Types of Missing Data

- Missing At Random
- Missing Completely At Random
- Missing Not At Random

Types of Missing Data

Missing Completely At Random

- There are several reasons why data may be missing.
- They may be missing because equipment malfunctioned, the weather was terrible, people got sick, or the data were not entered correctly.
- Here the data are **missing completely at random (MCAR)**.

Types of Missing Data

Missing Completely At Random

- When we say that data are missing completely at random, we mean that the probability that an observation (X_i) is missing is unrelated to the value of X_i or to the value of any other variables.
- Thus data on family income would not be considered MCAR if people with low incomes were less likely to report their family income than people with higher incomes.

Types of Missing Data

Missing At Random

- Often data are not missing completely at random, but they may be classifiable as **missing at random (MAR)**.
  (MAR is not really a good name for this condition because most people would take it to be synonymous with MCAR, which it is not. However, the name has stuck.)

Types of Missing Data

Missing At Random

- For data to be missing completely at random, the probability that X_i is missing is unrelated to the value of X_i or other variables in the analysis.
- But the data can be considered as missing at random if the data meet the requirement that missing-ness does not depend on the value of X_i after controlling for another variable.

Types of Missing Data

- MCAR : Completely at Random throughout the data
- MAR : Randomly Occuring within variables, but more likely to happen with some variables than other.

Types of Missing Data

Missing Not at Random

- If data are not MCAR or MAR then they are classed as Missing Not at Random (MNAR).
- MNAR data is data that is missing for a specific reason (ie. the value of the variable that's missing is related to the reason it's missing)

Types of Missing Data

Missing Not at Random

- When we have data that are MNAR we have a problem.
- The only way to obtain an unbiased estimate of parameters is to model missingness.
- In other words we would need to write a model that accounts for the missing data.
- That model could then be incorporated into a more complex model for estimating missing values.