

Coursework B

Methods for dealing with missing values in datasets.

There are no easy methods to deal with missing values. In order to deal with missing values, the type of missing values should be analyzed. If the missing values were random, the fastest way to deal with missing values might be a Deletion method. By simply deleting incomplete records, the dataset is ready to be used. But these Deletion methods will lead to bias if the missing values are not random. Another method that is commonly used is Multiple Imputation. The method can be done by various ways such as Maximum likelihood estimation or Markov Chain Monte Carlo simulation. Multiple Imputation is good for complex missingness of values of a large dataset.

References

1. http://en.wikipedia.org/wiki/Missing_data .
2. http://www.sagepub.com/upm-data/45664_6.pdf .
3. http://www.biostatistikanseura.org/kenward_pmm_talk.pdf .

Mr. Wanchana Ekakkharanon
MSc Data Science