Master Sampling Frames for Agricultural Surveys: Brazil overview

Flavio Pinto Bolliger
Marcos Paulo Soares de Freitas
Giuseppe de Abreu Antonaci
Maria Deolinda Borges Cabral

Expert Session: Master Sampling Frames for Agricultural and Rural Statistics
High Level Stakeholders Meeting on the Global Strategy: From Plan to Action
Rome, 3-5 December 2012
Outline

• Introduction

• Master Sampling Frame for the household survey system

• Master Sampling Frame for Agricultural Surveys

• Final remarks (research demand)
Introduction

Brazil is following GS recommendations:

- Building Master Frame for Agriculture
- Implementing a Integrated Ag Survey System:

  National System of Agricultural Establishment Sampling Surveys (SNPA Project)

Main surveys: National Survey on Agricultural Activity - PNAG (annual)
              National Survey on Agricultural Production - PNPA (quartely)
It had long been indicated:

It is proposed that the IBGE develop a master sample for agriculture (p.32) [...] It was noted above that the agricultural statistics program will have multiple purposes [...] It is therefore recommended an area sample selection with multiple stages [...] using composite measures of farm size for selection of primary sampling units, with probability proportional to size. The entire enumeration will then be concentrated within these area units, except for the additional sample (p.28-29) [...] Before the area sample selection [...] a special list of large and important production units will be developed and defined as an independent universe. They will be sampled separately. (p.32) (emphasis added)

Master Sampling Frame for the Integrate Household Survey System (SIPD)

Brazilian master frame: census enumeration areas

→ Master sample

- Recent experience: on going surveys after 2008
  - Household Expenditure Survey (POF) 2008/2009
  - Continuous National Household Survey (PNAD) 2010...
- The first problem was acceptance: resistance in several areas
- Other difficulty faced: size of the enumeration area
Master Sampling Frame for Agricultural Surveys
(under construction)

- Starting Information: 2006 Agricultural Census
- Adaptation to 2010 area map
- List update experience
- List frame x Area frame
- Enumeration areas with few cases
- Interaction between master samples frames
Starting Information: 2006 Agricultural Census

- 2006 Agricultural Census → 5.2 millions farms
- 2010 Population Census → 2.6 millions farms (only the ones with buildings)

Evidence of sub enumeration (difficulty in ensuring sufficient attention to a secondary aspect in a research focused in households and demographics)

- Agricultural Census:
  - farm address & household address (27%)
  - many agriculture-specific information
  - alternatives of scaling variables
  - generous database for analyzing sampling design
Adaptation to 2010 area map

For operational reasons, the AG Master Frame will be based on the 2010 geographic operational base, the same of HH Frame.

2006 base → 86,350 EAs with agricultural activity
2010 base → 105,221 EAs with agricultural activity (18,817 plus)
   (manly divided)

There alternatives for distribution of scaling variables were thought:
(a) Proportional of number of farms
(b) To distribute equally (in use for studies)
(c) Allocate using geographical coordinates (option for final construction)

This is a recurring problem!
Next year, 2013, IBGE starts a new big update of the operational base to be used for the Population Count (2015) and the 2015 Agricultural Census (in 2016).
List update experience

• Focus in large operations

• Update source: Central Register of Enterprises – CEMPRE 2010 and Annual Social Information Report – RAIS 2010 (employers)

→ 438,648 local units (ULs) classified in farming activities

• Match with 5,041,613 agricultural establishments with registration information (name of the producer, and name and address of the establishment), from 2006 Ag Census
List update experience

Procedures:

1. Standardization of the variables of both files for comparison;
2. Deterministic match considering the Tax Number – CNPJ (19,040 records relate)
3. Probabilistic match by UF and municipality, comparing the producer’s and the establishment’s names with the formal and the commercial names and the addresses of both files.

Results:
It was possible to identify a total of 50,871 agricultural establishments

Although advanced linkage techniques have been judiciously applied, the results achieved were quite frustrating due to problems of compatibility and quality of information found in different registers.

→ Need for major investment in the standardization of the information and the entry pattern between the administrative and survey records.
List frame x Area frame

• First approach to “on the list”:
  (a) formally constituted enterprise
  (b) 3 or more permanent employees
  (a) 200 ha or more in exploitation

  List ~300,000 farms
  Area ~5 millions farms

• Second approach to “on the list”:
  → specific criteria according to the farm type, in order to reduce list size, due to the difficulties in cadastral maintenance

  List ~115,000 farms

It’s under evaluation the inclusion of the establishments with semi-integration contracts for specialized activities, such as tobacco, pigs, poultry and others
List frame  (Second approach to “on the list”)

Farms according to selection criteria for composing sub-population for list sampling

<table>
<thead>
<tr>
<th>Type</th>
<th>Cut off</th>
<th>Farms</th>
<th>Item coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eggs</td>
<td>10,000 laying hens</td>
<td>1,310</td>
<td>80.0</td>
</tr>
<tr>
<td>Chicken</td>
<td>420,000 chickens sold</td>
<td>426</td>
<td>20.6</td>
</tr>
<tr>
<td>Pigs</td>
<td>40 sources or 880 pigs sold</td>
<td>10,090</td>
<td>34.3 &amp; 76.7</td>
</tr>
<tr>
<td>Milk</td>
<td>50 dairy cows</td>
<td>22,459</td>
<td>14.2</td>
</tr>
<tr>
<td>Cattle</td>
<td>1000 heads</td>
<td>20,365</td>
<td>28.3</td>
</tr>
<tr>
<td>Coffee</td>
<td>1,000,000 coffee bushes</td>
<td>214</td>
<td>8.6</td>
</tr>
<tr>
<td>Orange</td>
<td>300,000 trees or 200 ha of harvested area</td>
<td>425</td>
<td>15.3 &amp; 38.9</td>
</tr>
<tr>
<td>Cotton</td>
<td>300 ha of harvested area</td>
<td>313</td>
<td>84.4</td>
</tr>
<tr>
<td>Tobacco</td>
<td>100 ha of harvested area</td>
<td>376</td>
<td>12.9</td>
</tr>
<tr>
<td>Rice</td>
<td>100 ha of harvested area</td>
<td>3,593</td>
<td>42.5</td>
</tr>
<tr>
<td>Corn</td>
<td>300 ha of harvested area</td>
<td>1,633</td>
<td>10.7</td>
</tr>
<tr>
<td>Wheat</td>
<td>300 ha of harvested area</td>
<td>64</td>
<td>2.6</td>
</tr>
<tr>
<td>Soybean</td>
<td>300 ha of harvested area</td>
<td>10,245</td>
<td>57.6</td>
</tr>
<tr>
<td>Beans</td>
<td>100 ha of harvested area</td>
<td>3,147</td>
<td>16.2</td>
</tr>
<tr>
<td>Cassava</td>
<td>25 ha of harvested area</td>
<td>8,861</td>
<td>43.8</td>
</tr>
<tr>
<td>Wood</td>
<td>1000 ha of forest area planted</td>
<td>213</td>
<td>31.4</td>
</tr>
<tr>
<td>Sugar cane</td>
<td>500 ha of area harvested or CNPJ (corporation)</td>
<td>5,329</td>
<td>71.2</td>
</tr>
<tr>
<td>Gross production exceeding BRL 150,000 (~ USD 75,000)</td>
<td></td>
<td>25,117</td>
<td></td>
</tr>
<tr>
<td>Storage capacity over 1,200,000 tonnes</td>
<td></td>
<td>2,636</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>115,053</strong></td>
<td></td>
</tr>
</tbody>
</table>

Source: DPE/COAGRO; 2006 Agricultural Census
Enumeration area with few cases

Problem for area sample rotation and a rotation of selected establishments within each area

Solution → EA aggregation

The criteria for segregation planned involves:
• the vicinity
• the minimum number of farms
• homogeneity, taking into account the intraclass correlation coefficient of features investigated in the Ag Census

(Criteria equivalent was used for census dissemination)

<table>
<thead>
<tr>
<th>EA size</th>
<th>EA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 10 farms</td>
<td>28,374</td>
</tr>
<tr>
<td>11 to 50 farms</td>
<td>19,756</td>
</tr>
<tr>
<td>51 to 100 farms</td>
<td>18,610</td>
</tr>
<tr>
<td>101 to 200 farms</td>
<td>16,075</td>
</tr>
<tr>
<td>201 and more farms</td>
<td>3,535</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>86,350</strong></td>
</tr>
</tbody>
</table>

*Source: 2006 Agricultural Census*
Interaction between master samples frames

Ag Master Sample X HH Master Sample

The coincidence of enumeration areas:

If Allowed $\rightarrow$ easier list update & greater burden

If not allowed $\rightarrow$ increase the number EAs to be update
Thanks for your attention

Flavio Pinto Bolliger
flavio.bolliger@ibge.gov.br
www.ibge.gov.br