Incomplete Data Analysis

Vanda Inácio

University of Edinburgh

Semester 1, 2020/2021

Missing data patterns

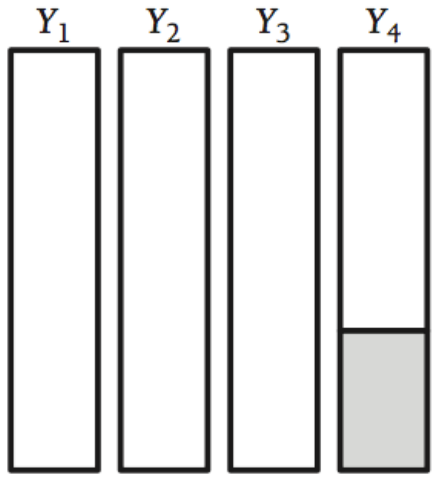
- A pattern of missing data describes the location of the missing values in a dataset.
- The missing data pattern describes the location of the 'holes' in the data but says nothing about why the data are missing.
- The pattern of missing values plays an important role with respect to the theoretical justification and the application of techniques for dealing with missing values.

Missing data patterns

- In a **univariate** pattern, data are missing only in one variable.
- For example, one could be interested in the relationship between the number of children living in a household and hourly wage.
- Suppose further that all households report the number of children but hourly wage is not observed for all households.

Missing data patterns

↪ A simple of extension of the previous bivariate example, as depicted in the figure below, includes the case where there are more than two variables, but only one variable is not completely observed.

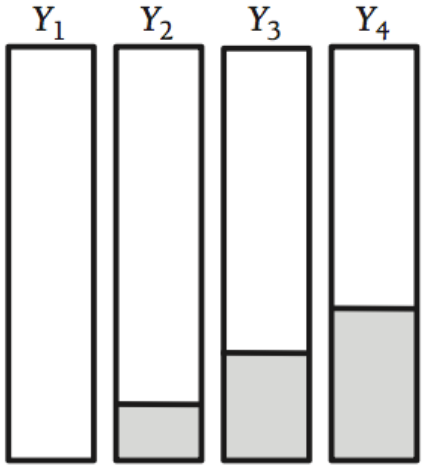


The shaded areas represent the location of the missing values in the data set. We are assuming a rectangular data matrix with rows representing subjects and columns representing variables. Figure from Enders, 2010, p. 4.

↪ The univariate pattern was one of the earliest missing data problems to receive attention in the literature.

Missing data patterns

↪ The figure below shows a **monotone** missing pattern.



The shaded areas represent the location of the missing values in the data set. Figure from Enders, 2010, p. 4.

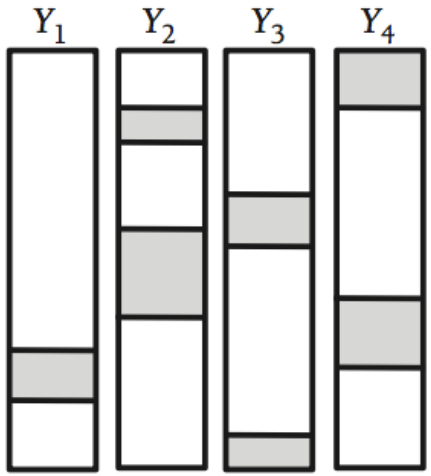
↪ A missing data pattern is called monotone if the dataset can be arranged by sorting rows and/or columns such that going from left to right if a missing value occurs in a row, all the following values in that row are missing as well.

Missing data patterns

- Visually, a monotone pattern resembles a staircase.
- A monotone missing data pattern is typically associated with a longitudinal study where participants drop out and never return.
- For example, consider a clinical trial for a new medication in which participants quit the study because they are having adverse reactions to the drug.

Missing data patterns

↪ An **arbitrary/general** pattern in which any set of variables may be missing for any subject is shown in the figure below.



The shaded areas represent the location of the missing values in the data set. Figure from Enders, 2010, p. 4.

↪ This pattern corresponds to the most common configuration of missing data and cannot be reduced to a univariate or monotone pattern.

Missing data patterns

- As a simple example, consider again the two variable example involving the number of children living in a household and hourly wage.

- The missing data pattern would be arbitrary if for some households the number of children is missing but hourly wage is observed and also, in contrast, for other households the number of children is missing but hourly wage is observed.