Submission for the 2022 IAOS Prize for Young Statisticians

Imputation for Sub-Sampling in Indonesian National Socioeconomic Survey

Ms. Atika Nashirah Hasyyati
Statistician
BPS – Statistics Indonesia
www.bps.go.id

Abstract

Collecting consumption and expenditure data might result in some measurement problems, such as potential recall bias. In addition, the respondent burden is another issue as a consequence of the interview for hours. Consumption and expenditure data in Indonesia is collected through National Socioeconomic Survey (Susenas). Indonesia is a country with many factors that can influence how long an interview may take, especially collecting consumption and expenditure data, so deliberate sub-sampling and imputation need to be considered. The focus of this study is to look at the possibility of using sub-sampling of expenditure data and imputing that deliberately missing data using a standard method of missing data imputation (mice), a multilevel approach (jomo), and two machine learning approaches (missRanger and miceRanger). The results show that only mice with reasonable imputation results, in particular when breaking down by some categories. Although missRanger is the fastest, it has a large bias compared to the actual data.

Note: This paper is a part of my MSc dissertation (Department of Statistics, The University of Auckland). My supervisor was Prof. Thomas Lumley, BSc (Hons), MSc, PhD, FRSNZ.