

EVALUATION OF THE CENSUS ERROR IN SPAIN*

Rodero-Cosano, ML¹, Salinas Pérez, JA², Carbonero-Ruz, M³

¹ *Department of Quantitative Methods Universidad Loyola Andalucía, Córdoba, Spain; mlrodero@uloyola.es*

² *Department of Quantitative Methods Universidad Loyola Andalucía, Sevilla, Spain; jsalinas@uloyola.es*

³ *Department of Quantitative Methods Universidad Loyola Andalucía, Córdoba, Spain; mcarbonero@uloyola.es*

Abstract

Into the accomplishment of the census 2011 the National Institute of Statistics INE has changed the methodology of capture of information with regard to the used one into 2001

The census 2001 was based on a general survey in every home, while the 2011 one was based on administrative records and a demographic and socioeconomic survey to a sample of the population.

Both of them have been treated statistically to assess their estimations, being compared with other known statistics. To do this, it has been used a methodology of contrast of hypothesis of equality of proportions on a sample for big populations.

The result confirms that the population estimations in Census 2011 are more precise than the 2001 ones. However, this methodology has done that is lost availability social and economic data in small units.

Consequently, the census has lost part of the usefulness as tool of socioeconomic analysis for small areas, such as, small isolated municipalities or marginal zones of big cities

Keywords: Cesus, statistical error, data collection, population estimate.

*This study is subsidized by the Carlos III Health Institute (Ministry of Health of Spain) [project PI15/01986] and co-funded by FEDER funds.

1. Introduction

The population and housing census, carried out every 10 years in Spain, provides data on the main demographic and social aspects of the population, as well as their distribution at any given time. The availability of reliable and updated information is crucial to improve the design of public policies.

So far Spain had conducted censuses by using the state-of-the-art methodologies, which includes two main strands. The first one consists on carrying out a general survey in every home, just as the census of 2001 was conducted (National Statistics Institute, 2007a). The second one has two elements, a short questionnaire administered to the complete population, and a long questionnaire only for a selected sample about socio-economic issues. The census conducted in 2001, due to its extension, meant broad inclusion of non-specialized personnel in the work of data collection. This fact joined to the lack of interviewed people and errors in the pre-census file increased the error associated with this part of the operation. This error had to be corrected by comparing the results with existing administrative records and other nationwide surveys as the Labour Force Survey (National Statistics Institute, 2007b). The census conducted in 2011 tries to reduce this error using all previously existing administrative data, and performing for the first time a census relied on this information (National Statistics Institute, 2011) This methodology was employed in previous censuses (National Statistics Institute, 1994, 1985)

Therefore, in order to improve the 17th census was carried out in Spain in 2011 (National Statistics Institute, 2014), within the 2010 World Population and Housing Census Programme (United Nations Statistics Division, 2008), based on the European Regulation 763/2008 that contained the main recommendations of this program (European Parliament and of the Council, 2008). Among the wide range of possible options for collecting information, Spain has opted for a model based on administrative records, verification of the information in situ and finally, the demographic and socioeconomic survey of around 12.3% of the population completed the statistical operation.

Taking into account all of the above, this work aims to assess the statistical error due to the application of the new methodology and to compare with those were in previous censuses. To

do this, we compare the survey of the census with other available data of demographic statistical operations such as Continuous Municipal Register Statistics and Economically Active Population Survey.

2. Methodology

2.1. Scope of the study

The census starts with a previous operation: the realization of the pre-census file to identify all households will be surveyed. This file is extracted from existing data in the municipal register. The regulation of this register was unified in 1996 and the first update in agreement with a new system was made at January 1998 (National Statistics Institute, 1996).

This connection and unification in the methodology of the municipal register bore to an enormous progress in the making of the pre-census file of the census 2001, since some errors and duplicities were eliminated. Due to being, the first one since this methodology existed the census it served like contrast of the register municipal. The consolidation of the Continuous register municipal in the decade of 2000, did that the Census of 2011 might be raised like an sampling operation using a pre-census file based on register municipal. This administrative record crossed with other records to add, if it was possible, the additional variables not included in the register municipal (National Statistics Institute, 2011) Nevertheless, this record still bears errors due to the lack of unification with other similar records like the electoral census. This unification would allow confirming the existing information in the register municipal, so the National Institute of Statistic (NSI) is working in this project currently (Jurado Jiménez, 2014).

Bearing in mind the previous thing, the study has centered on the autonomous region estimated population and the microdata of the census realized in Spain during November 2011, specifically on the basic information of the structure of population for genre and age.

To be able to estimate the committed error it is necessary to be provided with other reference estimations, so has been used the municipal register population to the year 2012 (National Statistics Institute, 2012). In addition, has been calculated a theoretical population in 2011

originated by the population estimation of the Census 2001 (National Statistics Institute, 2007a), the register municipal population of 2002 (National Statistics Institute, 2002), the vegetative balances from the birth and death statistics from 2002 to 2011 (National Statistics Institute, 2016a) and the migratory balances from 0. To estimate the theoretical population has used an the equation of demographic balance (1, 2)

$$TPC_{11} = Census\ pop.\ 2001 + Vegetative\ Balance_{02-11} + Migratory\ Balances_{02-11} \quad (1)$$

$$TPMR_{11} = Mun.\ reg.\ pop.\ 2002 + Vegetative\ Balance_{02-11} + Migratory\ Balance_{02-11} \quad (2)$$

Being:

TPC= Theoretical population based on Census 2001

Census pop.2001= Population estimate in Census 2001

Vegetative Balance = Annual different between birth and death statistics from year 2002 until year 2011.

Migratory Balance= Annual balance of the residential variation statistics from year 2002 until year 2011.

TPMR= Theoretical population based on Municipal Register 2001

Mun. reg. pop.2002= Population in Register Municipal 2002

This system has been used to value the reliability of the estimations for migration of the census 2011 previously (César Vila et al., 2015). The data taken into account are the following ones: the microdata the population of the Census 2011, the estimation of population of the Census 2011, the municipal register population of the 2012, the theoretical population in 2011 from the census 2001 and the theoretical population of 2011 from the municipal register of 2002 (Table 1)

Table 1: Information of population of the different statistical sources

	Microdata 2011	Census 2011	Municipal Register 2011	Teoretical population Census 2001	Teoretical population Padrón 2001
Nacional	4.108.843	46.815.916	47.265.321	46.177.822	47.168.345
Andalucía	649.697	8.371.270	8.449.985	8.292.001	8.412.875

	Microdata 2011	Census 2011	Municipal Register 2011	Teoretical population Census 2001	Teoretical population Padrón 2001
Aragón	176.739	1.344.509	1.349.467	1.324.838	1.338.137
Asturias, Principado de	76.735	1.075.183	1.077.360	1.062.235	1.073.208
Balears, Illes	68.421	1.100.503	1.119.439	1.046.915	1.122.214
Canarias	122.413	2.082.655	2.118.344	2.013.853	2.163.131
Cantabria	55.583	592.542	593.861	579.855	586.999
Castilla y León	419.653	2.540.187	2.546.078	2.497.562	2.521.457
Castilla - La Mancha	259.957	2.106.331	2.121.888	2.084.800	2.106.322
Cataluña	622.163	7.519.843	7.570.908	7.297.631	7.460.961
Comunitat Valenciana	381.403	5.009.931	5.129.266	4.990.713	5.154.645
Extremadura	145.162	1.104.499	1.108.130	1.083.476	1.098.023
Galicia	216.376	2.772.928	2.781.498	2.745.339	2.786.829
Madrid, Comunidad de	480.496	6.421.874	6.498.560	6.409.725	6.513.493
Murcia, Región de	103.711	1.462.128	1.474.449	1.467.429	1.496.776
Navarra, Comunidad Foral de	85.598	640.129	644.566	623.786	637.585
País Vasco	193.146	2.185.393	2.193.093	2.186.823	2.212.517
Rioja, La	40.801	321.173	323.609	318.905	323.817
Ceuta	5.618	83.517	84.018	78.956	83.603
Melilla	5.171	81.323	80.802	80.487	83.260

Source: INEbase 2001-2012 (National Statistics Institute, 2016c)

2.2. Methods

For the estimation of the error, due to the limitations of existing information since it is not provided the factor of correction used by the NSI for the calculation of the estimations, has been performed a preliminary analysis that allows us to corroborate if the general information between the different records is approaching.

A methodology of contrast of hypothesis of equality of proportions on a sample for big populations has been used, in which H_0 defends the proportions equality between the information and H_1 that the proportions are different. In these contrasts, have been compared the proportions of the microdata of census versus all others and the estimation of population of the Census 2011 and the rest.

3. Result and discussion

As it is shown at tables 2 and 3, the population proportions in the autonomous regions are different both if the microdata is analyzed, as if there is analyzed the estimation realized by the NSI for the census 2011. It is normal because this methodology is affected by the size of the population. Nevertheless, bearing in mind the value z, the differences between the proportions of population are minimal, considering the large population that are managing.

Focusing on the contrasts realized with the microdata of the census of 2011 (Table 2), it fits better with the population of the current municipal register, especially in autonomous regions with large population, and in most cases, the adjustment in respect to the studied theoretical populations is worse. This supports the hypothesis of which the census 2011, made with the new methodology, is more coherent with the rest of the current records, at least as for the information of structure of population. It also confirms that the population estimations are more precise than the 2001 ones.

Table 2: Results of the Contrast of hypothesis realized between the required microdata of 2011 and the rest of estimations of population.

	Micro. Census 2011-M Register 2012		Micro. Census 2011- Census Theo. pop 2001		Micro. Census 2011-M.Register Theo. Pop. 2002	
	Z value	Error	Z value	Error	Z value	Error
Andalucía	-100,82	0,000	-113,68	0,000	-107,40	0,000
Aragón	141,48	0,000	138,99	0,000	142,23	0,000
Asturias, P. de	-57,99	0,000	-61,53	0,000	-58,06	0,000
Balears, Illes	-99,23	0,000	-90,08	0,000	-106,70	0,000
Canarias	-160,29	0,000	-155,09	0,000	-180,06	0,000
Cantabria	16,73	0,000	16,37	0,000	18,28	0,000
Castilla y León	316,38	0,000	313,93	0,000	318,28	0,000
Castilla - La Mancha	151,09	0,000	146,22	0,000	150,33	0,000
Cataluña	-41,58	0,000	-35,78	0,000	-36,59	0,000
C. Valenciana	-87,81	0,000	-101,49	0,000	-109,57	0,000
Extremadura	127,67	0,000	126,56	0,000	128,63	0,000
Galicia	-52,06	0,000	-58,76	0,000	-55,64	0,000
Madrid, C.de	-113,89	0,000	-131,31	0,000	-127,17	0,000
Murcia, Región de	-69,96	0,000	-80,13	0,000	-79,67	0,000
Navarra, C. Foral de	100,63	0,000	101,05	0,000	100,99	0,000
País Vasco	7,08	0,003	-3,20***	0,001	0,92***	0,356
Rioja, La	62,26	0,000	59,99	0,000	60,85	0,000

	Micro. Census 2011-M Register 2012		Micro. Census 2011- Census Theo. pop 2001		Micro. Census 2011-M.Register Theo. Pop. 2002	
Ceuta	-20,79	0,000	-17,82	0,000	-21,07	0,000
Melilla	-24,98	0,000	-26,14	0,000	-27,35	0,000

*** H_0 confirmed $p\text{-value} > 0.001$

Source: NSIBASE 2001-2012. (National Statistics Institute, 2016c)

Fuente: EP a partir de Censo 2001, Censo 2011, MNP y EVR.

Analyzing the contrasts is observed that the estimation of the census fits more appropriately than the comparison with the microdata (Table 3). It is logical due to the lack of the weighting factor used in the estimation. We focus in the concrete results of the table 3, they ratify the hypotheses raised with the microdata. Therefore, there are more proximity between the estimations of the census 2011 and the municipal register 2012 than between this first and the studied theoretical populations.

Table 3: Results of the Contrast of hypothesis of population's proportion made between the estimations of census population 2011 and the rest of estimations.

	Census 2011- M.Register 2012		Census 2011-Census theor. Pop. 2001		Censo 2011-M.Register Theor. Pop. 2002	
	Z value	Error	Z value	Error	Z value	Error
Andalucía	0,44 ***	0,660	-9,48	0,000	5,75	0,000
Aragón	4,89	0,000	0,84 ***	0,400	10,18	0,000
Asturias, P. de	5,59	0,000	-1,19 ***	0,234	6,92	0,000
Balears, Illes	-5,66	0,000	26,83	0,000	-9,08	0,000
Canarias	-7,80	0,000	20,57	0,000	-32,06	0,000
Cantabria	4,02	0,000	4,32	0,000	9,23	0,000
Castilla y León	8,39	0,000	3,69	0,000	17,23	0,000
Castilla - La Mancha	2,31 ***	0,021	-3,61	0,000	7,88	0,000
Cataluña	5,91	0,000	34,16	0,000	32,42	0,000
C. Valenciana	-23,57	0,000	-16,54	0,000	-35,41	0,000
Extremadura	4,72	0,000	4,11	0,000	10,05	0,000
Galicia	7,86	0,000	-4,51	0,000	3,04 ***	0,002
Madrid, C.de	-4,48	0,000	-22,82	0,000	-12,91	0,000
Murcia, Región de	1,01 ***	0,312	-15,08	0,000	-13,91	0,000
Navarra, C. Foral de	1,51 ***	0,131	6,87	0,000	6,53	0,000
País Vasco	6,47	0,000	-15,40	0,000	-5,19	0,000
Rioja, La	0,80 ***	0,421	-2,66 ***	0,008	-0,28 ***	0,778
Ceuta	0,73 ***	0,465	8,56	0,000	1,32 ***	0,186

Melilla	3,22^{***}	0,001	-0,68^{***}	0,495	-3,26^{***}	0,001
----------------	---------------------------	--------------	----------------------------	--------------	----------------------------	--------------

^{***} H_0 confirmed $p\text{-value} > 0.001$

Source: NSIBASE 2001-2012. (National Statistics Institute, 2016c)

Obviously, this is the first approach to the aims of the investigation, but as it has already been commented, the census has been realized by means of sampling. This sampling it was made from the records of the pre-census file in which a percentage of doubtful population existed. This error was corrected by a weighting factor, which is not available, this has supposed a limitation to the research, since it is not possible to calculate the real estimation errors between the microdata and the census population estimation.

4. Conclusion

The previous analysis has demonstrated that the reliability of the general information of the census has improved with the different changes of methodology made by the NSI. In this way, the variation between the different records has been minimized and it is going to approach to the reality. Consequently the new methodology has improved the estimation, beyond reducing costs.

But this affirmation can only be made regard to the general information like the structure of the population. Nevertheless, the previous census was allowing to make a radiography from the demographic, economic and social point of view of the current situation of the country to the small units, such as, municipalities and census sections. While the current census is not capable due to the used methodology. The accomplishment of the census through data from administrative bases and a survey to a sample of the population has done that is lost availability socio-economics data in small units, since its free access would commit an outrage against the statistical secret.

Consequently the census has lost part of the usefulness as tool of socioeconomic analysis for small areas, such as, small isolated municipalities or marginal zones of big cities

This conclusion is necessary to be qualify, since we will not obtain evidences of the previous statements until the data of the whole census is not published and that the NSI allows the access to the weighting factors of the statistical classes.

In any case, appears to that the politics is geared towards that the census loses this function, as the NSI comments in the census project 2011, which it recommends " that the census of 2021 would be able to be achieved exclusively with information of administrative records and other specific considerations about variables to researching in the census operation "(National Statistics Institute, 2011). This raises a future question: Is the NSI planning tools that replace this function that was fulfilling the census?

Keeping in mind the obtained results, the future lines of research will be focus on the analysis of the microdata of the census 2011, knowing the weighting factors, and the analysis of the existing mistake between the sample data and estimations, not only of the demographic variables, but the socioeconomic variables as well.

5. References

- César Vila, M., Pereira López, X., Verdugo Matés, R.M., 2015. Fiabilidad de los datos de migraciones en el Censo de 2011 y su explotación para el estudio de la movilidad regional del capital humano. Presented at the XLI International Conference on Regional Science. Innovation and geographical spillovers: new approaches and evidence (AEER 2015), Reus.
- European Parliament and of the Council, 2008. REGULATION (EC) No 763/2008 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 9 July 2008 on population and housing censuses [WWW Document]. URL http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2008.218.01.0014.01.ENG&toc=OJ:L:2008:218:TOC (accessed 12.10.15).
- Jurado Jiménez, A.M., 2014. El Padrón y su relación con los Censos de Población y Viviendas. *Indice Rev. Estad. Soc.* 10–12.
- National Statistics Institute, 2016a. Vital Statistics [WWW Document]. URL http://www.ine.es/en/inebmenu/mnu_dinamicapob_en.htm (accessed 4.30.16).
- National Statistics Institute, 2016b. Residential Variation Statistics [WWW Document]. URL <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fp307&file=inebase&L=1> (accessed 4.30.16).
- National Statistics Institute, 2016c. INEbase: List of statistical operations included [WWW Document]. URL http://www.ine.es/en/inebmenu/indice_en.htm (accessed 4.30.16).
- National Statistics Institute, 2014. Population and Housing Census 2011 [WWW Document]. URL http://www.ine.es/en/censos2011_datos/cen11_datos_inicio_en.htm (accessed 12.10.15).
- National Statistics Institute, 2012. Revision of the Municipal Register 2012. National level Datas, Autonomous Community and Province. [WWW Document]. URL

<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe245&file=inebase&L=1> (accessed 4.30.16).

National Statistics Institute, 2011. Proyecto de los Censos Demograficos 2011. National Statistics Institute, Madrid.

National Statistics Institute, 2007a. Population and Housing Census 2001 [WWW Document]. URL <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe242&file=inebase&L=1> (accessed 12.10.15).

National Statistics Institute, 2007b. Evaluacion de la calidad de los datos del Censo de Población 2001. National Statistics Institute, Madrid.

National Statistics Institute, 2002. Revision of the Municipal Register 2002. National level Datos, Autonomous Community and Province. [WWW Document]. URL <http://www.ine.es/dynt3/inebase/en/index.htm?type=pcaxis&file=pcaxis&path=%2Ft20%2Fe245%2Fp04%2F%2Fa2002> (accessed 4.30.16).

National Statistics Institute, 1996. Informes Metodológicos Estandarizados. Estadística Padrón continuo. National Statistics Institute, Madrid.

National Statistics Institute, 1994. Population and Housing Census 1991 [WWW Document]. URL <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe243&file=inebase&L=1> (accessed 12.10.15).

National Statistics Institute, 1985. Population and Housing Census 1981 [WWW Document]. URL <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe241&file=inebase&L=1> (accessed 12.10.15).

United Nations Statistics Division, 2008. 2010 World Population and Housing Census Programme [WWW Document]. URL <http://unstats.un.org/unsd/demographic/sources/census/census3.htm> (accessed 12.10.15).