**Budapest 22-24 Aug. 2018**

**Victor Balsa Criado**

**Head of Unit of collection data in Households Surveys.**

**National Statistics Institute of Spain.**

**Improvements in CATI algorithm to reduce Non response. Treatment of inactive population**

The Spanish LFS is based in a mixed model CAPI – CATI.

Each dwelling in the sample remains for six consecutive quarters, after this period it is removed from the sample and replace by another dwelling from the same geographical section.

For each dwelling, first interview is made face to face and subsequent interviews are supposed to be conducted by CATI (only dwellings without a telephone number or people who expressly ask for it have their subsequent interviews face-to-face).

In order to assess the importance of CATI in the survey, we must know that every quarter the sample of Spain LFS consists of more than 80,000 households, and 70% of which are finally surveyed.

The effective sample, therefore, is composed of more than 55,000 households surveyed every quarter of which more than 40,000 come from CATI.

The sample is homogeneously distributed throughout the 13 weeks of the quarter, therefore, every week appear in the CATI about 3500 new households that must be managed with order and an optimal algorithm, which at each moment must offer to the interviewer the household with higher probability to be resolved.

A perfect functioning of the algorithm is vital for the Data collection. With the big amount of households we manage, a malfunction of the algorithm implies an increase in the stock of households that makes data collection unmanageable.

The basic rules that regulate the algorithm are, in order from most to least priority:

Households with:

* immediate appointment
* occupied phone more than 5 minutes ago.
* direct appointment (appointment for an specific date day/hour)
* indirect appointment (appointment for an indeterminate date, for example,.. “Call Tuesdays between 12:oo and 13:00)
* last call more than a week ago
* no calls

There are some other rules that the algorithm takes into account:

* Different time zones in Canary Islands
* Change of shift (mornings/afternoons) for a Non contacted household
* Maximum number of daily calls
* Maximum number of calls during an appointment
* Etc,..

With this algorithm, it was observed that the first days of the week were more productive, and on Thursdays and Fridays, the number of surveys conducted was much lower.

Specifically, the amount of surveys was distributed:

Approximately in 2 and a half days of work a whole round of calls is completed

Therefore, another 2 and a half days remain to complete the remaining dwellings of current week in addition to the outstanding stock of previous weeks.

The fact that has motivated the modification of the algorithm is the study of the average number of calls necessary to get a survey.

Currently we need 4.1 calls, on average, to complete a survey.

But it is an easily verifiable fact that the number of calls required is very different depending on the type of household and the relationship of its members with the labor market.

Thus, the average number of calls needed to complete a survey can range from 1.3 for a couple of retired people to more than 10 for a couple of young active workers.

So, in order to reduce the number of non-contacts, a modification has been made in the algorithm that is based on the type of household. The classification is made depending on:

- Type 2. Households composed only of people over 65 (theoretically retired)

- Type 1. Households with people over 65 or under 65 inactive.

- Type 0. Rest of households

Current sample consists approximatively in 20% of type 1 households and 5% of type 2 households, the rest are “active” households.

With the Households of types 1 and 2 are, in theory, easier to obtain, and require few attempts to call. On the other hand, type 0 households require more call attempts on more different days.

With this distinction, we can force the algorithm to offer us difficult housing first, leaving behind the simplest households, which are released on Thursdays and Fridays each week.

With this measure, carried out during the 13 weeks of which the quarter is composed, the number of calls to households with active people has increased significantly.

Now, a home type 0, is called during all the days of the week (if it remains not contacted), with what the average number of days with contact attempt, increases and the final success increases proportionally.

As a consequence of this intervention in the algorithm, it has been obtained that the number of "active" households surveyed has been improved by 5%, although on the other hand, it is detected that the number of livelihoods (type 2) surveyed has decreased by 10% (approximate figures).

In addition, the summary graph of the number of interviews completed per day of the week has changed significantly.

In general, the rates of non-contacted households in CATI have been improved and also the composition of the final surveyed sample.

Next steps in the near future will be:

* Differentiate type 1 households.

Interviewers have perceived that households with inactive (non-retired) people do not behave the same for the survey, depending on whether the inactive are students, housework or disabled people, for example.

* Improve the software.

Currently the classification of the household is done in final files external to the software.