International Journal of Population Data Science

Journal Website: www.ijpds.org

Data Integration for Assessing Census Coverage

Diniz Da Silva, Andrea^{1*}

¹National School of Statistical Science, Brazilian Institute of Geography and Statistics

Objective

Since the 70's census, the Brazilian Institute of Geography and Statistics -IBGE has been conducting a post enumeration survey -PES to assess census coverage. In 2010 the survey was conducted in a sample of enumeration areas in each of the 27 states and matching was performed for data from Census and PES. One of the biggest improvements of the 2010 Brazilian Census was the incorporation of new methodologies and technologies. Use of handheld devices in the 2010 Census and PES facilitated automatic matching of PES to the Census.

Method

A matching system was designed aiming to find as much as possible the enumerated units by both Census and PES - the true matches. An accurate matching process was essential as the number of matches/unmatches had an effect on the coverage rates so that the levels of false positive (false matches) was strongly controlled during matching operation and the number of false negative (missed true matches) was minimised by successive steps in the matching system. The matching system comprised three stages: automatic, assisted and reconciliation. The automatic matching step was based on the probabilistic linkage theory and a probabilistic model was developed to identify true matches of persons and housing units from census and post enumeration survey data files. Scores were computed according to agreement and disagreement probabilities of selected variables in the pairs of records. The assisted step was held for all housing units and persons classified as unmatched or possible match at the end of automatic step. The procedures included revision of possible matches and matching "unmatched" pairs. This step was run through an application developed in house. The last step was the field reconciliation. Field team double checked the data collected on the unmatched housing units and persons from both Census and PES and searched for new matches.

Results

New true matches were found while carrying out field checks, especially in rural areas where the addressing system is not standardized. The matching system has been fully implemented immediately after the completion of data collection in each enumeration area. The performance of automatic step was impressive as Brazil is eight million squares kilometers country with huge regional differences and the automatic step was based in a single model for the whole country. Automatic matching resulted in 76% of the total of pairs, with regional differences under 10%, while assisted allowed for 20% and reconciliation 3% of the final pairs.

*Corresponding Author:

Email Address: andrea.diniz@yahoo.com (A. Diniz Da Silva)