IBGE
Instituto Brasileiro de Geografia e Estatística

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)

KING, A. J. ; LINDQUIST, M. V. Programa de estatística agrícola para Fundação IBGE. [S.l.] : Agência Norte-Americana para o Desenvolvimento Internacional, **1969**. 108 f.

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 larges 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

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

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 agregation

The criteria for sgregastion 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)

Frequency of Enumeration Areas (EA) by and size-Brazil-2006

EA size	EA
1to 10 farms	28,374
11 to 50 farms	19,756
51 to 100 farms	18,610
101 to 200 farms	16,075
201 and more farms	3,535
Total	**86,350**

Source: 2006AgriculturalCensus

Interaction between master samples frames

Ag Master Sample X HH Master Sample

The coincidence of enumeration areas:

If Allowed → easier list update & greater burden

If not allowed → increase the number EAs to be update

Thanks for your attention

Flavio Pinto Bolliger

flavio.bolliger@ibge.gov.br

www.ibge.gov.br