

Imputation of income data with generalized raking procedure and GB2 law: illustration with the 2009 Swiss SILC data

Eric Graf*

Institut de Statistique, Université de Neuchâtel, Pierre-à-Mazel 7, CH - 2000
Neuchâtel, eric.graf@unine.ch

Yves Tillé

Institut de Statistique, Université de Neuchâtel, Pierre-à-Mazel 7, CH - 2000
Neuchâtel, yves.tille@unine.ch

In sample surveys of households and persons, questions about income are often sensitive and thus subject to a higher non-response rate. Nevertheless, the household or personal incomes are among the important variables in surveys of this type. The distribution of such collected incomes is not normal, neither log-normal. Hypotheses of classical regression models to explain the income (or their log) aren't fulfilled. Imputations using such models modify the original and true distribution of the data. This is not suitable and may conduct the user to wrong interpretations of the results computed from data imputed in this way.

The generalized beta distribution of the second kind (GB2) is a distribution of four parameters. Empirical studies have shown that it adapts very well to income data; often better than other four parameters distributions. Results of the European project AMELI (2011) confirm this for the data collected by the EU-SILC (Survey on Income and Living Conditions).

The advantage of a parametric income distribution is that there exist explicit formulae for the inequality measures like the Laeken indicators as functions of the parameters.

We'll present a parametric method of imputation, based on the fit of a GB2 law on the income distribution by the use of suitably adjusted weights obtained by generalized raking. These weights correct better for the non-ignorable non-response that affects the variable of interest. We'll illustrate with the 2009 Swiss SILC data.

Key Words: Imputation, GB2, generalized raking, non-ignorable non-response, inequality measures, SILC.