Designing household samples in Brazil using the 2010 census enumeration area frame

Sâmela Batista Arantes*
IBGE, Rio de Janeiro, Brazil – samela.arantes@ibge.gov.br
Pedro Luis do Nascimento Silva
ENCE/IBGE, Rio de Janeiro, Brazil – pedronsilva@gmail.com

Complex sample designs are used in most of the household sample surveys conducted by the Brazilian Institute of Geography and Statistics. One of the key steps of sample design comprises the sample size determination, possibly for several variables of interest. This paper presents an approach for calculating sample sizes using design effects (deff) and intraclass correlation coefficients estimated for four different sampling strategies using data from the census 2010 enumeration area frame. All four sampling strategies comprise two stage cluster sampling, where in the first stage census tracts are the primary sampling units (PSUs) and the households are selected in the second stage. One of the sampling strategies considered selects census tracts (enumeration areas) using probability proportional to size (PPS) via Pareto sampling. This strategy is included here because it was used for selecting the Master Sample of the Integrated Household Survey of IBGE. For all sampling strategies, it is assumed that households are sampled within each census tract by simple random sampling without replacement (srswor). Estimation of intraclass correlations and design effects was carried out considering microdata for selected variables from both the short and long (sample) questionnaires of census 2010. The results indicate that PPS sampling designs are generally more efficient (i.e. lead to smaller sample sizes) because they take advantage of the correlation between target variables and the size of the PSUs. Pareto PPS sampling design (without replacement sampling of PSUs) appears to be more efficient when compared to PPS sampling with replacement of PSUs. R functions were developed for calculating the design effects and intraclass correlation coefficients, and can be used for other variables of interest.

Key Words: Household sample survey, design effect, sample size, Pareto PPS sampling.