

Imputation Methods Effect on the Goodness of Fit of the Statistical Model

İsmail Yenilmez

Eskişehir Technical University

Abstract

In the analysis of financial and economic data, the problem of missing data can directly affect model fit and therefore inferences. Missing data should not be confused with censored, truncated, and rounded data types. The missing data situation causes different problems and therefore requires different applications than the solutions used for censoring, truncation and rounding problems. This study investigated the effects of some notable methods used for imputing missing data on the fit of the statistical model. For this purpose, frequently used linear interpolation, spline interpolation, and Stineman interpolation were utilized. In addition, traditional and basic methods such as imputation by mean and random sample, two other methods frequently used in the literature, were also conducted. For this purpose, synthetic data was produced, and the study was carried out by controlling some arguments such as missing data rate and data set size assignments. Missing data in the data sets produced according to the determined procedure and algorithm were imputed and modeling was made for the new data sets imputed according to the methods used. Models and observed-predicted values based on the models were examined. Considering statistical criteria, the findings are presented comparatively with tables and graphs.

Keywords: linear interpolation, missing data, spline interpolation, Stineman interpolation, synthetic data