

Incomplete Data Analysis

Vanda Inácio

University of Edinburgh



Semester 1, 2020/2021

A little more on checking MCAR versus MAR

- ↪ MCAR is the only missingness mechanism that, to a certain extent, yields testable propositions.
- ↪ Several tests have been proposed in the literature to check whether missingness is consistent with the MCAR assumption.
- ↪ These tests are not routinely used and their practical value is still not clear. Remember that we can never rule out the possibility of MNAR.
- ↪ Enders (2010, pp. 17–21) contains a good discussion of this topic.

A little more on checking MCAR versus MAR

- ↪ One popular and simple option is to perform a series of t -tests.
- ↪ This approach separates the missing and observed values on a particular variable and uses a t -test to examine group mean differences in the two groups induced by such splitting in the other variables in the dataset.
- ↪ The MCAR mechanism implies that such two groups should be similar on average.
- ↪ As a consequence, a non significant t -test (i.e., not rejecting the null hypothesis that the means of the two groups are equal) provides evidence that data *can be* MCAR .
- ↪ The main advantage for implementing the t -test approach is to identify (auxiliar) variables that we can later adjust for in the missing data handling procedure.
- ↪ Alternatively, I personally prefer to use density plots and boxplots to visualise the distributions of the two groups.

A little more on checking MCAR versus MAR

- ↪ As the number of variables grow, computing the t -test statistics can be cumbersome.
- ↪ Little (1988) proposed a multivariate version of the t -test that simultaneously evaluates mean differences on every variable in the data set. It is a global test of MCAR that applies to the entire dataset.
- ↪ For details see Little (1988) or Enders (2010, pp. 19–20). This can be carried out by the `LittleMCAR` function in the `BaylorEdPsych` R package.

How to prevent MNAR missingness?

- ↪ As we had already quoted, the ideal solution to the missing data problem would be to have none.
- ↪ Missing data prevention requires a careful experiment's design and a very careful execution as well.
- ↪ Most of the methods we will cover assume MAR data. However, we cannot be sure whether the data are really missing at random, or whether the missingness depends on unobserved variables or the missing data themselves.
- ↪ The idea is to start the study with a data collection strategy that will turn MNAR missingness into MAR missingness.
- ↪ This, so called inclusive analysis strategy, incorporates variables that are known to be correlated with the missing prone variables. Then, missing values will be more likely to be MAR than MNAR.
- ↪ These correlates variables are called auxiliary variables in the missing data literature.

How to prevent MNAR missingness?

- ↪ Note that auxiliary variables might not be of substantial interest in the sense that they would not have been included in the analysis had the data been complete.
- ↪ Theory and past research, as well as the MCAR tests/visualisation checks, can help to identify auxiliary variables.
- ↪ Note that the inclusion of auxiliary variables *per se* does not guarantee that the MAR assumption is satisfied, but it certainly improves the chances of it.
- ↪ For instance, it may be a strong assumption that nonresponse to an income question in a survey depends only on gender, race and education, but this is certainly a lot more plausible than assuming the probability of nonresponse is constant, or that it depends only on one of these variables.

Planned missing data designs

- ↪ In the first week, we have seen an example (nutrition study) where missing, instead of out of the researcher control, was 'induced' on purpose.
- ↪ The idea of planned missing data design is to intentionally generate MCAR or MAR data.
- ↪ For example, in a randomised study with two treatments (e.g., active treatment vs. placebo), each individual has a hypothetical score on both treatments, but participants only provide a response to their assigned treatment. The unobserved response to the other treatment is MCAR.
- ↪ A classic example of intentional MAR data occurs in selection designs where values on one variable determine whether respondents provide data on a second variable.
- ↪ For instance, universities often use exam(s)' marks as a selection tool for admissions, so first year marks are subsequently missing for students who scored below the admissions threshold.