# Types of missing data

It is usual to define three kinds of missing data:

- missing completely at random (MCAR);
- missing at random (MAR);
- missing not at random (MNAR).

When we say data are **missing completely at random**, we mean that the missingness has nothing to do with the person being studied. For example, a questionnaire might be lost in the post, or a blood sample might be damaged in the lab. In CADET, sex might be MCAR. Of course, this is not truly random, but means that whether something is missing is not related to the subject of the missing data.

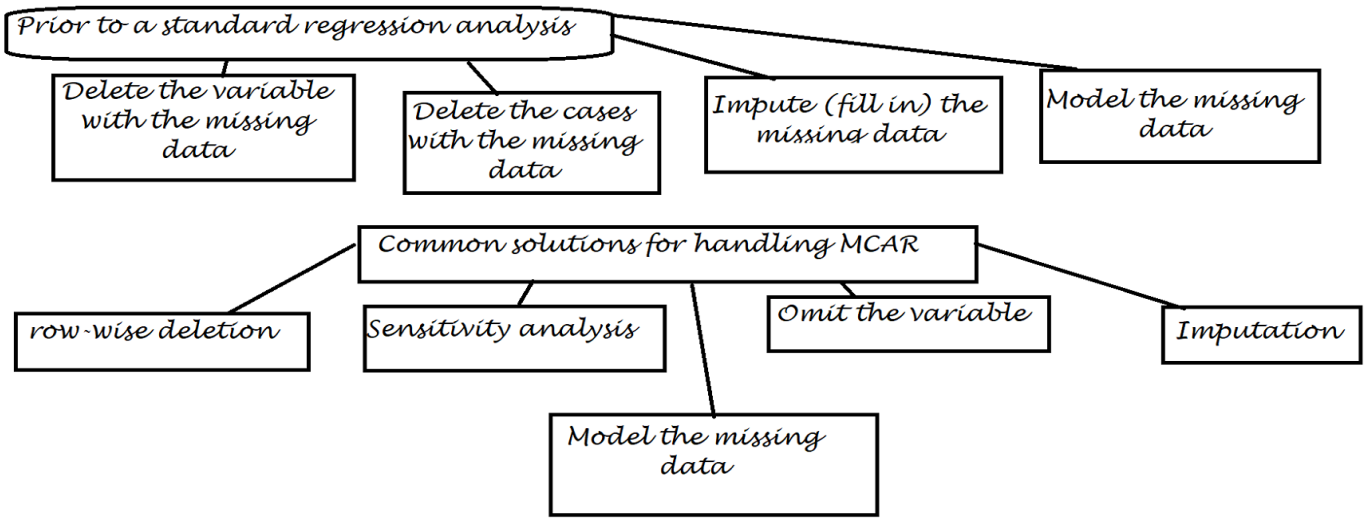
When we say data are **missing at random**, we mean that the missingness is to do with the person but can be predicted from other information about the person. It is not specifically related to the missing information. For example, if a child does not attend an educational assessment because the child is (genuinely) ill, this might be predictable from other data we have about the child's health, but it would not be related to what we would have measured had the child not been ill.

When data are **missing not at random**, the missingness is specifically related to what is missing, e.g. a person does not attend a drug test because the person took drugs the night before.

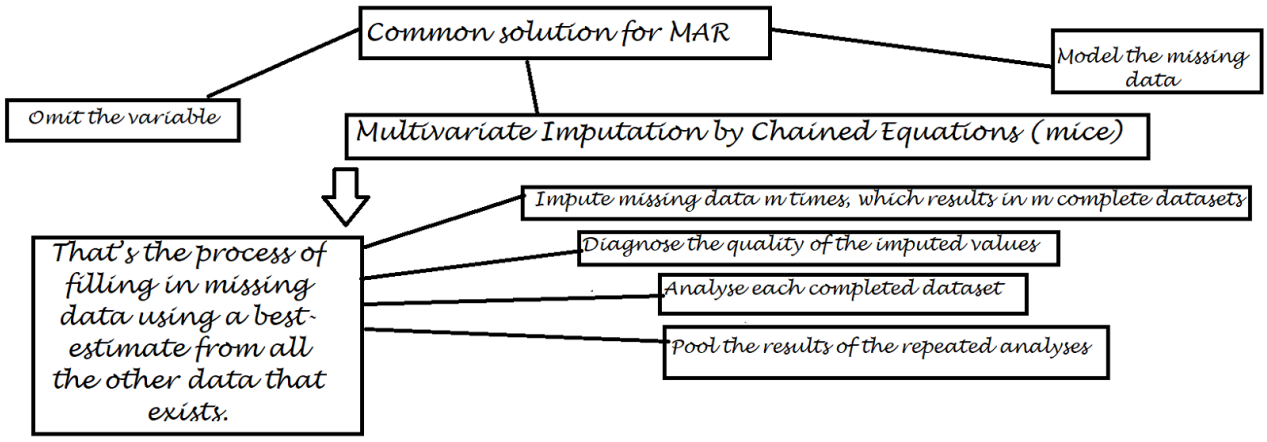
There are several strategies which can be applied:

- try to obtain the missing data;
- leave out incomplete cases and use only those for which all variables are available;
- replace missing data by a conservative estimate, e.g. the sample mean;
- try to estimate the missing data from the other data on the person.

# Handling missing data: MCAR



# Handling missing data: MAR



# Handling missing data: MNAR