Paradata on the 2010 Brazilian Census: Analysis of the field work supervision process

Luciano Tavares Duarte
IBGE, Rio de Janeiro, Brazil – luciano.duarte@ibge.gov.br

Denise Britz do Nascimento Silva
ENCE, Rio de Janeiro, Brazil – denise.silva@ibge.gov.br

Jose Andre de Moura Brito
ENCE, Rio de Janeiro, Brazil – jose.m.brito@ibge.gov.br

The relevance of a Census for many National Statistical Systems is unquestionable in terms of its thematic and territorial coverage. Nonetheless, its complexity leads to challenges for ensuring timeliness and quality of the results .

The aim of this work is to identify potential determinants of nonsampling errors associated to the data collection process of the 2010 Brazilian Census based.

We used data obtained from the field work monitoring system that provided information about divergences observed between data collected by enumerators and the supervisors. The latter carried out follow-up interviews in those households selected on the supervision/monitoring plan. In addition, human resources databases containing sociodemographic information of enumerators and supervisors, such as gender, age, educational level were also brought to bear to enhance the analysis. Moreover , in order to investigate associations between paradata and sociodemographic profiles of survey respondents, Census microdata was also used.

The statistical analysis employed generalized hierarchical models in which three nested hierarchical levels were taken into account, namely the supervisor, the enumerator and the respondent levels. The model response variable was defined as an occurrence of a discrepancy between information collected by enumerators and their supervisors for at least one of the following respondent characteristics: sex, age and literacy.

The results indicate that the different hierarchical levels investigated are relevant to decompose the data variability and hence have to be accounted in the analysis. We also found evidence of significant associations between the occurrence of divergences (on data collected by enumerators and supervisors) and sociodemographic characteristics of enumerators, supervisors and respondents. In addition, the results revealed notable regional differences regarding the levels divergences and the associations with explanatory variables considered in the study.

Keywords: Hierarchical Models ; Paradata; Census; Nonsampling Errors.