ABSTRACT

COMPARISON OF METHOD HOT DECK, REGRESION AND MULTIPLE IMPUTATION IN HANDLING MISSING DATA

(DATA MINI SURVEY INDONESIA 2014 PASURUAN CITY-EAST JAVA)

The problem of data loss in value is more commonly referred with missing data. Missing data or missing value is information that is not available to a subject (case). One method of handling the problem is with imputation of missing data. Multiple imputation methods that can be used: Replaced missing data with a constant value, hot deck, regression, EM (Expectation maximization), and multiple imputation method. The purpose of this study is to analyzed, compared and defined the best imputation methods of missing data between hotdeck, regression and multiple imputation.

Type of researched is non-reactive researched which is a typed of researched for secondary data. The data used is data pasuruan urban respondents who participated KB taken from in monitoring data of fertile couples through a mini survey of Indonesia in 2014. The variables used were age and age of childbirth. Analysis of missing data used hotdeck methods, regression and multiple imputation. Compared the results with the original data imputation used T test, Pearson Correlation and RMSE test.

Conclusion, the best views of the results were not significant p value, the value of r near +1 and smallest RMSE value. Regression method produced p-value not significant data lost 5% and 10%, but this method has big value even approached +1 and RMSE were small. In addition to see the results of the analysis of the consistency of results seen also repeated values of p, r and RMSE value of the three methods. The results of analysis T test, Pearson Correlation, RMSE repeatability and consistency analysis showed that the regression method is the best method for the analysis of missing data.

Keywords: missing data, hotdeck, regression, multiple imputation