ARTIC: Automatic Random error Treatment and Imputation for Categorical variables

Romina Filippini, Simona Toti [1]

Abstract

The identification and treatment of non-sampling errors is a fundamental step in the production of Official Statistics, aimed at ensuring higher accuracy and quality of the resulting estimates. Traditionally, the Editing and Imputation process represents one of the most time-consuming phases. In the context of Surveys, developing generalized tools that can be easily adapted to different questionnaires offers valuable support throughout the entire process. To this end, a standardized procedure in R, called ARTIC (Automatic Random error Treatment and Imputation for Categorical variables), has been developed for the detection and automatic correction of random errors for categorical variables, based on the Fellegi-Holt methodology. Fully implemented in R, the procedure is customizable, extendable, and easy to share.

Keywords: non-sampling errors, editing and imputation, categorical variables, Fellegi-Holt, R packages for E&I

[1] Filippini Romina (filippini@istat.it), ISTAT, Italy; Toti Simona (toti@istat.it), ISTAT, Italy. The views and opinions expressed are those of the author and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat