Data Analytical Methods

1. Missing values

The most common and simple solution to this problem is if any case has missing data for any of the attribute to be analyzed we can simply ignore it. This will give us a dataset which will not contain any missing value and we can then use any standard methods to process it further.

But this method has **a major drawback** which is deleting missing values sometimes might lead to ignoring a large section of the original sample.

**Missing Completely at Random**

Missingness does not depend on any other attribute.

It happenes arbitrarily

Ignore missing values in this case

Not Missing at Random

Data is not missing at random or **informatively missing** then it is termed as Not missing at random

Solution:

We need to write a model for missing data and then integrate it into a more complex model to estimate missing values.

Missing data in the training data set can reduce the power / fit of a model or can lead to a biased model because we have not analysed the behavior and relationship with other variables correctly. It can lead to wrong prediction or classification.