Dealing with missing data: A list

Description

Tweet

In this post â??missing dataâ?• does not mean absence of whole categories of data, which is a common enough problem, but missing data values within a given data set.

While this is a common problem in almost all spheres of research/evaluation it seems particularly common in more qualitative and participatory inquiry, where the same questions may not be asked of all participants/respondents. It is also likely to be a problem when data is extracted from documentary source produced by different parties e.g. project completion reports.

Some types of strategies (from Analytics Vidhya):

1. Deletion:
    1. Listwise deletion: Of all cases with missing data
    2. Pairwise deletion: : An analysis is carried out with all cases in which the variable of interest is present. The sub-set of cases used will vary according to the sub-set of variables which are the focus of each analysis.
2. Substitution
    1. Mean/ Mode/ Median Imputation: replacing the missing data for a given attribute by the mean or median (quantitative attribute) or mode (qualitative attribute) of all known values of that variable. Two variants:
        1. Generalized: Done for all cases
        2. Similar case: calculated separately for different sub-groups e.g. men versus women
    2. K Nearest Neighbour (KNN) imputation: The missing values of an attribute are imputed using those found in other cases with the most similar other attributes (where k = number of other attributes being examined).
    3. Prediction model: Using a sub-set of cases with no missing values, a model is developed that best predicts the presence of the attribute of interest. This is then applied to predict the missing values in the sub-set of cases with the missing values. Another variant, for continuous data:
        1. Regression Substitution: Using multiple-regression analysis to estimate a missing value.
3. Error estimation (tbc)

References (please help me extend this list)

Note: I would like this list to focus on easily usable references i.e. those not requiring substantial knowledge of statistics and/or the subject of missing data

- Gene Shackmanâ??s list of 23+ references on missing data (updated 12/11/2016)
- Wikipedia entry on Missing Data (2016)
- www.missingdata.org.uk (2016) London School of Hygiene and Tropical Medicine

- 7 Ways To Handle Missing Data (2015) Jeff Sauro
- Cochrane Collaboration (2011): General principles for dealing with missing data
- Statistical analysis with missing data. (2002) By Roderick J. A. Little, Donald B. Rubin

Category

1. Lists of sources re ...

Date
04/11/2026
Date Created
20/11/2016
Author
admin