Household Survey Sampling in the Annual Population Census

Sébastien Faivre

This article draws largely on the first part of Olivier Haag's paper "The tax source: a required frame for household and individual surveys", presented at the 2013 Statistics Canada Symposium.

1. Description of the Annual Census

At INSEE, the survey frame for household surveys is usually taken from the population Census (RP)[1].

Over the past decade, INSEE has moved from a policy of conducting an exhaustive French population census at semi-regular intervals (seven to nine years) to a new rotation approach to census activity. This was permanently introduced in January 2004.

Under the new approach, municipalities with a population of less than 10,000 (based on the 1999 RP), or "small municipalities", are surveyed exhaustively in rotation every five years: to do this, five groups in rotation were defined at random, within which these small municipalities were distributed.

Municipalities with a population of 10,000 or more, or "large municipalities", are subject to a more complex census survey every year, with an average rate of 8%. Addresses in these municipalities are divided at random into five rotation groups: each year, dwellings within a sample of addresses taken from the "active" rotation group are surveyed. Therefore, in total, over a five-year cycle, approximately 40% of dwellings in the municipality are surveyed.

2. Introduction of Octopusse, a New System for Drawing Samples [1]

This change in methodology, whilst no longer retaining the exhaustive nature of the survey, does bring new benefits, chief among which is the freshness of data collected: using this collection method, an exhaustive census is carried out every year in approximately 7,000 small municipalities, with a survey census conducted in approximately 900 large municipalities. For census data collectors and users alike, the increased freshness of the survey frame, achieved by minimising the temporal lag between the data collection date and the census date, is an undoubted advantage.

As a result, **the guiding principle involves selecting survey samples only from areas surveyed in the previous year.** The "freshness" principle offers a number of advantages:

- It makes it possible to **minimise the number of dwellings demolished or under construction in samples** as well as the number of "conversions" of primary residences into secondary residences (out-of-scope), which are often problems raised by data collectors in relation their working conditions, as well as the cause of cost overruns (as a larger sample must be drawn to compensate for this).
- Furthermore, the **quality of targeting in surveys of some population groups** (through over-representation of these groups in samples drawn) **is considerably improved** by the freshness of information available regarding dwellings in the survey frame (for example, this has made it possible to conduct the Mode of Care survey with parents of young children).
- Lastly, it enables **maximum survey separation** by ensuring that a single household form can only be selected once by Octopusse for "household" surveys during a five-year period.

However, the change in context linked to the ***introduction of this new census methodology*** since January 2004 led to a complete overhaul of the existing household survey sampling process. The "new census" led to radical changes in the methods used to construct survey samples, because the frame providing the list of units that can be used in samples every year, allowing the inclusion of "fresh" data, will, on the other hand, no longer be exhaustive over the whole territory.

3. The Introduction of Primary Survey Units in Sampling: ZAEs

1 In the following, we will use the term RP frame to refer to the survey frame.

However, territorial constraints in France associated with *one-on-one collection*, namely, the need to avoid spreading survey locations over too wide an area to minimise travel by data collectors, remains an issue with this new census method. It is therefore still necessary when sampling to apply a "master sample", which requires the introduction of primary units for collection.

A direct result of the freshness principle is the need to review the construction of primary units (Survey Zones (Zones d'Action Enquêteur - ZAEs)) from which samples of dwellings can be drawn. These zones must be constructed based on the following principles:

- as in the past, ZAEs must be fixed areas so that they can be assigned to a data collector who is relatively close and stable over time;
- however, the key new feature is that ZAEs must contain dwellings from the five rotation groups in order to be able to produce annual surveys using a sample drawn from the dwellings surveyed the previous year.

In order to minimise the scope of ZAEs, the following rules have been defined for their construction:

- ZAEs must respect regional boundaries,
- ZAEs are separated into ZAEs for large municipalities (ZAEGCs) and ZAEs for small municipalities (ZAEPCs),
- a large municipality on its own constitutes a ZAEGC,
- each ZAEPC rotation group must contain at least 300 primary residences,
- with the aim of minimising their scope.

4. Sampling [2]

INSEE household survey sampling therefore takes place in two phases:

- primary units are selected (i.e. ZAEs from the master sample)
- dwellings are then selected from this sample, while aiming to minimise weight dispersion for dwellings drawn using this method.

4.1. Selecting Primary Units from the Master Sample

For all ZAEs typically used by INSEE, we make the distinction between exhaustive and non-exhaustive ZAEs.

- There are 79 exhaustive ZAEs: 34 large municipalities and the arrondissements of Paris, Lyon and Marseille. These correspond to the 37 largest municipalities in France based on the number of primary residences. They account for 18% of all primary residences. In view of the size of INSEE samples, these exhaustive ZAEs are interpreted as those which must be selected to ensure the quality of estimation. Their probability of inclusion in the master sample is therefore 1.
- Lastly, ZAE sampling to construct the master sample should only be carried out on non-exhaustive ZAEs, of which there are 3,706. For reasons primarily linked to the cost of collection, the master sample includes 488 non-exhaustive ZAEs. These ZAEs are selected using stratified sampling by region and balanced across regional totals for the five rotation groups.

4.2. Selection of Dwellings in Primary Units

In each primary unit, the number of dwellings to be selected is calculated in such a way as to achieve equal weighting for dwellings in the sample. These units are then selected in the manner described above.

In a given year, the Octopusse annual survey frame is constructed using the following lists of dwellings from the most recent available annual census survey:

- In the ZAEGCs selected (exhaustive), the list of dwellings in the municipality surveyed at the last annual census survey (approximately 8% of all dwellings in the municipality).

- In the ZAEPCs selected, the list of dwellings in the exhaustive census of municipalities belonging to the surveyed portion of ZAEPCs (municipalities in the ZAEPC belonging to the rotation group impacted by the last available annual census survey).

Dwellings in ZAEs are drawn systematically after sorting dwellings within each ZAE. The maximum number of sorting criteria is five, which depend on the survey subject. These criteria may correspond to information on the dwelling (e.g. surface area, number of rooms, etc.), or characteristics of the household living there (e.g. household size, socio-professional category of the head of household, number of children aged under five, etc.) All information is taken from the population census.

5. Present Limitations of this Survey Frame

5.1. Fluctuations from One Year to the Next

The new census, especially the rotation of census zones over five-year periods, somewhat complicates the sampling process. As it was considered preferable only to use the most recent annual census survey due to the freshness of data, the municipal composition of the RP survey frame in rural areas may vary considerably from one year to the next, for example in terms of the level of urban development.

Alternatives were therefore found for random survey sampling (e.g. the employment survey uses an alternative frame taken from the tax source [3]), as well as annual surveys used to measure changes over time, which therefore require stability in terms of geographic typology for the areas covered (e.g. quality of life and security surveys, drawn from five census campaigns). For these surveys, a survey frame that is stable over time is thus preferable.

5.2. Impossibility of "Individual" Sampling

Population census files contain no named data. Furthermore, the Octopusse system is based on a survey frame for ordinary dwellings[2]. This system was not intended for sampling of individuals, for which some adjustments would need to be made.

Here also, it was necessary to find an alternative based on individuals' DGFP tax records for sampling in the PIAAC (Program for the International Assessment of Adult Competencies) for example.

5.3. The Absence of Certain Stratification Variables in Population Censuses

For very specific surveys, population census data may be missing for variables used to identify (or over-represent) the target population. For example, no variable is available regarding household income or wealth. In such cases, it is necessary to use external sources to construct the survey frame.

2 It does not cover individuals living in communities (e.g. university halls of residence, army barracks, etc.)

References

[1] Marc CHRISTINE, Sébastien FAIVRE, « Le projet OCTOPUSSE de nouvel Échantillon-Maître de l'Insee », *actes des journées de méthodologie statistique 23, 24 et 25 mars 2009*.

[2] Fabien GUGGEMOS, « Simulation de tirages de zones d'action enquêteurs pour les enquêtes-ménages de l'Insee », *actes des journées de méthodologie statistique 23, 24 et 25 mars 2009*

[3] Vincent LOONIS, « La construction du nouvel échantillon de l'Enquête Emploi en Continu à partir des fichiers de la Taxe d'Habitation », *actes des journées de méthodologie statistique 23, 24 et 25 mars 2009*.

Department of statistical methods
Version 1, published 10 October 2017