire)
stress	self reported stress score $(0 - 5)$
inclusion	date of inclusion in the study
BMI	self reported maternal BMI before pregnancy
bd_mom	birth date of the mother
trimester	trimester of measurement

Analysis model of interest

The analysis model of interest is a linear mixed model for `weight` with random intercept and slope for `gage`. Covariates are `smoke`, `kcal`, `stress`, `preterm`, `parity`, `educ` and `income`.
`weight` is assumed to have a non-linear (quadratic) slope over time.