

Methods for dealing with missing values in datasets

Many methods exist for dealing with missing values in datasets. We can regroup each method in different categories. The first category included methods which consist to delete missing data.

- Listwise deletion: this method removes the lines of data where the missing values are.
- Pairwise deletion: with this method, only specific missing values are deleted and not the line of data like in Listwise deletion.

The second category included methods which consist to substitute missing data by a new value dependent of the method chosen.

- Mean substitution: missing data are substituted by mean value of the field.
- Multiple regressions: this method uses all elements known to estimate the value of missing value.
- Maximum likelihood estimation: this method allows finding a value maximize the probability of the model.
- Approximate Bayesian bootstrap: this method computes the value by logistic regression.
- Multiple imputations: this method is iterative and generates several values for missing data and analyse what is the better value.

Bibliography:

- Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage Publications.
- http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html, September 2009.
- <http://faculty.chass.ncsu.edu/garson/PA765/missing.htm#concepts>, September 2009.
- http://www.psychwiki.com/wiki/Dealing_with_Missing_Data, September 2009.
- http://en.wikipedia.org/wiki/Listwise_deletion, September 2009.
- http://en.wikipedia.org/wiki/Maximum_likelihood, September 2009.
- http://en.wikipedia.org/wiki/Approximate_Bayesian_computation, September 2009.