2010 Brazilian Census Post Enumeration Survey

Diniz da Silva, Andrea
Brazilian Institute of Geography and Statistics, Brazil
Avenida Republica do Chile 500 – 10 andar
Rio de Janeiro, CEP 20.031-170, Brazil
E-mail: adiniz@ibge.gov.br

1. Introduction

Censuses are the main source of information on the population and its living condition at local level. Usually the local reality in both urban and rural areas depends on censuses results to be updated. A census produces essential information to support public policies and the decision-making concerning future investments by the private sector or by the government. Apart from providing information to identify areas for priority investments on health, education, housing and transport, census results are also used for growth monitoring; market analyses to set up shops, theatres, restaurants; and to analyse employment, among many other uses. In many countries the census results are also used for funds sharing, electoral boundary delimitation and assignment of the number of elected officials who will represent people in the country's legislature.

The national statistics offices make investments in order to reach census high response rates but it is known that complete coverage is not feasible and a part of the population will not be counted. The undercount does not usually occur uniformly across all geographical areas of the country and it is also affected by population characteristics such as age and sex.

The coverage of a census can be measured by different means. Comparisons with other data sources such as previous censuses, surveys or administrative sources are techniques used to evaluate census coverage, although demographic analysis and post enumeration surveys are the two widely applied methods to estimate census coverage. *Two thirds of the countries or territories conduct a Post Enumeration Survey to evaluate the coverage of the census, among them 75% to evaluate also some content errors. In Africa, Asia and South America, almost 80% of the countries undertake a PES, versus only 40% of the Oceanian countries*[1].

The objective of the 2010 Brazilian Census is to provide high quality population figures, however the continental dimensions of Brazil with its vast territorial extension, areas with no standardized addressing system and problems to access certain remote locations are serious issues to be dealt with in the census process. Under such conditions the census is subject to certain types of errors, which may affect the data collection coverage. As a result, dwellings and persons can be omitted. Nevertheless, such omissions do not invalidate the data and knowing the coverage quality is fundamental for its appropriate use.

Since the 1970 Census, IBGE has been conducting an evaluation of coverage of the census data collection. On the ground of the experience accumulated and considering the commitment to transparency regarding the quality of statistics produced by the Institute as well as the need to provide elements for the population estimates, IBGE decided to undertake a Post Enumeration Survey as part of the 2010 Census.

As in previous censuses, the fifth edition of the PES aims at providing elements to estimate the coverage of 2010 census data collection. The PES was conducted on a sample of enumeration areas of each of the 26 States and the Federal District. The data obtained were compared to the ones collected by the Census in the same areas. The idea was to provide information to estimate the coverage rate.

---

[1] DURR, J.M, 2010.

To ensure comparability of data between the Census and the PES as much as to guarantee independence among them, which is fundamental for the application of the estimation model chosen2, the following aspects were taken into consideration:

- Concepts and definitions: same wording, concepts and definitions used in the Census for the common questions;
- Operational Conditions: equipment, vehicles and additional financial resources provided under the same rules used for the Census;
- Technology: perform PES data collection using handheld device, same layout application, and same systems used in the Census for the transmission and monitoring data collection; and
- Independent PES team: national coordination of PES directly linked to the front-office of the Directorate of Surveys and state coordinators directly linked to heads of State Offices.

2. Scope

The survey was conducted in 26 States and in the Federal District and in both urban and rural areas. The PES did not target special EAs such as campsites, military bases and indigenous areas, institutions such as shelters, prisons, penal colonies, jails, asylums, orphanages, convents and hospitals. Also mobile home such as ships, boats, cars and similar were out of scope. Coverage rates were estimated in respect of occupied private dwellings and people living in there.

3. Sample

The sample was designed to ensure a 95% confidence level for a relative error of 0.2, for estimating the following rates at the State level:

- Omission rate of occupied private dwellings;
- Erroneous inclusion rate of occupied private dwellings;
- Omission rate of people living in occupied private dwellings; and
- Erroneous inclusion rate of people living in occupied private dwellings.

The sample size was 4,011 EAs, i.e. 1,3% of the total EAs in the 2010 Census operational database. The number of enumeration areas per State and Federal District is shown in table below.

---

[2] Further presented in the section 7.

Fig 1 – Number of urban and rural enumeration areas per State and Federal District in 2010 PES sample

State/Federal District	Enumeration Area in the PES sample		
	Total	Urban	Rural
Brasil	4011	2954	1057
Rondônia	75	44	31
Acre	42	26	16
Amazonas	179	123	56
Roraima	76	59	17
Pará	62	37	25
Amapa	92	77	15
Tocantins	83	55	28
Maranhão	152	74	78
Piauí	58	30	28
Ceará	185	128	57
Rio Grande do Norte	116	82	34
Paraíba	113	73	40
Pernambuco	235	164	71
Alagoas	116	75	41
Sergipe	162	100	62
Bahia	222	129	93
Minas Gerais	345	264	81
Espírito Santo	131	105	26
Rio de Janeiro	173	161	12
São Paulo	393	351	42
Paraná	76	55	21
Santa Catarina	172	136	36
Rio Grande do Sul	186	138	48
Mato Grosso do Sul	115	90	25
Mato Grosso	167	126	41
Goiás	88	68	20
Distrito Federal	197	184	13

A two-stage sampling scheme was used in the PES. At the first stage, the EAs were selected with probability proportional to size (PPS), in an attempt to account for the variability in the sizes of the EAs on the variance estimator and, therefore, to improve the accuracy of the estimates. At the second stage, 10% of the private dwellings were randomly selected through an algorithm running on the handheld device. The sizes of the EAs were estimated using the number of private dwellings provided by the cartographic operation prior to the 2010 Census.

Two questionnaires were used: a short and a long one. The short form was applied to 90% of the occupied private dwellings and the long one to 10%.

4. Planning

The 2010 Planning began in December 2008 when the project was prepared and the schedule of the research was set up. During 2009, issues relating to the survey sample, questionnaires, manuals, training model, methodologies and tools to perform the matching of census data with the one collected by the PES and methods for estimating the coverage rates of the 2010 census have been studied and developed. In 2010, the systems for implementing and monitoring the data collection and transmission were developed. In addition, a test of the PES was conducted after each of the two census rehearsals in order to test and improve all methods and processes planned.

The 2010 PES is part of the 2010 Census so its core issues have been submitted to the Census Committee. In weekly meetings the group, comprising experts from the directorates of Research, IT and Geosciences, discussed issues related to technical aspects of the 2010 Census. For technical matters relating specifically to the PES, was also formed a group with representatives from the front-office of the Researches Directorate, Coordination of Methods and Quality, Coordination of the Census apart of the Coordinator of the 2010 PES. Besides, operational issues linked to the Specific demands of the PES associated with the differentiated period of collection and the need for greater mobility of the supervisors, as the PES is a sample survey, were discussed by the 2010 Census Coordination of Planning and Budget.

The questionnaires were designed to collect the necessary data for comparing Census and PES and also to register the mobility of the population in between the Census and the PES data collection. The questions were selected among the ones asked in the Census short form, which was applied to all population. Since the PES sample is independent from the Census sample, to which the census long form is applied, questions asked only in the census long form were not considered.

In total, seven questions were asked on dwelling addressing and one question to identify if people moved in between the Census and the PES operation. To address the type and name of the dwelling street and the dwelling number, complements and reference information were registered. There were three questions asked about persons in the short form and fifteen questions in the long one. The short form included complete name of the responsible for the household and its partner, as well as the number of men and women living in there. The long form included information on complete name, sex and age of the out movers; relation to the responsible for the household, name, sex, month and year of birth and age, month and year when the person moved in, race and literacy.

The objective of the two rehearsals was to refine the processes and procedures to be adopted in the 2010 data collection. The operation was used to evaluate and improve the material, techniques and time allocated for training, as well as the quality of the cartography and the handheld device. Besides, assessment of the coverage levels and patterns in the EAs could be done once the operation comprised all the PES steps: data collection, matching and reconciliation.

In order to ensure the independence between the Census and the PES, the survey counted on staffs working exclusively in the PES project, i.e. an independent team conducted the PES. In each State, the PES team had a coordinator who had one or more assistants according to the number of EAs in the PES sample. Each enumerator was responsible for one EA and one supervisor followed the work in three to six EAs, according to the distance and accessibility of the EA.

The training started in September 2010 in order to allow enough time to the cascade strategy adopted and to be as close as possible to the data collection agenda. It was developed according to the 2010 Census methodology with focus on consolidating the PES concepts. The training was structured in two parts: self-study and presential activities. The self-study comprised reading the manuals and completing exercises found on the Exercise Notebook. The presential training was an important step as it offered the opportunity to deepen the theoretical content and check the trainees' understanding.

All training material was developed based on the material prepared for the 2010 Census in order to maintain not only the same concepts and definitions used, but also the same structure, language and layout. The video and manual were reproduced under the responsibility of the Centre for Documentation and Information Dissemination of IBGE and sent to State Units for distribution in local training.

A training cascade was implemented as follows:
- Regional Training: four poles (North, Northeast, Southest and South), 27 coordinators and 49 assistants trained;
- State Training: twenty six States and 1 Federal District and 1,179 supervisors trained; and
- Local Training: one thousand seven hundred localities, 3,877 enumerators trained.

5. Data Collection Process

The data collection process was conducted from 14 November 2010 to 7 February 2011, taking about two weeks to be completed in each EA. In order to keep proximity to the reference date of the 2010 Census and to take advantage of the recent mobilization of the population, the PES data collection started just after census data collection had finished in the municipality. In the capitals of the States and in a few other large towns where the census data collection took longer due to a lower availability of the respondents, the PES data collection started after census data collection had finished[3] in the EA. This strategy allowed the necessary time not only for the census data collection to be finished in time but also for its supervision and revision to be completed, in order to assure that census quality indicators were in accordance with the IBGE standards.

The data collection team was formed by 27 coordinators (one per State or Federal District), 49 assistants, 1,179 supervisors and 3,877 enumerators. Permanent staffs, appointed by the IBGE State Offices, occupied the State coordinator and the assistant positions. All of them attended the Census training and were previously in the Census project as training coordinator, cartography coordinator or other position related to preparatory activities, or had experience in household surveys. The supervisors and the enumerators were hired in a short-term contract through the 2010 Census public competition. The supervisors were either hired for the PES or transferred from the census team to work in the PES after the census task had been completed. All the enumerators were transferred from the census team to the PES once the data collection had finished. For both supervisor and enumerator the *sine qua non* condition to work in the PES was to work in different EA from the ones they had worked in the Census.

The data collection monitoring was made through the Data Collection Management Indicators System (SIGC), which displayed the number of EA by status (not started, in progress and finished), the number of dwelling and people counted. A central coordinator, a Survey Directorate staff, as well as the 27 State Coordinators, carried on the monitoring. The State Chiefs also followed up all the data collection trough the system.

6. Matching and Reconciliation

The coverage estimates will be based on the dual system estimation methodology (see section 7), so that the matching aimed to provide figures for units (dwellings and persons) enumerated based on:

- Census and PES;
- Census Only; and
- PES Only.

One of the biggest improvements of the 2010 Brazilian Census Post Enumeration Survey is the incorporation of new methodologies and technologies that have been developed or improved throughout the last decade. The use of handheld devices for the data collection, already experienced in the 2007 population count, was one of the successful innovations in the Census project, allowing improvement of quality and timeliness in the data collection process. The use of such technologies had an effect on the structure and format of the collected data, which was delivered immediately after its collection, facilitating the automatic matching of the census data and the post enumeration survey data.

A matching system was designed to find as much units that were enumerated by both Census and PES (the true matches) as possible. An accurate matching process was essential as the number of matches (or unmatches) affects the 2010 census coverage rates. The level of false matches was strongly controlled. The undesirable false positives (false matches) were emphasised in the training, supervision and revision of the matching job. The false negatives (missed true matches) were minimised by successive steps in the matching

---

[3] The State census team signed the conclusion of the data collection.

system. Both false positive and false negative were controlled and minimized in each step of the matching system through the use of exact and probabilistic matching, as described below.

To set the matching system up two main research lines were developed: one covering the mathematical models to perform the automatic and part of the assisted matching and another one with tools to implement the different steps of the matching system. The studies were jointly conducted by the PES coordination and the methodology coordination, both in the Surveys Directorate, and the developers from the IT Directorate.

The matching system comprises three stages: automatic matching, assisted matching and reconciliation.

The automatic matching was performed by running a program in R language, which was developed by the IBGE Survey Directorate. The program implements Fellegi-Sunter model, Jaro-Winkler distance and EM algorithm, already available in different R libraries, and also uses some other functions, which were necessary for the Brazilian PES project. The program was inserted in the production process managed by the IT Directorate by running it in batch so that the automatic matching was performed for each enumeration area as soon as the PES data collection had been finished.

The assisted matching included revision on suggested pairs and matching, through an application developed by the IBGE IT Directorate staffs. The operators were trained to evaluate the suggested pair and undo its match in case of detection of being a false match, and also to search for new true pairs. Fifteen operators and five supervisors were in charge of the job.

The last step of the matching system was the reconciliation stage. State coordinators were in charge of this task. They were advised to double-check the data collected on the unmatched dwellings and people by both Census and PES and search for new matches, as they know better the field and could detect some systematic errors in the enumeration that could not be done by the clerical staffs. Besides, new true matches could also be found while carrying on the field check, especially in the rural area where the addressing is not standardised.

The reconciliation phase was an opportunity to complete the matching process, however its main objective was to double-check the information on dwellings and people when any discrepancy was found in the data registered by either the Census or the PES. Such discrepancies could be, for example, a different number of residents in a dwelling in the census and in the PES data collection or a complete dwelling found neither in the Census nor in the PES. The job was done by the supervisors that worked in the data collection and staffs of other household surveys. A two stage supervision was implemented, the first stage in charge of the local branch chief and the second stage in charge of the state coordinator.

7. Estimation

The 2010 Census coverage estimation will be similar to the one developed in the 2000 Census PES, shown in SOUZA OLIVEIRA, 2003. The Dual System Estimate will be used to estimate the state level coverage and the national rate estimates will be obtained by summing up the estimated State rates.

Tables as in Figure 1 will be built for each State showing the results of the matching process for dwellings and persons.

Fig 2 – Table of counts of units (dwellings or persons)

		Census	
		Counted	Missed
PES	Counted	a	b
	Missed	c	d

Where:
  a is the number of units counted in both – Census and PES
  b is the number of units counted in the PES
  c is the number of units counted in the Census
  d is the number of units NOT counted, neither in the Census nor in the PES

Results of the matching process will be used to estimate the omissions and erroneous inclusions of the 2010 Census.

8. Final Remarks

The Post Enumeration Survey is part of the Brazilian Census since the 70's. In every new edition, IBGE implements changes in order to improve the process as a whole and, consequently, to obtain better estimates. In this 5th edition of the PES, the main improvement was the introduction of new technologies and methodologies. However, investment in organisation and qualified staffs still remains as a key factor for high quality results. Synchronised planning with the Census process to ensure the material and technical resources necessary to conduct the survey is the key for the success of the PES. Besides, a publicity campaign to encourage as many people as possible to answer the PES is essential, since the respondents have recently answered the Census and may not cooperate with the PES.

REFERENCES

DINIZ DA SILVA, A., et al. Inovações no Sistema de Pareamento de Domicílios e Pessoas para a Pesquisa de Avaliação da Cobertura da Coleta do Censo 2010. Rio de Janeiro: IBGE, Diretoria de Pesquisas, 2010.

DURR, J.M. The 2010 round of Population and Housing Censuses in the world, Caxambu: 19º Simpósio Nacional de Probabilidade e Estatística, 2010.

IBGE. Metodologia do Censo Demográfico 2000. IBGE, Rio de Janeiro, 2003.

ROMEO, O. Situação do Pareamento Automático após a PA Experimental de Rio Claro/SP. Rio de Janeiro: IBGE, Diretoria de Pesquisas, 2010.

SOUZA OLIVEIRA, L.C, et al. Censo 2000: Pesquisa de Avaliação da Cobertura da Coleta: apresentação dos resultados. Rio de Janeiro: IBGE, Coordenação Técnica do Censo Demográfico, 2003.

UNITED NATIONS Post Enumeration Surveys - Operational guidelines. Technical Report. Department of Econimic and Social Affairs. Statistics Division. New York, 2010.

UNITED NATIONS Principles and Recommendations for Population and Housing Censuses. Revision 2. Department of Econimic and Social Affairs. Statistics Division. New York, 2008.