## INVESTIGATING THE EFFECTS OF IMPUTATION NUMBERS ON VARIANCE OF ESTIMATES

Nwakuya, M. T. PhD Department of Mathematics / Statistics University of Port Harcourt, P.M.B 5323 Port Harcourt, NIGERIA &

Nwabueze Joy C. PhD Department of Statistics, Michael Okpara University of Agriculture Umudike P.M.B. 7267, Abia State, NIGERIA

## ABSTRACT

The number of imputations to use in a multiple imputation analysis is always in question, knowing that missing data is a major problem in most research works. Most software's have five number of imputations as their default setup for any datasets with any percentage of missing values. Some researchers recommend that percentage of missing data should equate the number of imputations. But this is argued, as high number of missing values will attract very high number of imputations, which will take more computing time. Multiple imputation method imputes multiple values into a single missing point generating multiple complete data sets. In this paper we compared the variances of multiple regression estimates gotten from complete data sets imputed using 6 different imputation numbers, namely 50, 40, 30, 20, 15 and 5.The sample sizes investigated are 20000, 8000 and 30, each having 30%, 20% and 10% missing values. This work was analyzed in R software. Each of the complete datasets was analyzed and results pooled to give a single inference. The variances of the estimates were compared to each other to determine if they were significantly different from each other based on the imputation number used to impute the missing values and the percentage of missing value. The paired comparison was done in SPSS and the analysis showed that the variances were not significantly different from each other irrespective of the number of imputation used. But when it was compared based on the percentage of missing values, the variances were found to be significantly different.

**Keywords:** Imputation number, Missingness, Comparison, Variance and Multiple Imputation.