Chapter 5

Conclusions

This study focuses on missing data treatment on cluster performed on Sudan Household survey. Initially, missing data mechanism and treatment rules are presented. Two-Step Cluster Analysis is be chosen over a wide range of approaches of statistical pattern-recognition available for clustering household health data.

Using the multiple imputation procedures, have analyzed patterns of missing Values and found that much information would likely be lost if simple listwise deletion were used. After an initial automatic run of multiple imputations, found that constraints were needed to keep imputed values within reasonable bounds. The run with constraints produced good values, and there was no immediate evidence that the FCS method did not converge. Using the "complete" dataset with multiply imputed values, fit a Multinomial Logistic Regression to the data and obtained pool regression estimation and also discovered that the final model fit, in fact, not have been possible be using listwise deletion on the original data.

Based on these results, I propose the following conclusions:

1. Linear imputation with rounding should never be used. It is usually inferior and never superior to linear imputation without rounding, which is computationally simpler.

2. For estimating proportions, the principal benefit from imputation is reduction in bias when data are MAR but not MCAR. Imputation methods have standard errors that are no smaller than those for complete case analysis are. In the MCAR condition, where bias is not an issue, there is no particular benefit to imputate.

3. For estimating proportions in the MAR condition, linear imputation without rounding is inferior to the logistic and discriminant methods, but may be acceptable if the proportion to be estimated is above .20.

4. For estimating the coefficient of a dummy variable with missing data, all the methods except for linear imputation with rounding are about equally good.

5. For estimating the coefficient of the covariate that has no missing data, linear imputation with rounding is about as good as the logistic or discriminant methods. Complete case analysis has better coverage than the other methods when p is near zero, but it always has substantially higher standard errors.

Recommendations

1- Single imputation methods like last observation carried forward and baseline observation carried forward should not be used of as the primary approach to the treatment of missing data unless the assumptions underlie them are scientifically justified.

 2- Weighted generalized estimating equations Methods should be more widely used in settings when missing at random can be well justified and a stable weight model can be determined, as a possibly useful alternative to parametric modeling. Data collection and information about all relevant treatments and key covariates should be recorded for all initial study participants, whether or not participants received the intervention specified in the protocol.

3- Statistical methods for handling missing data should be specified by clinical trial sponsors in study protocols, and their associated assumptions stated in a way that can be understood by clinicians.

4- When substantial missing data are be anticipated, auxiliary information should be collected that is believed to be associated with reasons for missing values and with the outcomes of interest.