Abstract

The treatment of missing data has become a mandatory step for performing valid data analysis in most scientific research fields. In fact, researchers have found that dealing with missing data avoids misleading data analysis and improves the quality and power of the research results [1]. According to the authors in [2,3], the missing values in a data set could be missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR), a categorization that should be taken into consideration to deal with the problem of missing data. The number of observations, the types of variables, and the percentage of missing values in a data set are also important characteristics that should be contemplated before dealing with missing values. Understanding the missing data case helps the researchers to identify the imputation techniques that best handles the missing data problem. However, the development of procedures to impute categorical data is not significantly available as the procedures focused on continuous data imputation [1]. This study compares six different imputation methods to find the one that performs the most appropriate treatment for categorical data, type ordinal, in a breast cancer dataset.
References

Index Terms

Computer Science  Information Sciences

Keywords

MCAR; categorical data; ordinal data.