Economic and Social Council

Economic Commission for Europe

Conference of European Statisticians

Sixtieth plenary session
Paris, 6-8 June 2012
Item 3(a) of the provisional agenda
2010 round of censuses – innovations and lessons learned

A general approach to the importance and use of registers in the Spanish Census

Note by the National Statistical Institute of Spain

Summary

The paper presents the 2011 Spanish Census which was based on a mixed approach of a large sample survey of 12% of the population (3 million dwellings and 5.8 million people) and the use of administrative registers.

The paper explains the challenges and benefits of this kind of an approach from a statistical point of view. First of all, an in-depth study of the different registers and statistical sources must be carried out. Features such as accuracy, coverage and correspondence between statistical and administrative concepts should be analysed. Different strategies should be defined in order to link all this information. Furthermore, other aspects such as inconsistencies between different data sources and level of detail of the information disseminated should be considered.

This paper summarizes the main aspects of these points, stressing the importance of the strong and weak areas of this methodology, from the experience of the National Statistical Institute of Spain.

Please recycle

I. Registers in Spain

1. Focusing on the sources of the 2011 Spanish Census, two of the most important ones include: the population administrative register (PADRON) and the 2001 Population and Housing Census.

2. PADRON is an administrative register with its own legal regulation on how it must be managed. The operational part is based in a central database located in our national statistics office (INE) where, every month, all the municipalities that are responsible for managing the registration of their inhabitants, send the changes noted in their local registers. INE, in a joint effort with municipalities and other local authorities, works to improve PADRON's quality by de-registering duplicates, communicating changes of residence between municipalities in order to avoid duplicates, de-registering registers of people that have moved abroad, etc. One of the main limitations of PADRON is the lack of demographic variables in the register. Only very basic information like date of birth, place of birth and citizenship make up the population register[1].

3. Nevertheless, PADRON does present some problems regarding its use as a direct statistical data source. It is an administrative register, and registration therein gives people fundamental rights such as voting, access to health and education services. Therefore, INE cannot decide to register or de-register a person in PADRON based on only statistical criteria. This means, for example, that to de-register a person who has moved abroad without having informed the municipality could take several months, following all the steps that are established in legal regulations. For census purposes, the register requires some statistical treatment.

4. PADRON is an essential register for obtaining demographic features about the population. Although it does contain territorial information of where people are living, this information is not organized from a computing point of view. In other words, it is a register of people, not of dwellings. Because of this, the frame of buildings is taken as a starting point for the 2001 Census.

5. Since 2001, this dwelling frame has been updated using PADRON information with all the people that arrive or move to new dwellings. Thus, for the purposes of the 2011 Census, we have a frame that contains:

(a) Information from dwellings that existed in the 2001 Population and Housing Census;

(b) Information from dwellings that were occupied at least once between 2001 and 2011.

6. This dwelling frame must be updated in order to include all those dwellings that have been built since 2001 and have never been occupied since then. For this reason, and to collect all the geographical coordinates of the buildings, we carried out an almost exhaustive[2] 2011 Building Census.

7. Other important registers in the 2011 Census project are national identity cards and the resident card register (they contain important information for verification of inhabitant

---

[1] It also contains a topic related to educational attainment, but information is not very accurate.
[2] In 20% of enumeration areas, the information was obtained directly from our updated dwelling frame and coordinates of those dwellings were obtained from Cadaster information.

identification), vital statistic bulletins (they allow us to check information from births, deaths and marriages), tax agency and social security.

8. The next table sums up the registers used in the 2011 Census.

Table 1
Registers used in the 2011 Census

Name of the register	Responsible for administration
Population register	INE
Census 2001	INE
National identity cards and resident cards	Home Office
Cadaster: Dwelling register	Treasury Department
Tax administration	Treasury Department
Social Security	Ministry of Labour and Social Affairs
Vital statistic bulletins: Births, deaths and marriages	INE

II. Spanish methodology: the mixed approach

9. Censuses are the biggest challenge that official statistical offices have to deal with. In fact, they are not only a question of data collection but a logistic operation as well. To illustrate this, the 2001 Spanish Census budget was more than 200 million euros (today it would be more than 300 million) and it required more than 40.000 people to be involved as enumerators, supervisors, etc. More than 70.000 contracts were signed in order to make replacements when certain staff members decided not to continue with the job.

10. For the 2011 Census, for the first time, a European Regulation regarding the population and housing census exists. This regulation provides a description of different sources of data acceptable in the Census. We took the following into account when deciding which of these sources would best suit our census.

11. Population figure is one of the most remarkable result of a census. Furthermore, other types of information are also very important depending on the interest of different users: demographic and social variables, dwellings and buildings information and many others.

12. Regarding the first point, counting population, the population administrative register must be taken into account because of its characteristics described briefly above.

13. As mentioned before, PADRON should not be used without verification of its information. For this reason, PADRON data are linked with other administrative sources, for example Social Security and Tax Agency, to accumulate proofs of real residence in Spain. Using this approach, we can be sure that around 97% of PADRON´s records are in fact people who live in Spain. At the same time, we have some information (based on sex, place and date of birth, place of residence and citizenship) to characterize those records where it is not clear if residence is in Spain or not (for example, Spanish citizens aged 65 or older registered in PADRON but not registered in Social Security or Tax Agency).

14. So, INE decided to concentrate the efforts on people whose residence in Spain is in doubt rather than on those citizens who clearly reside in the country. We therefore do not need to carry out an exhaustive enumeration of people.

15.   There are also some administrative registers that can provide us with information on dwellings and buildings. As mentioned before, PADRON only contains information on occupied dwellings, and we need information on dwellings that have been built since 2001 and have never been occupied since them. Cadaster has such a different structure of data that the attempts for linking these two registers produced acceptable results only in 20% of the enumeration areas. For this reason, INE decided to conduct an exhaustive enumeration of buildings and dwellings.

16.   Taking into account all these considerations, the 2011 Spanish census is a combination of administrative registers and a large sample. It is based on three main pillars:

(a)   A pre-Census file as a result of a linkage process between different administrative registers. It will be used:

(i)   To decide which population will be counted (or not);

(ii)   To characterize those groups of records where it is not clear whether or not they should be counted;

(iii)   To be the initial frame of Building Census;

(iv)   To provide data about, at least, the basic population topics: sex, age, citizenship and place of birth;

(b)   A large survey (around 10% of population) to collect data about population and housing topics:

(i)   To estimate the size of those groups of records where it is not clear whether or not they should be counted;

(ii)   To provide information about variables that are not included in the pre-Census file;

(c)   An exhaustive enumeration of buildings:

(i)   To obtain a complete frame of buildings (a basic need for future other statistical surveys);

(ii)   To obtain geographical coordinates of buildings.

17.   Information has been collected through several channels: Internet, mail and CAPI. First of all, a letter was sent to occupied dwellings giving them the opportunity to answer by Internet. For those that do not use this option, a second letter is sent with a paper questionnaire to give both options: answer by Internet or send the questionnaire by mail. Finally, we visited those who do not answer, to collect the data with enumerators with hand-held computers.

III.   Linkage between registers

18.   Because of its quality, PADRON plays the major role, and information from other registers is linked to PADRON for verification purposes. PADRON contains identification data for each person, and this information is stored in three different fields: number of national identification document (DNI), number of passport and number of residence card. Personal data for each person in PADRON is contained in one of those three topics.

19.   PADRON contains information about personal identification number for each person. Although a rare occurrence, we sometimes find two or more people with the same or with a wrong personal identification number or even people without one at all. In order to solve this problem and compare information available in PADRON, a linkage between

information from PADRON and information from the Home Office should be made. The Home Office is the organization that assigns national identity cards, passports or residence cards to citizens.

20.   For a linkage to be made between these two sources, there has to be an exact match of personal identification, name, family name, sex, date and place of birth between the two sources. For those records where this match is not perfect (maybe information from one of the two sources is wrong), probabilistic algorithms with an internal threshold or control parameter that measure similarities among people are executed. For Spanish citizens, the percentage of personal identification number of records from PADRON linked with the Home Office databases is 99%. For foreigners, it is 80%.

21.   The same process is applied to other administrative sources in order to be sure that the identification numbers are of good quality in all the registers involved.

22.   Information from PADRON and those registers will be linked if there is a match between the identification data. For those records where this technique does not work, probabilistic methods are executed in order to increase the percentage of linkage.

IV.   How to obtain the population figure

23.   As mentioned, the pre-Census file contains records from PADRON linked with other administrative registers. In order to obtain the population figure, we will assign to each record a counting factor, the weight of each record in total population.

24.   First of all, the records in PADRON that have also been found in other registers should be counted. In this case, we give this record a counting factor of 1.

25.   Secondly, if the record has been linked with some of the register of deaths provided by the Civil Register, we assign to that register a counting factor of 0.

26.   Finally, if the record has not been linked with any of the administrative registers, we should check if that record is marked with a "warning"[3]. If this is the case, the register will have a provisional counting factor of "X", where X is a value greater than 0 that should be calculated using the results of survey.

27.   The direct way to calculate "X" is by estimating the size of "warning" records using the sample data. But it is impossible to do that because these records are defined by its administrative situation and this information, obviously, is not a part of the Census. So, an alternative approach is needed.

28.   The strategy is to define clusters based on social and demographic variables (for example, EU citizens aged over 65 living in mediterranean provinces). In such clusters there will be "warning" records and records with a counting factor equal to 1. Each "warning" record must be in one cluster. The total population included in clusters must not be too big, approximately around 5% of total population. Ideally, there should be as many clusters as we need but all of them with a small amount of population.

29.   Next, we can estimate the size of these clusters. The counting factor for every "warning" record included in a cluster will be the estimated size of the cluster minus the records included in the cluster with a counting factor equal to 1 divided by the number of

---

[3] This is a simplistic approximation of some of the different rules that assign a warning to a register:
- if a record belongs to a Spanish person older than 13 years who does not have a proper ID number;
- if a record has an identification number but it was not found in the police databases;
- if a record that belongs to a foreigner remains unchanged for a long period of time in PADRON.

initial "warning" records. These values could be greater than 1 if the cluster is under-represented in PADRON or less than 1 if it is over-represented.

30.   The aggregation of the counting factors will give us the population figure of a geographical area and the results of the sample will be calibrated to this weighted frame.

V.   Conclusions

31.   This strategy has the advantage of making the most of the information contained in registers and disseminating detailed information based on a relatively large sample.

32.   It reduces the costs, the staff involved and improves the quality of field work. It also presents a strategy to treat the differences between administrative concepts and statistical variables.

33.   Because only demographic variables of a sample are collected, geographical dissemination is reduced in comparison with previous censuses.

34.   Finally, of course, this methodology is completely new for Spain. Therefore, some unexpected difficulties must be faced.

––––––––––––––––