

INSTITUTO NACIONAL DE ESTADÍSTICA



Anteproyecto de los Censos de Población y viviendas 2021

Subdirección General de Estadísticas
Sociodemográficas
Julio 2017

Índice

| | |
|-------|---|
| 1 | Introducción |
| 1.1 | EL CENSO DE 2011 |
| 1.2 | EL CONTEXTO INTERNACIONAL ACTUAL |
| 2 | El censo de población y viviendas 2021: perspectiva general |
| 3 | El censo de población: personas y hogares |
| 3.1 | EL PRIMER OBJETIVO: ¿CUÁNTOS SOMOS? |
| 3.1.1 | Los datos sobre población en España desde 1996 |
| 3.1.2 | La cifra de población del censo 2011 |
| 3.1.3 | Las cifras de población desde 2012 |
| 3.2 | LA CONSTRUCCIÓN DE UN FICHERO CENSAL DE POBLACIÓN PARA EL CENSO DE 2021 |
| 3.3 | EL FICHERO CENSAL DE POBLACIÓN: QUÉ VARIABLES INTERVIENEN |
| 3.3.1 | Variables demográficas (sexo, edad, nacionalidad, lugar de nacimiento) y relacionadas con las migraciones |
| 3.3.2 | Estado civil |
| 3.3.3 | Composición de hogares / relaciones de parentesco |
| 3.3.4 | Estudios en curso y nivel de estudios alcanzado |
| 3.3.5 | Variables económicas |
| 3.3.6 | Estudio de las variables relacionadas con la movilidad |
| 4 | Censo de viviendas y edificios |
| 4.1 | ASPECTOS GENERALES DEL CENSO DE VIVIENDAS |
| 4.2 | EL FICHERO CENSAL DE VIVIENDAS Y EL MARCO DE DIRECCIONES GEORREFERENCIADAS |
| 4.3 | DETERMINACIÓN DEL TIPO DE VIVIENDA SEGÚN USO |

| | |
|--------|--|
| 4.3.1 | Viviendas principales y no principales |
| 4.3.2. | Clasificación de viviendas no principales según consumo eléctrico |
| 4.4 | RÉGIMEN DE TENENCIA |
| 4.5 | OTRAS VARIABLES SOBRE LAS VIVIENDAS PRINCIPALES |
| 4.6 | CARACTERÍSTICAS DE LOS EDIFICIOS |
| 4.7 | VIVIENDAS, LOCALES Y HUECOS EN EL CENSO DE 2021 |
| 4.8 | ALOJAMIENTOS |
| 5 | Establecimientos colectivos |
| 5.1 | LA POBLACIÓN RESIDENTE EN ESTABLECIMIENTOS COLECTIVOS EN EL CENSO DE 2011 |
| 5.2 | PLANTEAMIENTO PARA EL CENSO 2021 |
| 6 | Incorporación de nuevas variables no investigadas en censos anteriores |
| 7 | Evaluación de la calidad. Operaciones de campo |
| 7.1 | ¿POR QUÉ TIENE MAYOR CALIDAD UN CENSO BASADO EN REGISTROS? |
| 7.2 | UN ANÁLISIS DE LA CALIDAD DEL FICHERO CENSAL A PARTIR DE LAS FUENTES |
| 7.3 | UNA ENCUESTA DIRIGIDA A LA POBLACIÓN Y OTRAS POSIBLES OPERACIONES DE CAMPO |
| 8 | Calendario de tareas |

Anteproyecto de los Censos de Población y viviendas 2021

Versión julio 2017

1 Introducción

1.1 EL CENSO DE 2011

En 2011 se realizó el último censo de población hasta hoy. Esta edición introducía una importante novedad respecto de todos los anteriores: por primera vez no se contemplaba en el proyecto el envío de cuestionarios a todas las viviendas del territorio nacional sino solo a una muestra de ellas, cercana al 10%.

La operación de 2011 dejó de ser un “censo” desde la perspectiva del proceso seguido para obtener los datos, si bien sí puede considerarse que el producto obtenido es un verdadero censo, al menos para una parte de las variables, en tanto que los datos procedían de un fichero que contenía todas las personas.

El producto censal es el fruto de una combinación de fuentes. Por un lado se construyó, mediante combinación de ficheros administrativos, el llamado “**fichero censal**”, que era un registro de población formado por casi 47 millones de individuos.

Por otro lado, se realizó una encuesta a una muestra de hogares, que recopiló información directa, mediante cuestionarios, de algo más de 4 millones de personas.

Además se realizó una operación de censo de edificios. Se recorrió todo el territorio nacional con el fin de identificar todos los edificios que contenían viviendas. En realidad, un 20% de las secciones censales no se visitaron porque se disponía de una información de partida suficientemente buena para ellas, lo que hacía innecesario el trabajo de campo.

Este censo de edificios, que consumió la mitad del presupuesto global dedicado al censo, era necesario para poder completar un marco de edificios y viviendas de calidad. Ya se disponía de un marco, a partir del censo de 2001 y de diversos métodos de actualización (sobre todo de la información proveniente de Catastro y del propio padrón continuo del INE). Pero era necesario completar esa información con un recorrido a pie de calle que permitiera resolver información incompleta y también proporcionar coordenadas geográficas a muchos edificios para los cuales no se disponía de ese dato.

Figura 1. Esquema global del Censo de Población y viviendas 2011.



1.2 EL CONTEXTO INTERNACIONAL ACTUAL

El censo de población es seguramente la operación estadística que se realiza en un mayor número de países del mundo. Según datos de Naciones Unidas, en la ronda 2005-2014, ha habido 214 países (un 93% de la población) que realizaron un censo, la mayoría en los años 2010 y 2011, mientras que solo 21 países no lo pudieron llevar a cabo.

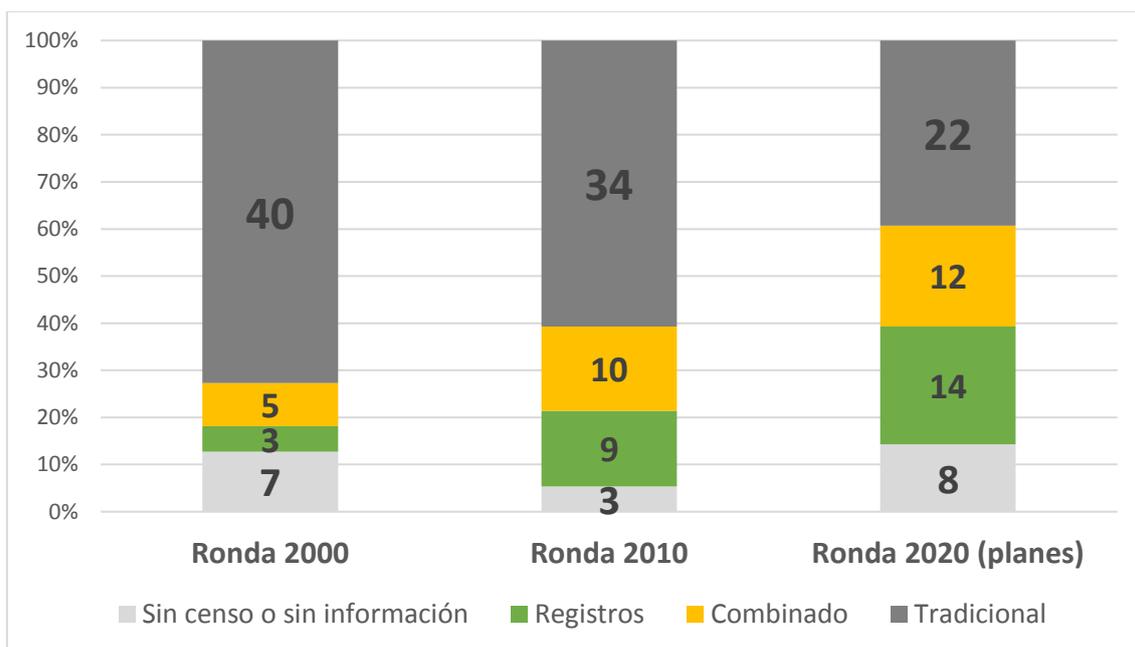
España no fue, ni mucho menos, el único país que renunció en 2011 a recoger cuestionarios de todos los hogares. Aunque el censo tradicional sigue siendo el método más frecuente (está presente en un 85% de los países), su importancia ha perdido peso en los últimos años.

La tendencia de construir la información censal sin necesidad de enviar cuestionarios a toda la población, iniciada por algunos países pioneros en los años 80, ha ido creciendo paulatinamente. Esta tendencia se hace todavía más relevante en el ámbito de países UNECE que está formado actualmente por 56 países¹ entre los que se encuentran muchos de los más desarrollados del mundo.

Según una encuesta realizada por UNECE en abril de 2015 sobre la metodología que tienen previsto llevar a cabo los países en la ronda de 2020 (comparada con las dos ediciones anteriores) se tienen los siguientes resultados:

¹ UNECE está formado por todos los países europeos, Turquía, Rusia y todos los estados post-soviéticos, Canadá, Estados Unidos e Israel. En la ronda del 2000 aparecen solo 55 países debido a que Montenegro pasó a formar parte de UNECE en el año 2006.

Figura 2. Métodos censales en distintos países (ámbito UNECE)



Puede verse como a lo largo de las dos últimas décadas el número de países con un Censo tradicional prácticamente se ha reducido a la mitad mientras que el número de países que optan por Censos basados en registros o combinado no ha parado de crecer.

España tiene previsto llevar a cabo en la ronda de 2020 un Censo en el que se utilicen todos los registros administrativos posibles, que sería completado, en su caso, con encuestas. Este planteamiento, que parte del llevado a cabo en 2011, dista poco del realizado por otros países como las 3 repúblicas bálticas (Letonia, Estonia y Lituania) y también Turquía. Todos estos países optaron por realizar un Censo combinado en la última ronda de 2010.

Hay 8 países que tienen pensado pasar del Censo tradicional en la ronda de 2010 al Censo combinado en la ronda de 2020. Son los siguientes: Canadá, Chipre, República Checa, Italia, Luxemburgo, Montenegro, Rumanía y Reino Unido.

2 El Censo de población y viviendas 2021: perspectiva general

El punto de partida para el censo de 2021 es el propio censo de 2011. En aquel momento, si bien se disponía de dos fuentes con datos de personas, el fichero censal y los datos de la muestra, el fichero censal solo pudo utilizarse para obtener de él las principales cifras: los totales de población por municipios (ver apartado 3.1). Todas las explotaciones detalladas del censo se construyeron a partir de la muestra, que a efectos de los usuarios fue el único producto censal.

Esta limitación se debió al hecho de que el fichero censal apenas tenía cuatro variables para cada persona. Así pues, el objetivo que se marcó el INE en 2014, al terminar los trabajos del censo 2011, fue estudiar hasta qué punto se podía aumentar el número de variables en ese fichero censal a partir de distintos registros administrativos y disminuir, por tanto, la información contenida en los cuestionarios censales.

El análisis de fuentes no es inmediato porque son muchas las variables recogidas en un censo como éste, que no solo censa población sino también edificios, viviendas y hogares y además recoge para cada una de estas unidades distintos tipos de información. Algunas de las variables son muy fáciles de obtener de registros, pero en otros casos el reto es mucho mayor.

Durante los años 2014 y 2015 se inició la investigación. Se contactó con multitud de departamentos de la Administración y se recibieron decenas de ficheros distintos con los que componer el fichero censal.

La investigación continuará durante los próximos años porque quedan muchos elementos por estudiar, pero la conclusión a la que se puede llegar ya desde finales de 2015 es que se dispone de mucha más información y de mucha mayor calidad de la que se preveía al inicio del análisis. Con seguridad se podrá construir más del 90% de la información que contenía el cuestionario censal de 2011 a partir de registros administrativos y estos ofrecerán mayor calidad que los datos basados en recogida directa mediante cuestionarios.

Esto implica que se pueda dar un paso muy importante en la mejora de calidad respecto del censo de 2011. Si el número de variables contenido en el fichero censal es muy superior al de 2011, la muestra, tal cual estaba diseñada en 2011, con el objetivo de extraer de ella toda la información para toda la población, deja de tener sentido.

En lugar de usar un fichero de población casi solo como marco y una muestra para el grueso de los datos, se tendría en 2021 que el grueso de los datos saldría ahora del fichero censal y la encuesta dirigida a la población se usaría solo para unas pocas variables y probablemente no dirigida a cubrir todo el territorio sino quizá solo para resolver situaciones concretas donde los registros no lleguen a dar información de calidad.

De forma resumida, se puede concluir, con la información que se dispone a primeros de 2017, que se plantea un Censo de Población y Viviendas 2021 en el que destacan las siguientes características:

Se construirá un **verdadero “producto censal”, basado en** cuantos registros administrativos sea posible.

La operación censal daría como resultado un conjunto de ficheros de microdatos: un fichero de personas formado por tantos registros como habitantes, un fichero de hogares, y un fichero de viviendas, todos ellos completos. A partir de ellos se calcularán todos los resultados agregados que componen los productos de difusión del censo.

El producto que se pretende obtener sería similar al que se dispondría en caso de procesar los cuestionarios de todos los hogares de España en un censo clásico.

Una parte de la información que falta en los registros se puede imputar a partir de encuestas ya existentes.

Esta es la solución que ya se utiliza en los países que basan sus censos en registros administrativos para algunas variables que no se encuentran, o no se construyen del todo bien a partir de registros.

Por ejemplo, el tipo de hogar es una variable difícil de construir a partir de registros. En algunos casos es muy sencillo: hogares unipersonales o parejas casadas con o sin hijos. Suponen la mayoría de casos. En otros, el uso de fuentes administrativas (datos tributarios, permisos de residencia de **extranjeros, datos de filiación del DNI, Registro Civil,.**) **permiten resolver** la incógnita sobre parentescos.

Pero subsistirá con total seguridad un residuo de hogares para los que no sea posible determinar vínculos entre sus miembros. La solución utilizada por otros países en esos casos es la imputación probabilística para estos hogares a partir de datos de encuestas existentes. En ese sentido el INE dispone de muestra suficiente tomando datos de la EPA o ECH, incluso acumulando muestra de ambas.

Como en todo censo, se dejará de publicar alguna información contenida en censos anteriores.

En todos los censos se ha dejado de publicar información que se recogía en censos anteriores, normalmente por su falta de actualidad o necesidad. Así, los censos de 2001 se eliminaron preguntas sobre la residencia de hecho y se recoge solo información de residencia de derecho. De cara a este censo hay información recogida en censos anteriores que con seguridad no se podrá

encontrar en registros administrativos. Pero para algunas de estas variables cabe preguntarse si tiene sentido mantenerlas en el proyecto censal.

A día de hoy, para edificaciones nuevas no censadas en 2011, si bien no se descarta aún obtenerla, se podría perder información relativa al equipamiento de las viviendas tal como la disponibilidad de cuarto de aseo y baño o ducha, calefacción, acceso a Internet, sistema de suministro de agua, número de habitaciones, ascensor en el edificio, garaje, número de plazas del garaje, gas, tendido telefónico, agua caliente central y evacuación de aguas residuales.

Para algunas de ellas se puede asumir que están presentes en el 100% de las viviendas o edificios y no parece grave dejar de investigar con carácter general estas variables (los porcentajes recogidos en el Censo de 2011 fueron cercanos al 99,5% en casi todos los ámbitos territoriales). En algunas otras nos enfrentamos a dificultades, aunque todavía no insalvables.

Un caso particular es el régimen de tenencia de las viviendas ocupadas. Se está trabajando para intentar elaborarlo a partir de datos administrativos pero a día de hoy no se puede asegurar si se conseguirá una información de suficiente calidad de los registros.

En algunos de estos casos una posible solución podría ser obtenerlas de una encuesta pero eso implicaría no publicar los datos para cada persona o cada hogar, sino de forma agregada por ámbitos superiores (provincia).

Se incorporan nuevos temas y nuevas formas de medir algunas variables.

El uso de datos administrativos abre la puerta a publicar nuevas variables que no se habían recogido en censos anteriores por no sobrecargar el cuestionario o por las dificultades que suponía. Nos referimos, por un lado, a variables que se podrían recoger en el censo a nivel individual, como por ejemplo, la discapacidad.

Pero también hay información nueva que, por diversas razones, no se va a poder incorporar de forma individual pero sí agregada, aunque con mucho detalle. Por ejemplo, no es previsible que se pueda proporcionar individualmente el consumo eléctrico de las viviendas o el nivel del ingresos del hogar o de la persona pero sí se plantea proporcionar el nivel medio de ingresos de cada sección censal o municipio o datos sobre consumo eléctrico para ámbitos geográficos muy reducidos (normalmente secciones censales): porcentajes de viviendas por tramos de consumo, por ejemplo.

En algunos casos se renunciaría a recopilar la información tal como se publicó en el censo 2011 pero a cambio se ofrece información distinta, que sirva para medir el mismo fenómeno y que sí pueda obtenerse de registros. Este es el caso de la información sobre movilidad cotidiana (lugar de estudios y trabajo). No se puede dar esta información tal cual se ofrece en censos anteriores pero a cambio se trabaja en dar una aproximación diferente, mucho más rica en cuanto a cantidad y calidad, a partir de datos de localización de teléfonos móviles.

Al tratarse de movimientos de telefonía móvil anónimos, sin identificación del portador del teléfono, se perdería la capacidad de analizar esa movilidad cotidiana clasificada por variables demográficas (sexo, edad, etc.) pero se dispone de mucho mayor detalle en el conocimiento de los movimientos lo que a efectos de estudios de movilidad parece ser prioritario.

Se ofrecerá una información de mayor calidad y una calidad más contrastable

La disponibilidad de registros administrativos proporciona, en general, un salto cualitativo en la calidad de la información respecto de la recogida mediante cuestionarios.

Por ejemplo, si fuéramos capaces de obtener a partir de datos administrativos el nivel educativo que ha alcanzado toda la población, es evidente que este dato fehaciente ofrecería mucha mayor cantidad y calidad de información que el que puedan aportar los entrevistados.

El problema es que no siempre va a ser posible, para todos los ficheros administrativos, encontrar datos de toda la población, y el nivel educativo es uno de ellos.

Para poder asignar el nivel educativo a toda la población va a ser necesario acudir a distintas fuentes, con distinta calidad. Otro tanto va a ocurrir con variables como las de migraciones (lugares de residencia anteriores) o con las relacionadas con la actividad económica (ocupación, relación con la actividad económica).

Se puede, en cierto modo, cuantificar la calidad de las variables censales mediante un método sencillo. Aunque la forma detallada de evaluar la calidad todavía no está del todo decidida, la estrategia que se pretende seguir se basará en estas líneas generales:

Supongamos que asignamos una puntuación de calidad (con valores 1, 2 y 3) a cada una de las fuentes que intervienen para construir una determinada variable. Volviendo al caso del nivel educativo, supongamos que cuando para una persona hemos determinado su nivel educativo alcanzado a partir de datos del sistema educativo (por ejemplo, un grado universitario o un título de bachiller), a este dato de le podemos asignar **un valor "1"**, que significaría que el dato se considera de mucha calidad por obtenerse de una fuente directa relacionada con la variable a medir.

Para otra parte de la población, el dato se va a poder asignar, no a partir de registros del sistema educativo sino de otros registros administrativos (por ejemplo, otros registros donde figura el nivel educativo alcanzado como es el propio padrón). En estos casos podríamos **asignar a ese dato un valor "2"**, lo que significaría que el dato tiene una calidad media porque proviene de una fuente indirecta o no directamente relacionada con el tema a medir.

Por último, habrá un porcentaje de población para la cual no tendremos ninguna información sobre nivel educativo alcanzado y tendríamos que imputar ese nivel alcanzado de forma probabilística usando, por ejemplo, distribuciones de frecuencias a partir **de encuestas. Este dato tendría entonces un valor "3"** que significaría dato imputado.

Se puede cuantificar así que calidad tiene la asignación de una variable como el nivel educativo alcanzado a partir de los porcentajes de población que tienen calidad 1, 2 o 3.

Pero también se puede analizar, para una determinada persona, cuantas variables sobre ella se han obtenido con calidad 1,2 o 3, es decir, se puede hacer un análisis del **el fichero censal de personas "por columnas" (calidad de las variables a lo largo del fichero) o "por filas" (calidad de los datos de las distintas variables obtenidas para una misma persona).**

Un análisis de calidad así se puede realizar durante estos próximos años con el fin de llegar a 2021 con el mejor fichero posible. Por eso el censo de 2021 es un proceso que culmina en ese año pero que se entiende ya iniciado en 2014 y que sobre todo se concentra en los años previos a 2021.

Pero además, esta cuantificación de la calidad nos puede servir para el análisis **"por filas", es decir, para qué personas los registros administrativos no ofrecen información de suficiente calidad.**

Se prevé realizar una encuesta específica

El análisis **de calidad anterior "por filas"** nos podría servir para dirigir una posible encuesta específica de manera mucho más eficiente que si la encuesta se dirige a toda la población.

Idealmente se podría diseñar una operación de campo para obtener información directa, mediante cuestionario, solo de aquellas personas o viviendas cuyos datos a partir de registros no tienen calidad suficiente. Pero seguramente una operación así sería costosa y difícil; en definitiva, poco eficiente.

Si se concentra el problema de falta de calidad en los registros, como es esperable, en algunos municipios o secciones censales, se puede diseñar una encuesta específica, que se desarrollaría a primeros de 2021, para recoger mediante cuestionarios, la información en aquellos lugares específicos. La mejor forma de recopilar datos en esos lugares sería el método clásico de envío de cuestionarios censales. En cualquier caso se trataría de una operación de un tamaño muy inferior al de la encuesta de 2011.

La encuesta estaría dirigida a resolver situaciones concretas. Más bien se trataría una investigación de campo específica dirigida solamente en aquellos lugares donde pudieran darse mayores problemas.

Lo mismo se puede pensar del censo de edificios, para el cual quizá tuviera sentido plantearse solo una operación destinada al estudio de determinadas secciones censales. Por ejemplo, se pueden realizar tareas de campo en aquéllas donde haya alojamientos no convencionales, que también constituyen un elemento de difícil solución a partir de los registros administrativos.

La “realidad demográfica” y la realidad “administrativa”. La estimación de la población no registrada en el padrón

No se puede soslayar que hay cierta información de interés demográfico que no figura en los registros administrativos y que difícilmente va a figurar, al menos en algunos casos. Las relaciones de convivencia o de pareja que no sean declaradas ante una instancia administrativa; o el trabajo informal, no recogido en los ficheros laborales; o el régimen de tenencia de la vivienda en ciertos casos; son varios los ejemplos de información que solo puede obtenerse mediante preguntas directas a los ciudadanos.

Cabe hacerse también una reflexión sobre hasta qué punto una parte de la realidad que no se registra en ámbitos administrativos llega a ser declarada en un cuestionario estadístico. No siempre será así y no debemos suponer que el cuestionario estadístico es percibido por los ciudadanos como algo tan confidencial como es en realidad. El cuestionario que recoge el INE en un censo clásico seguramente no refleja toda la realidad sino solo aquella que los informantes quieren revelar.

Con todo, es cierto que una entrevista con un cuestionario estadístico puede aproximarse mejor en ciertos ámbitos a la realidad. Pero para eso precisamente se han creado las encuestas sociales, que siguen teniendo todo su vigor. Un censo basado en registros administrativos debe complementarse con un buen sistema de encuestas sociales.

Por poner un ejemplo, los registros administrativos no van a ofrecer información correcta sobre convivencia de parejas no casadas, y seguramente subestimariamos el número de parejas del mismo sexo. Seguramente una encuesta pueda ayudar a suplir esa carencia. Así, encuestas como la actual Encuesta Continua de Hogares, se erigen en un instrumento necesario para ofrecer información, no tan detallada en el territorio, pero si complementando al censo en temas que no puedan cubrirse desde los registros.

Considerar como habitantes de un territorio a la población que figura registrada es solo una de las aproximaciones posibles a medir la población residente. Evidentemente siempre existe la posibilidad de que haya población que reside en España y que no figure empadronada, por diversas causas. De la misma forma podemos argumentar que un censo clásico que intente visitar todos los hogares puede no encontrar a toda la población, de forma que tampoco ese sistema garantiza una contabilidad exacta de la población residente. Además,

el propio concepto estadístico de población residente hace que no toda la población que en un momento dado vive en un determinado territorio deba contabilizarse.

Según el reglamento comunitario que establece el marco básico para la realización de los censos, medir población registrada es un método válido de estimación de la población residente.

En cualquier caso podemos hacer una primera estimación de hasta qué punto una encuesta puede ayudar a encontrar población no empadronada en España, usando para ello las evidencias que ofrecen propias encuestas del INE.

Una prueba de la calidad del padrón la ofrece la Encuesta de Población Activa (EPA). Para realizar cada trimestre esta encuesta se selecciona una muestra a partir de un marco de viviendas, independientemente del padrón. Posteriormente se cruzan las personas que han respondido a la encuesta con el padrón, con el fin de complementar la información con datos administrativos.

Por ejemplo, en el cuarto trimestre de 2016, ese cruce da como resultado que de un total de 134.412 personas de 16 o más años recogidos en los cuestionarios de la EPA ha solo hay 179 registros no encontrados en padrón (un 0,13%) lo cual no quiere decir que esas personas no figuren en el padrón sino que, a falta de identificadores tipo DNI, NIE o pasaporte, la búsqueda por otros campos **(nombre y apellidos, fecha de nacimiento,..)** no ha permitido enlazar unívocamente con el padrón.

En todo caso, de cara al proyecto censal cabe la posibilidad de estudiar la incorporación de población que no estando en padrón figure como residente en otros ficheros (tributarios, seguridad social), como por ejemplo, población extranjera cuya inscripción en padrón ha caducado o que se ha dado de baja o que simplemente nunca se ha empadronado. En principio se analizará la posible incorporación de esa población como residente, si esos otros ficheros ofrecen suficientes evidencias de tal residencia.

3 El censo de población: personas y hogares

3.1 EL PRIMER OBJETIVO: ¿CUÁNTOS SOMOS?

El objetivo original de los censos es contar a la población. En la actualidad en España este objetivo principal deja de tener tanta importancia, pues debido a la existencia del padrón continuo no se tiene la incertidumbre en las cifras de población que se da en otros países durante los periodos intercensales, que normalmente son de diez años.

En muchos lugares del mundo, durante el periodo que media entre dos censos, no hay más que unas proyecciones basadas, en algunos casos, en muy pocas evidencias. En España, la existencia de un registro de población permanentemente actualizado, como es el padrón continuo, hace que se puedan proporcionar cifras de población con mucha frecuencia y que, por tanto el objetivo original del censo, actualizar las cifras de habitantes, se convierta casi en secundario.

3.1.1 Los datos sobre población en España desde 1996

Durante los años 70 y 80 del siglo pasado en España se disponía de una cifra oficial de población cada diez años (el Censo de 1970 y el de 1981) y en los periodos intercensales se procedía a la actualización de esos datos mediante la renovación o rectificación padronal.

Esencialmente el método consistía en solicitar a los ayuntamientos datos agregados de su población residente cada año, que éstos obtenían de los recuentos de sus padrones municipales, existentes en España desde hace siglos en muchos casos. Este sistema se conocía como **“rectificaciones padronales”**².

Así mismo, el año central de cada periodo intercensal se realizaba lo que se **conoce como “renovación padronal”**. **En 1986, entre los censos de 1981 y 1991, se produjo la última renovación padronal. La “renovación” de 1986, al contrario de la “rectificación” anual era un verdadero procedimiento censal.** En 1986 se enviaron a los hogares los volantes de empadronamiento para actualizar sus datos de vecinos residentes en cada municipio.

En 1996 se cambió la normativa padronal y se puso en marcha un sistema muy novedoso de gestión continua e informatizada de los Padrones municipales, que pasaban a estar coordinados por el INE. Era el inicio de un registro de población en España.

Las antiguas rectificaciones anuales dieron así paso a cifras anuales de población a partir del padrón, a uno de enero de cada año, desde 1998.

Pero a diferencia de cómo venía sucediendo hasta entonces, de acuerdo con la nueva normativa, el censo de población y viviendas no venía a corregir la serie

² Ley 70/1980, de 16 de diciembre, por la que se modifican las fechas de referencia para la formación de los Censos Generales de la Nación y de renovación del Padrón Municipal de Habitantes.

padronal sino que las cifras del padrón se actualizaban por medio de actuaciones de los ayuntamientos, no mediante una operación de envío de cuestionarios censales.

Así pues, desde la entrada en vigor del padrón continuo, se sientan las bases para que en España haya dos series de datos de población independientes entre sí. Es a partir de 2001, con el censo cuando se consuma esta separación en dos series de datos:

- Por un lado, cada año, se tienen las cifras poblacionales basadas en recuentos directos de padrón ([Estadística del Padrón Continuo](#)) que se publican, como cifras oficiales para todos los municipios de España a uno de enero.

- Por el otro, los datos del censo de 2001 requieren de una actualización con fines estadísticos. Se construye la **que podemos denominar "serie estadística"** ([Cifras de Población](#)). Durante la década 2001-2011 esta serie estadística se **conoce como "Estimaciones de la población actual" (EPOBA)** y desde 2011, con alguna mejora metodológica, **bajo la nueva denominación de "Cifras de Población"**. Esta serie se actualiza también cada año (de hecho, se publica cada seis meses), sumando a la cifra censal los flujos demográficos estimados.

En realidad esas dos fuentes no son del todo independientes: lo son en cuanto a que la segunda toma como punto de partida el censo, pero la serie estadística también se apoya en el padrón, dado que tanto los flujos migratorios como el propio censo 2011 se obtienen del padrón. No obstante, todavía a 1 de enero de 2016 subsiste una diferencia a nivel nacional entre ambas fuentes de unos 110.000 habitantes (un 0,2%).

3.1.2 La cifra de población del censo 2011

El recuento de la población de España a 1 de noviembre de 2011 que se publicó como principal dato del censo (46.815.916 habitantes) se obtuvo mediante un sistema completamente nuevo: no era el producto de un simple recuento de cuestionarios recogidos, como ocurría tradicionalmente con los censos, ni tampoco era el número de habitantes que se obtenía del padrón. Se ideó un procedimiento específico para contar la población a partir del padrón pero usando cuanta información había disponible para acercar esa cifra al concepto de población residente habitual, propia de unos censos de población, y no al de población empadronada, que cumple otros fines.

Se construyó el llamado **"fichero precensal" a 1 de noviembre de 2011**, que es la fecha de referencia del Censo, a partir del aprovechamiento de los diversos registros administrativos disponibles, tomando al Padrón como elemento básico de su estructura.

El objetivo fundamental del fichero precensal era disponer de información suficiente para realizar el recuento de la población y de su estructura según las principales variables de clasificación (sexo, edad, nacionalidad, país de nacimiento).

Para ello, se partió de la base de datos del Padrón y de la información estadística y de fuentes administrativas (registros del DNI y Tarjetas de Residencia del Ministerio del Interior, nacimientos y defunciones del Movimiento Natural de la Población y del Registro Civil, Seguridad Social, Agencia Tributaria).

Se comprobó la presencia en otras fuentes de la población que figuraba en el padrón y también se comprobó que las defunciones y nacimientos estaban bien incorporados. No solo se cruzaron los registros vivos en padrón sino también registros eliminados por diversas razones, con el fin de hacer el análisis lo más exhaustivo posible.

Para la gran mayoría de las personas que figuraban en el fichero precensal se consideró confirmada su residencia en España ya que se consiguió encontrar esos registros de padrón en otras fuentes: eran afiliados dados de alta como trabajadores o pensionistas o figuraban como contribuyentes en las bases de datos de la Agencia Tributaria, todo ello durante el periodo de referencia (2011).

Los factores de recuento

A estas personas, cuya residencia en España se consideró segura, se les asignó un "factor de recuento" igual a 1. **Este "factor de recuento", ideado para el censo de 2011, puede entenderse como la probabilidad de que una persona resida en España a partir de la evidencia encontrada en distintas fuentes. Un valor 1 es una probabilidad del 100% y por tanto, la persona se cuenta como residente.**

Otros registros tendrán un factor de recuento 0, y por tanto esas personas no se cuentan como residentes. Son aquellos que se han encontrado en los ficheros de defunciones y que aún figuran en el padrón, generalmente porque se trata de defunciones recientes aún no procesadas y en algunos casos porque no se ha podido determinar perfectamente el enlace entre persona fallecida y empadronada. Esta es una ventaja de la aproximación con fines estadísticos a los registros: no se puede borrar un registro del padrón al no tener certeza pero sí puede descontarse un residente a efectos estadísticos.

Así mismo, como consecuencia de este proceso se incorporan al fichero precensal aquellos nacimientos que aún no han sido procesados para su incorporación al padrón pero de los cuales se tiene ya constancia por los boletines del Movimiento Natural de la Población.

Como consecuencia de este cruce de ficheros, hubo un porcentaje pequeño de la población -pero no despreciable (inferior al 3%)- para el cual no se disponía de evidencia alguna en los registros. La única presencia de una persona en el

padrón, en algunos casos registrada con mucha antigüedad, no era suficiente para contarla como residente. Esta población se consideró “**dudosa**” y se introdujo un método estadístico para contarla.

Se contabilizó esa población asignándoles “**factores de recuento**” distintos de 0 o 1. Los factores de recuento se calcularon a partir de la encuesta dirigida a la muestra del 10% de la población que formó parte del proyecto censal.

Para ello se realizó una partición del fichero de registros dudosos (aproximadamente 1.046.000 personas) en clases o grupos de población definidos por características sociodemográficas: edad, según intervalos de edades agrupadas, la nacionalidad y la provincia de residencia.

Las clases se configuraron de forma que al menos contuvieran 1.000 registros dudosos y que todo registro dudoso perteneciera a una clase. Las clases que resultan con menos de 1.000 dudosos se agruparon con el objetivo de llegar al menos a ese umbral: primero, para una misma nacionalidad y provincia, se agrupaban en mayores intervalos de edad, si esto fuera insuficiente se agruparon para una misma nacionalidad a nivel de la Comunidad Autónoma y si aun así no se consigue el mínimo se agrupaba para dicha nacionalidad a nivel nacional, por último y residualmente se llegaba a agrupar nacionalidades a nivel nacional. De esta forma el fichero precensal quedó configurado en 724 clases con al menos 1.000 dudosos cada una. La más numerosa fue la de ciudadanos del Reino Unido residentes en la provincia de Alicante con edades entre 65 y 69 años, con casi 6.000 dudosos, de unos 21.000 que figuraban en el fichero precensal.

Se estableció el mínimo de 1.000 dudosos para tener garantías de que se encontrasen representados en la encuesta ya que esta iba dirigida a un 10% de la población. Las características que definen las clases (edad, nacionalidad, provincia de residencia) son características observables al realizar la Encuesta ya que se recogían en el cuestionario

A partir del número teórico de dudosos que se debían haber recogido en la encuesta y del número real de ellos que se encontró (corrigiendo por falta de respuesta) se asignó a cada clase de dudosos un factor de recuento, es decir, una probabilidad de residir en España para las personas de ese grupo.

Así, por ejemplo, para los 6.000 dudosos de Reino Unido entre 65 y 69 años en la provincia de Alicante, resultó un factor de recuento de 0,36 (se encontraron en la muestra el 36% de los esperados)

A cambio, hubo clases con factores de recuento superior a 1: por ejemplo, los pakistaníes entre 25 y 30 años en la Comunidad de Madrid tienen un factor de recuento de 1,67 (se encuentran en la encuesta más de los esperados).

El factor de recuento medio resultó ser 0,424, lo que implica que los 1.046.000 dudosos se contabilizaron como una población de 444.000 personas aproximadamente.

Al final de este proceso todos los registros tienen asignado un factor de recuento, con valor 1 para el 97% de los registros, y con valor distinto de uno

(casi siempre menor), para el resto. El fichero así formado se conoció como Fichero Censal Final Ponderado (FCFP), que contenía unos 47,4 millones de registros, pero que sumaban 46,8 millones de residentes estimados: la cifra censal

3.1.3 Las cifras de población desde 2012

A partir del censo 2011, cuya fecha de referencia es 1 de noviembre, se construye la serie estadística de cifras de población, publicada cada semestre, con inicio en 1 de enero de 2012.

Esta serie se actualiza **mediante un método “de flujos”**, es decir, al stock poblacional referido a 1 de enero de cada año se le suman (o restan, según corresponda) los siguientes flujos:

- Nacimientos y defunciones, obtenidos de las estadísticas de movimiento natural de la población.
- Migraciones interiores y exteriores: provenientes de la Estadística de migraciones.
- Adquisiciones de nacionalidad española de residentes (operación estadística del mismo nombre) para proporcionar cifras por nacionalidad.

Entretanto, cada año se siguen publicando las *Cifras oficiales de población de los municipios españoles* que se derivan del Padrón. Los resultados son ligeramente diferentes debido sobre todo a la naturaleza administrativa del padrón, que introduce ciertas restricciones. Con las diversas mejoras puestas en marcha en los últimos años en la contabilización de población esa brecha entre serie estadística y padrón se ha hecho casi insignificante y es de esperar que esa distancia se reduzca. Con todo es previsible que siga habiendo alguna diferencia en el futuro.

Mantener los dos conjuntos de cifras de población no debería producir ninguna alarma si no se separan mucho entre sí y se pueden explicar las diferencias que tienen que ver con los distintos fines de ambas fuentes, padrón y serie estadística. Si bien el padrón tiene como propósito también cuantificar la población residente, es cierto que por los muchos usos y derechos que se derivan para los ciudadanos a partir de él, no pueda ser siempre inmediata y automática su actualización. Por poner un ejemplo sencillo, una defunción de una persona sobre cuyos datos de identificación no figuren correctamente no generará inmediatamente una baja en el padrón pero sí se contabiliza en la serie estadística la existencia de esa defunción.

En la actualidad se prevé que en 2021 las cifras de población censales se calculen con un método parecido al de 2011: fundamentalmente se partirá de los datos del padrón pero se estimará con métodos distintos la población

dudosa. Si ya en el censo de 2011 esa población dudosa suponía en torno al 3% del total es de esperar que para 2021 sea muy inferior, dadas las mejoras producidas en los padrones municipales durante estos años. No sería eficiente basar el recuento de una población dudosa tan pequeña en una encuesta dirigida a una muestra de hogares.

3.2 LA CONSTRUCCIÓN DE UN FICHERO CENSAL DE POBLACIÓN PARA EL CENSO DE 2021

En primer lugar se debe elegir la fecha de referencia. Se contempla para 2021, de forma aún tentativa, tomar como fecha el 1 de enero. El principal elemento en la estrategia de trabajo para el censo de 2021 pasa por la construcción de un fichero censal de población (y uno de viviendas, como se explica en el apartado correspondiente).

Esencialmente un fichero censal se puede ver como un “fichero plano” así:

Figura 3. Esquema del Fichero Censal de Población

| Fichero censal de población | | | | | | | | | | | |
|-----------------------------|------|------|--------------|-----------|-----|----------------------------------|-------------------------------------|--|-----------|--|---|
| Número | sexo | edad | nacionalidad | país | | lugares de residencia anteriores | estudios en curso y nivel alcanzado | situación en relación con la actividad | ocupación | lugar trabajo, desplazamientos, vacaciones | relaciones con resto de miembros de hogar |
| 1 | H | 17 | España | Ecuador | | AA-BB-CC | YYY | ZZZZ | xxx | YY | xxxx |
| 2 | M | 21 | España | España | | AA-BB-CC | YYY | ZZZZ | xxx | YY | xxxx |
| 3 | M | 13 | Marruecos | Marruecos | | AA-BB-CC | YYY | ZZZZ | xxx | YY | xxxx |
| ... | .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 47.000.000 | M | 89 | España | Francia | | AA-BB-CC | YYY | ZZZZ | xxx | YY | xxxx |

Las incógnitas a despejar en la construcción de un fichero así están en las filas (¿qué población se incluye?) y también en las columnas (¿cómo obtener todas las variables que se recogen en un censo a partir de registros administrativos?).

Respecto de la primera incógnita, durante estos últimos años se han realizado estimaciones sobre la población dudosa que podría resultar de utilizar los mismos criterios seguidos en 2011. Pendientes de realizar un cruce completo con todas las fuentes disponibles similar al realizado entonces, se estima que el número de dudosos a enero de 2016 estaría por debajo del 2% de registros.

Asignar factores de recuento a la población en 2011 trajo aparejado un recuento poblacional con decimales, lo cual resultaba poco natural para los usuarios menos expertos en estadísticas demográficas, acostumbrados a que la población solo puede medirse con números enteros. El uso de decimales es habitual en la encuestas por muestreo en las cuales la muestra representa a una población a partir de unos factores de ponderación o de elevación que son inherentes al proceso estadístico, pero no lo es en recuentos exhaustivos.

Para 2021, el método que se plantea haría innecesario el uso de factores de recuento y por tanto de decimales. Dado que, además, se contempla para 2021 que el producto censal sería un fichero de microdatos, es decir, un registro por cada persona, asignar factores de recuento a la población complicaría la lógica del producto censal y su utilización. Por ejemplo, pensemos que un determinado municipio está habitado por 100 personas con factor de recuento medio de 0,5, es fácil determinar que la población estimada para el municipio sería de 50 habitantes. Pero determinar cuál sería el número de viviendas ocupadas resultaría complejo y con cierto grado de arbitrariedad.

La condición de dudoso de un registro se determina a partir del cruce con otros registros para cuantificar la verosimilitud de la residencia en España de cada persona. Otros países que cuentan la población a partir de registros de población determinan si cada registro se contabiliza o no a partir de una regresión logística. Así, no tendríamos registros seguros y dudosos sino que, para cada registro del padrón, a partir del cruce con otros registros, tendríamos que decidir, mediante un modelo matemático, si el registro se considera válido o no.

El fichero precensal contendría todos los registros válidos a partir de la verosimilitud obtenida del cruce de registros.

De esta manera cualquier recuento de otras variables, como el número de hogares en que viven, se haría a partir de ese fichero censal de forma natural y siempre usando números enteros.

Como resultado de todo este proceso se obtendría un fichero precensal, construido a partir del padrón, que contendría tantos registros como habitantes saldrían del recuento de población (la cifra censal).

El siguiente paso consistiría en recoger, para cada uno de esos registros, todas las variables que típicamente se estudian en un censo utilizando diferentes registros administrativos.

3.3 EL FICHERO CENSAL DE POBLACIÓN: QUÉ VARIABLES INTERVIENEN

El segundo elemento de la estrategia de construcción del fichero censal consiste en obtener las variables de distintos registros administrativos. El punto de partida es el padrón, que contiene únicamente variables básicas: sexo, edad, nacionalidad, lugar de nacimiento, además de las variables de identificación (DNI, NIE o número de pasaporte, necesarias para efectuar los enlaces con otros ficheros).

A partir del padrón se ha elaborado un primer fichero precensal a 1 de enero de 2016, que tiene un identificador para cada vivienda y que en él se dispone de variables migratorias que se obtienen directamente del Padrón sobre lugares de residencia y años de llegada (lugar de residencia actual, anterior, hace un año y hace diez años así como los años de llegada a España, a la CCAA, al municipio y a la vivienda).

Se trata de construir a partir de él un verdadero fichero censal que contenga todas las variables que son objeto de investigación en un censo.

A continuación analizamos por grupos temáticos las columnas de este fichero, es decir, aquellas variables que forman parte de la investigación censal sobre personas.

3.3.1 Variables demográficas (sexo, edad, nacionalidad, lugar de nacimiento) y relacionadas con las migraciones

En el cuestionario del censo 2011 figuraban estas preguntas:

Cuestionario Individual de la Persona 1

Escriba los siguientes datos para la persona que aparece en primer lugar (Persona nº 1) en la Lista de personas.

Nombre y apellidos:

Fecha de nacimiento: día mes año Sexo: Hombre Mujer

País de nacimiento:

España. *Escriba municipio y provincia:*

Municipio:

Provincia:

Otro país:

¿Cuál es su nacionalidad?

Si tiene doble nacionalidad, española y otra, marque ambas opciones y escriba el país correspondiente.

Si tiene doble nacionalidad, pero ninguna es la española, escriba únicamente una de ellas.

Española

De otro país:

1 ¿Desde qué año reside en esta vivienda?

Desde el año Desde que nació (Pase a la pregunta 3)

¿y en este municipio?

Desde el año Desde que nació (Pase a la pregunta 3)

¿y en esta comunidad autónoma?

Desde el año Desde que nació

¿y en España?

Desde el año Desde que nació

¿Dónde residía antes de llegar por última vez a este municipio?

En otro municipio:

Provincia:

En otro país:

2 ¿Dónde residía hace 1 año?

En este municipio (o no había nacido aún)

En otro municipio:

Provincia:

En otro país:

¿Dónde residía hace 10 años?

En este municipio (o no había nacido aún)

En otro municipio:

Provincia:

En otro país:

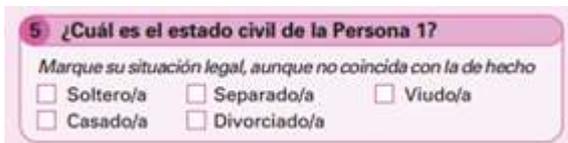
Todas estas preguntas se pueden responder a partir de la información que se almacena en la base de datos del padrón y asignarlas a toda la población.

Por un lado las variables básicas del individuo (preguntas iniciales del cuestionario del Censo 2011) se obtienen directamente de la descarga de la base de datos del Padrón y por otro lado, las variables migratorias (preguntas 1 y 2 del Censo 2011) se obtienen a partir de una explotación de la información de la base padronal y, en aquellos casos en los que falte información, se puede acudir a los datos de la propia persona del Censo del 2001.

El padrón inicia su historia en el año 1996; por tanto, para cualquiera de las cuatro cuestiones relativas a los años de residencia de la pregunta 1, el año más antiguo sería siempre 1996. Sin embargo, durante el año 2015 se ha completado ya el enlace con el censo de 2001, que fue exhaustivo, de forma que para toda la población se puede construir su historia residencial más allá del año 1996.

Por tanto, los trabajos para estas variables los podemos considerar como ya finalizados y este grupo de preguntas se pueden incorporar con calidad contrastada al fichero censal.

3.3.2-Estado civil



5 ¿Cuál es el estado civil de la Persona 1?

Marque su situación legal, aunque no coincida con la de hecho

| | | |
|------------------------------------|---------------------------------------|----------------------------------|
| <input type="checkbox"/> Soltero/a | <input type="checkbox"/> Separado/a | <input type="checkbox"/> Viudo/a |
| <input type="checkbox"/> Casado/a | <input type="checkbox"/> Divorciado/a | |

La variable estado civil no forma parte del conjunto de variables recogidas en Padrón ni existe una única fuente de datos que disponga de información sobre ella para toda la población. Por ello se tiene que acudir a múltiples fuentes de datos para tratar de obtener información de esta variable. Las más destacadas son:

1. Fichero de contribuyentes del IRPF de la Agencia Tributaria. Además se utilizarán ficheros específicos de los territorios forales (País Vasco y Navarra). Todas las personas que realizan la declaración están obligadas a declarar su estado civil legal. Esta fuente de datos es la que dispone de una mayor cobertura de información.
2. Prestaciones de viudedad de la SS
3. Boletines de matrimonios del Movimiento Natural de la Población del INE. Sirve para detectar las personas casadas.
4. Ficheros de separaciones/divorcios de INFOREG (ficheros informatizados de los Registros Civiles). Actualmente una gran mayoría de las separaciones y divorcios se registran en dicha aplicación.
5. Información del registro central de extranjeros para conocer matrimonios de extranjeros no formalizados en España.

6. Censos de Población 2001 y 2011 del INE. A pesar de que ya ha pasado bastante tiempo desde la realización de estos Censos, se usará esta información para considerar casadas a aquellas personas que figuran como casadas en dichos censos y que sigan compartiendo vivienda en el fichero censal con las mismas personas y para determinadas situaciones en las que no se disponga de información de otras fuentes.

Es necesario establecer unas prioridades para aquellos casos donde el estado civil de una persona pueda obtenerse, y no necesariamente ser igual, de diferentes fuentes de datos. Como regla principal, se otorga más credibilidad a los ficheros que tienen una contraprestación económica que a otros que no la tienen.

Después de integrar la información de las fuentes mencionadas anteriormente, quedará un porcentaje pequeño de población sin información del estado civil que deberá ser imputado mediante un modelo probabilístico basados en distribuciones de frecuencias condicionadas por la edad, tamaño del hogar, etc.

Es importante resaltar que la variable estado civil (legal o de derecho) está correlacionada con otra variable para la que también tenemos que proporcionar información, que es saber si una persona vive con su cónyuge/pareja (estado civil de hecho, ver 3.3.3), por lo que es importante que la información de ambas variables sea coherente siempre que sea posible.

3.3.3 Composición de hogares / relaciones de parentesco

En el censo de 2011, las principales variables utilizadas para investigar sobre la composición de hogares fueron **variables “derivadas”, es decir, no se preguntan tal cual en los cuestionarios sino que se construyen a partir de esta pregunta que figura en el cuestionario del censo de 2011.**

6 ¿Viven en esta vivienda los siguientes familiares de la Persona 1?

Cuando marque Sí, escriba el número con el que este familiar aparece en la Lista de personas de la página 2:

| | | | | |
|--|-----------------------------|-----------------------------|---------------------------------|----------------------|
| Padre | <input type="checkbox"/> NO | <input type="checkbox"/> Sí | → Persona nº | <input type="text"/> |
| Madre | <input type="checkbox"/> NO | <input type="checkbox"/> Sí | → Persona nº | <input type="text"/> |
| Cónyuge o pareja | <input type="checkbox"/> NO | <input type="checkbox"/> Sí | → Persona nº | <input type="text"/> |
| Otros familiares (hijos, hermanos, etc.) | <input type="checkbox"/> NO | <input type="checkbox"/> Sí | → Nº de persona de uno de ellos | <input type="text"/> |

Hay cierta dificultad para disponer de esta información a partir de registros administrativos. Por ejemplo, tomemos la clasificación detallada de estructura del hogar usada en el censo de 2011:

Censos de Población y Viviendas 2011. Hogares

Resultados nacionales

Hogares según la estructura del hogar (detallada)

Unidades: hogares

| | |
|---|-----------|
| Hogar con una mujer sola menor de 65 años | 1.054.513 |
| Hogar con un hombre solo menor de 65 años | 1.429.621 |
| Hogar con una mujer sola de 65 años o más | 1.279.486 |
| Hogar con un hombre solo de 65 años o más | 429.700 |
| Hogar con padre que convive con algún hijo menor de 25 años | 180.826 |
| Hogar con madre que convive con algún hijo menor de 25 años | 693.167 |
| Hogar con padre que convive con todos sus hijos de 25 años o más | 153.055 |
| Hogar con madre que convive con todos sus hijos de 25 años o más | 666.209 |
| Hogar formado por pareja sin hijos | 3.804.677 |
| Hogar formado por pareja con 1 hijo menor de 25 años | 2.023.841 |
| Hogar formado por pareja con 2 hijos en donde algún hijo es menor de 25 años | 2.552.052 |
| Hogar formado por pareja con 3 ó más hijos en donde algún hijo es menor de 25 años | 538.471 |
| Hogar formado por pareja con 1 hijo de 25 años o más | 919.634 |
| Hogar formado por pareja con 2 hijos de 25 años o más | 252.663 |
| Hogar formado por pareja con 3 ó más hijos de 25 años o más | 35.262 |
| Hogar formado por pareja o padre/madre que convive con algún hijo menor de 25 años y otra(s) persona(a) | 894.956 |
| Otro tipo de hogar | 1.175.560 |

Fuente: Instituto Nacional de Estadística

Construir esta información a partir del padrón es inmediato para los hogares unipersonales, pero en hogares de más de una persona se precisa disponer de relaciones de parentesco entre los miembros que no figuran en el fichero del padrón, que solo nos informa de qué personas comparten domicilio. En concreto, para poder agregar los hogares según esa clasificación detallada del censo se precisa conocer las relaciones de pareja y las de filiación (hijos) dentro del hogar.

La idea con la que se trabaja de cara a 2021 consiste en tratar de obtener información del padre, madre, cónyuge o pareja y otro pariente de cada persona dentro del hogar a partir de registros administrativos de la siguiente manera:

-PADRE o MADRE: La idea principal consiste en buscar a alguien dentro del hogar que cumpla determinadas condiciones en las variables sexo, edad y apellidos y que tenga como nombre de pila del padre/madre el mismo valor que el que aparece en los ficheros de DNI del Ministerio de Interior.

-CONYUGE: A diferencia de las variables padre/madre de una persona, conocer quién es el cónyuge de una persona es más complicado. Para este propósito se utiliza información de la AEAT, de MNP, de **censos anteriores**. Además para los casos en donde las fuentes anteriores no proporcionen respuesta, intentamos ver si existen hijos en común y si se puede detectar que dos personas son cónyuge fijándonos en los apellidos y edades de los hijos.

-OTRO pariente: Se busca dentro del hogar a candidatos que cumplan determinadas condiciones en los apellidos (p.ej: apellidos comunes en hermanos). Además se tienen en cuenta determinadas reglas de reciprocidad que tratan de capturar a todos los parentescos. Por ejemplo: si A es padre de B

y B es padre de C, entonces A es otro pariente (abuelo) de C y C es otro pariente (nieto) de A.

Una vez que hayamos identificado para cada persona quién es su padre, madre, cónyuge y otro pariente, seremos capaces de generar de forma derivada todas las variables de hogar: tipo de hogar, estructura de hogar, número de núcleos **en el hogar, número de familias...**

La laguna más importante que se tiene en los registros administrativos, en España y en otros países, a la hora de identificar los tipos de hogar está en la diferenciación de los hogares formados por una pareja sin hijos de los otros tipos de hogar. ¿Cómo sabemos si dos personas que están compartiendo un piso son pareja o no? La solución para estos casos en los que no tengamos información consiste en emplear modelos que basados en encuestas existentes proporcionen una aproximación estadística a este fenómeno.

El algoritmo para determinar los tipos de hogar concretos a partir de toda esta información está siendo objeto de mejoras en la actualidad por el equipo de Censos del INE.

3.3.4 Estudios en curso y nivel de estudios alcanzado

En el Censo 2011, las variables censales relacionadas con los estudios eran las siguientes:

-Para menores de 16 años información de si está acudiendo o no a un centro escolar (pregunta 9)

-Para personas de 16 años o más: el máximo nivel de estudios alcanzado (pregunta 10), el campo al que pertenecen dichos estudios (pregunta 11) e información sobre los estudios en curso (pregunta 18). Las preguntas del cuestionario censal de 2011 que recogían esta información eran:

Si la Persona 1 tiene 16 años o más, pase a la pregunta 10.
Si tiene menos de 16 años, continúe en la pregunta 8.

9 ¿Va a algún centro escolar (incluyendo guarderías)?

SÍ → (Pase a la pregunta 10)

NO → Ya ha terminado el Cuestionario Individual de la Persona 1

10 ¿Cuál es el mayor nivel de estudios que ha completado?

Marque sólo una opción

No sabe leer o escribir

Sabe leer y escribir pero fue menor de 5 años a la escuela

Fue a la escuela 5 o más años pero no llegó al último curso de ESO, EGB o Bachiller Elemental

Llegó al último curso de ESO, EGB o Bachiller Elemental o tiene el Certificado de Escolaridad o de Estudios Primarios

Bachiller (LOE, LOGSE), BUP, Bachiller Superior, COU, PREU

FP grado medio, FP I, Oficialía Industrial o equivalente, Grado Medio de Música y Danza, Certificados de Escuelas Oficiales de Idiomas

FP grado superior, FP II, Maestría Industrial o equivalente

Diplomatura universitaria, Arquitectura Técnica, Ingeniería Técnica o equivalente

Grado Universitario o equivalente

Licenciatura, Arquitectura, Ingeniería o equivalente

Máster oficial universitario (a partir de 2006), Especialidades Médicas o análogos

Doctorado

(Pase a la pregunta 11)

Las personas de 16 años o más que no se encontrasen en ninguna de las 5 primeras situaciones y sin embargo sí estuviesen estudiando, se clasificarían en **la categoría "Otra situación", por lo que de forma indirecta, esta pregunta también recoge información de los estudiantes.**

Es importante resaltar que en el Censo 2011 se recogió información de todas las variables descritas anteriormente, pero el reglamento de Eurostat sólo establece como obligatorias las de las preguntas 10 y 14³.

De cara al futuro Censo 2021, teniendo en cuenta la última propuesta recibida por parte de Eurostat y las variables recogidas en censos anteriores, estaremos obligados, como ocurrió en 2011, a dar respuesta a las preguntas relacionadas con el nivel de estudios y la situación laboral de la persona. Adicionalmente, si las fuentes de datos lo permiten, se obtendrá información del resto de variables para así seguir con la trayectoria censal de difusión de estas variables.

Las fuentes administrativas con las que se está trabajando a día de hoy para asignar el nivel de estudios a todas las personas de 15 o más años son:

- Información de Padrón
- Información de los Censos 2001 y 2011
- Los ficheros de títulos que el propio Ministerio de Educación remite anualmente al INE (a la unidad de Padrón)
- Los ficheros de demandantes de empleo y de certificados de profesionalidad del SEPE
- Ficheros de graduados proporcionados por el MECD
- Ficheros de matriculados en enseñanzas del sistema educativo también proporcionados por el MECD

En caso de no disponer de información, se procede a imputar ésta de forma probabilística.

Este algoritmo, aunque no está cerrado totalmente en sus especificaciones de cara al futuro Censo 2021, se ha aplicado ya a una publicación que se ha realizado por primera vez en diciembre de 2015⁴ (Estadística de Defunciones).

El procedimiento para la obtención del nivel educativo tiene cuatro fases principales:

1. Recuperación de la información de las fuentes disponibles

Para cada una de las personas de 15 años o más de Padrón se recopila toda la información de nivel de estudios disponible por medio de enlaces entre los distintos ficheros.

³ Eurostat exigía dar respuesta a las preguntas 10 y 14 para las personas de 15 años o más, pero tanto en el cuestionario del Censo 2011 como en censos anteriores, se establecía el límite en personas que tuviesen como mínimo 16 años.

⁴ Para más información consultar: http://www.ine.es/metodologia/t20/t2030306_niveduc.pdf

2. Normalización del nivel de estudio de las distintas fuentes de datos

A la hora de asignar el nivel de estudios, el hecho de que cada una de las fuentes involucradas en el algoritmo utilice una clasificación distinta genera problemas, por lo que es preciso establecer una clasificación homogénea. En concreto en el Censo 2021 tenemos previsto utilizar una adaptación específica de la clasificación la CNED-A⁵.

3. Tratamiento y depuración de la misma

Existe una correlación entre la edad de la persona y el nivel educativo alcanzado, por lo que debemos garantizar su coherencia. En los casos en los que la información no sea compatible, habrá que imputar un valor válido.

4. Obtención del nivel educativo mediante el algoritmo

Una vez depurada la información, se asigna a cada persona el nivel de estudios, asignando el valor más alto si hay discrepancia entre las fuentes de datos disponibles. Si no se dispone de información, o si el nivel de estudios se corresponde con varios valores posibles, se imputará un valor utilizando una fuente de datos externa⁶.

La principal fuente de información que deberá servir en un futuro para obtener el nivel de estudio de las personas es el Ministerio de Educación, pero a día de hoy no se dispone de la información de títulos completa y actualizada.

El desfase temporal existente entre la emisión de un título y la comunicación del mismo al INE provoca que, en algunos grupos de edad, en concreto los de edades más jóvenes, tengan un nivel de estudios en muchos casos inferior al que tienen en la realidad.

Además, el uso de distintas clasificaciones en cada una de las fuentes involucradas hace que a veces sea difícil establecer correspondencias entre las distintas clasificaciones y la CNED-A adaptada y que en algunos casos las correspondencias no sean directas.

En cuanto a la variable de estudios en curso, aunque no es obligatoria y actualmente no se dispone de ningún fichero que recoja esta información, se intentará dar respuesta a ella si las fuentes de información administrativas lo permiten.

Con el objetivo de obtener información de calidad para todas estas variables, se trabaja con el Ministerio de Educación. Es necesario recabar información de los muchos agentes involucrados en el sistema educativo.

⁵ Se puede consultar la clasificación en: http://www.ine.es/metodologia/t20/t2030306_niveduc.pdf

⁶ Para el Censo 2021 se deberá utilizar como fuente externa una encuesta con información actualizada (por ejemplo la EPA o la ECH).

Las fuentes administrativas con las que se está trabajando a día de hoy para obtener información de variables económicas son:

-TGSS: Obtenemos información de los trabajadores que cotizan a la Seguridad Social y de otras variables (código de actividad del establecimiento, situación profesional)

-SEPE: Nos proporciona información de los parados y del código de ocupación a partir de los contratos de las personas.

-INSS: Recoge información de las personas del régimen de la Seguridad Social que cobran una pensión y la causa de la **misma (invalidéz, jubilación.)**.

-Muface, Mugeju e Isfas: Obtenemos información de los funcionarios que no cotizan a la Seguridad Social.

-Clases pasivas: Recoge información de las personas que no son del régimen de la Seguridad Social y que cobran una pensión.

A la hora de trabajar con múltiples fuentes para una variable concreta pueden darse situaciones de conflicto (por ejemplo, que una persona tome valores diferentes) que se deben resolver primando las fuentes según diversos criterios.

Al igual que en otros casos mencionados anteriormente, para aquellas personas para las cuales, tras la integración de todas las fuentes no se disponga de información, deberá imputarse el valor de estas variables mediante un modelo probabilístico basado en información de otras encuestas.

A día de hoy se han realizado pruebas para analizar el resultado de estas variables y comparar la información con los datos obtenidos en el Censo 2011. El primer resultado es esperanzador: se puede afirmar que para las variables "situación laboral", "situación profesional" y "actividad del establecimiento", se puede construir la información con suficiente nivel de calidad a partir de las distintas fuentes.

Por otra parte, aunque la información relativa a la variable "ocupación" presenta las mayores dificultades; ya se ha empezado a trabajar con información existente en el SEPE relativa a los contratos laborales.

Es prematuro a día de hoy aventurar si se va a poder responder a esa variable a partir de registros administrativos, pero los primeros análisis realizados son prometedores. En el peor de los casos, se tendría que proporcionar esa información a partir de encuestas existentes, como la EPA.

Esa estrategia se ha seguido ya en otros países como en el caso de Holanda. La acumulación de muestra correspondiente, por ejemplo a dos años completos de la EPA, daría lugar a un tamaño muestral cercano a las 250.000 personas. Ello permitiría proporcionar cifras de ocupación a dos dígitos (en algunos casos a tres) a nivel provincial, lo cual puede ser suficiente para satisfacer la demanda del Reglamento comunitario pero parece un punto de partida no del todo satisfactorio. Una estrategia de obtención de la ocupación para un porcentaje elevado de los ocupados a partir de ficheros administrativos complementado

con la mejora a partir de imputación probabilística, tomando como referencia la encuesta, pudiera ser finalmente la solución adoptada.

3.3.6-Estudio de las variables relacionadas con la movilidad

La información relacionada con las preguntas de movilidad fue de la más demandada del Censo 2011. Tanto los sistemas de peticiones automáticos como los de peticiones a medida reflejaron que un gran porcentaje de los usuarios estaban interesados en este tipo de información.

En el Censo de 2011 se preguntó a las personas por la ubicación de su lugar de trabajo/estudio (pregunta 19) y también por si pasaba 14 noches o más (por motivos laborales o de veraneo) en un municipio diferente del de residencia (pregunta 4). Con estas dos preguntas se calculó la población vinculada, incluida por primera vez en el Censo 2001, que permite disponer de una mejor estimación de la carga poblacional real que soporta un municipio.

Adicionalmente se incluyeron otras tres preguntas relacionadas con el número de viajes diarios, el medio de transporte y el tiempo transcurrido para llegar al lugar de trabajo/estudio.

4 En los últimos doce meses, ¿ha pasado más de 14 noches (aunque no fueran seguidas) en otro municipio de España o en otro país?

Puede ser por razones de trabajo, estudio, fines de semana, vacaciones o porque reside en más de un municipio

NO → (Pase a la pregunta 5)

Sí → Indique el lugar donde ha pasado más noches y el número aproximado de noches que ha pasado allí:

Otro municipio:

Provincia:

Otro país:

Nº de noches:

¿Dispone en este lugar de una segunda vivienda (ya sea en propiedad, alquiler o cedida gratis)?

Sí NO

19 ¿Dónde está su lugar de trabajo o de estudio?

Si trabaja y estudia, le preguntamos por el sitio donde trabajó la semana pasada

En el propio domicilio

En varios municipios (comercial, conductor, ...)

En este municipio

En otro municipio:

Provincia:

En otro país:

Código postal de su centro de trabajo o de estudios:

Ya ha terminado el Cuestionario Individual de la Persona 1

20 ¿Cuántos viajes de ida y vuelta realiza normalmente cada día desde esta vivienda hasta el lugar de trabajo o estudio?

Ninguno (porque tiene una segunda residencia desde la que se desplaza a trabajar o estudiar)

Uno diario (es decir, uno de ida y otro de vuelta)

Dos o más diarios

Ya ha terminado el Cuestionario Individual de la Persona 1

21 ¿Cómo va normalmente desde su casa hasta el lugar de trabajo o estudio?

Si utiliza varios medios de transporte, marque los dos con los que cubre más distancia

| | |
|--|---------------------------------------|
| <input type="checkbox"/> En coche o furgoneta como conductor | <input type="checkbox"/> En moto |
| <input type="checkbox"/> En coche o furgoneta como pasajero | <input type="checkbox"/> Andando |
| <input type="checkbox"/> En autobús, autocar, minibús | <input type="checkbox"/> En tren |
| <input type="checkbox"/> En metro | <input type="checkbox"/> En bicicleta |
| | <input type="checkbox"/> Otros medios |

22 ¿Cuánto tiempo tarda normalmente desde su casa hasta el lugar de trabajo o estudio?

| | |
|--|--|
| <input type="checkbox"/> Menos de 10 minutos | <input type="checkbox"/> Entre 45 y 59 minutos |
| <input type="checkbox"/> Entre 10 y 19 minutos | <input type="checkbox"/> Entre 1 hora y hora y media |
| <input type="checkbox"/> Entre 20 y 29 minutos | <input type="checkbox"/> Más de hora y media |
| <input type="checkbox"/> Entre 30 y 44 minutos | |

Conseguir toda esta información a partir de fuentes administrativas no parece posible.

En la TGSS se dispone de información de la cuenta de cotización que, indirectamente y posiblemente con algunos problemas de calidad, puede proporcionar información sobre el lugar de trabajo. No obstante la ubicación geográfica de la cuenta de cotización no siempre coincide con el lugar físico del puesto de trabajo, lo que limita el análisis.

No obstante, existe una nueva y prometedora fuente que podría dar respuesta a gran parte de la información proporcionada en 2011: el uso de la información de coordenadas espacio-temporales de los teléfonos móviles.

El INE se encuentra en la actualidad estudiando las posibles formas de acceder a esta información que obra en poder de las compañías operadores de telefonía móvil que cuentan con red propia de antenas y es prematuro anunciar la cantidad y calidad de la información de la que se dispondrá para el censo de 2021. En cualquier caso, lograr disponer de información de esa naturaleza se sitúa como una de las líneas prioritarias de investigación dentro del proyecto censal.

Si se pudiera analizar la información proporcionada por la señal de los teléfonos móviles, bajo ciertas hipótesis, como por ejemplo asumir que el lugar de residencia de una persona es aquel en donde se encuentra habitualmente por la noche o que su lugar de trabajo es aquel en el que está a ciertas horas del

día, se puede llegar a reconstruir información bastante parecida a la que se ofrece para el Censo 2011.

Adicionalmente esta fuente abre la puerta a nuevos productos como, por ejemplo, poder conocer diferencias entre la población diurna y nocturna de una ciudad o poder analizar día a día la evolución de la población de un determinado lugar geográfico.

4 Censo de viviendas y edificios

4.1 ASPECTOS GENERALES DEL CENSO DE VIVIENDAS

El censo de viviendas consiste en una cuantificación exhaustiva de todas las viviendas, teniendo en cuenta las siguientes definiciones:

Se entiende por vivienda todo recinto estructuralmente separado e independiente que, por la forma en que fue construido, reconstruido, transformado o adaptado, está concebido para ser habitado por personas o, aunque no fuese así, constituye la residencia habitual de alguien en el momento censal. Como excepción, no se consideran viviendas los recintos que, a pesar de estar concebidos inicialmente para habitación humana, en el momento censal están dedicados totalmente a otros fines (por ejemplo, los que estén siendo usados exclusivamente como locales).

Tipos de vivienda: A continuación podemos observar un pequeño esquema de los distintos tipos de viviendas que nos podemos encontrar.

- **Viviendas familiares**
 - **Viviendas convencionales**
 - **Viviendas principales**
 - **Viviendas secundarias**
 - **Viviendas vacías**
 - **Alojamientos**
- **Viviendas colectivas**

Vivienda familiar: Vivienda destinada a ser habitada por una o varias personas, general pero no necesariamente unidas por parentesco, y que no constituyen un colectivo.

Vivienda familiar convencional: Es una vivienda familiar que cumple todos los requisitos para ser habitada y en la fecha censal no se utiliza totalmente para otros fines. La vivienda convencional puede ser principal cuando es la residencia habitual de sus componentes. Si está destinada a ser ocupada sólo ocasionalmente (por ejemplo durante las vacaciones) se denomina vivienda secundaria. Cuando permanece sin ser ocupada se denomina vacía.

Alojamiento: Vivienda familiar que presenta la particularidad de ser móvil, semipermanente o improvisada, o bien que no ha sido concebida en un principio con fines residenciales pero, sin embargo, constituye la residencia de una o varias personas en el momento del censo (por tanto, los alojamientos **vacíos no se censan**). **Entran en este grupo las cuevas, chabolas, caravanas...**

Vivienda colectiva (o establecimiento colectivo): Vivienda destinada a ser habitada por un colectivo, es decir, por un grupo de personas sometidas a una autoridad o régimen común no basados en lazos familiares ni de convivencia. La vivienda colectiva puede ocupar sólo parcialmente un edificio o, más

frecuentemente, la totalidad del mismo. Por ejemplo, son viviendas colectivas las prisiones, las residencias de mayores o los conventos.

En el caso de las viviendas familiares, el fichero que se construyó para el censo de 2011 incluía datos para toda la muestra de viviendas relativos a la propia vivienda y al edificio que la contiene.

El contenido del fichero de viviendas en el censo de 2011 es:

- Datos de la vivienda (*)
- Tipo de vivienda
- Régimen de tenencia
- Calefacción
- Cuanto de aseo con inodoro
- Baño o ducha
- Acceso a Internet
- Sistema de suministro de agua
- Superficie útil
- Número de habitaciones
- Datos del edificio
- Número de plantas sobre rasante
- Número de plantas bajo rasante
- Tipo de edificio
- Número de inmuebles
- Año de construcción
- Estado del edificio
- Ascensor en el edificio
- Accesibilidad del edificio
- Garaje
- Número de plazas de garaje
- Gas
- Tendido telefónico
- Agua caliente central
- Evacuación de aguas residuales

(*) Figura también el factor de elevación de cada vivienda dado que se trata de una muestra

Análogamente a lo realizado con el fichero de personas, de cara al próximo censo de 2021 está previsto construir un fichero completo de viviendas familiares. Para ello es necesaria la integración de diversas fuentes administrativas.

Por otro lado se construiría otro fichero para las viviendas colectivas (ver capítulo 5. Establecimientos colectivos).

4.2 EL FICHERO CENSAL DE VIVIENDAS Y EL MARCO DE DIRECCIONES GEORREFERENCIADAS

La estrategia consiste en hacer uso de la información existente en todas las fuentes disponibles para determinar el total de viviendas dentro de cada ámbito de población, siendo el más básico la sección censal.

El producto que se pretende obtener, el *Fichero Censal de Viviendas (FCV)*, sería un fichero que contendría un conjunto mínimo de datos para todas las viviendas familiares. Además de este fichero, para las viviendas principales (aquellas donde reside población) se tendría un conjunto adicional de variables tales como el régimen de tenencia o datos de equipamiento. Este es el *Fichero Censal de Viviendas Principales (FCVP)*.

La información a construir figura en el siguiente esquema.

Figura 4. Esquema de ficheros censales sobre viviendas para 2021

Fichero censal de viviendas

| Número (identificador de vivienda) | Datos de localización (hasta sección censal) y coordenadas xy | | | ... | características del edificio |
|--|--|-----|-----|------|---------------------------------|
| | superficie | | | | |
| 1 | xxx | 45 | ... | zzzz | |
| 2 | xxx | 67 | ... | zzzz | |
| 3 | xxx | 120 | ... | zzzz | |
| ... | ... | .. | | | |
| ... | ... | .. | | | |
| ... | ... | .. | | | |
| 26.000.000 | ... | 30 | ... | zzzz | |

Fichero censal de viviendas principales

| Número (identificador de vivienda) | Otras características | | |
|--|--------------------------|----------------|--------------------------------|
| | régimen de tenencia | de la vivienda | Equipamiento de la vivienda |
| 1 | yyy | xxxxxx | xxxxxx |
| 3 | yyyy | xxxxxx | xxxxxx |
| 6 | yyyy | xxxxxx | xxxxxx |
| ... | ... | | |
| ... | ... | | |
| 19.000.000 | yyyy | xxxxxx | xxxxxx |

Como fichero de partida para la construcción del FCV, el INE dispone desde noviembre de 2011 de un marco territorial de referencia para sus operaciones ("**marco de direcciones** georreferenciadas", **MDG**) consistente básicamente en la relación de vías, tramos, aproximaciones postales (direcciones postales hasta

número, bloque/portal/escalera) y huecos en ellas (detalle hasta piso y puerta) existentes en la totalidad del territorio y contrastado y actualizado en campo en el último censo de 2011.

Este marco está relacionado con la población puesto que está cruzado con los ficheros padronales a través de las direcciones en que están registradas las personas en el Padrón y este cruce, que ya se ha revisado a fecha 2014 y se está revisando a 2016, está previsto que se actualice anualmente desde 2017, a la vez que se incorporan las modificaciones producidas en el callejero, comunicadas por la Oficina del Censo Electoral.

También contiene información derivada del cruce con los ficheros catastrales y del Instituto Geográfico Nacional (IGN) por lo que respecta a las coordenadas geográficas o a las referencias catastrales de los inmuebles. Aproximadamente se dispone tanto de las coordenadas como de las referencias catastrales de las aproximaciones postales en un 83% del fichero. Respecto a las viviendas, estas cifras superan el 90% en el caso de coordenadas.

Este MDG constituye el punto de partida para la construcción futura del FCV, que se construiría añadiendo variables a las viviendas contenidas en él.

4.3 DETERMINACIÓN DEL TIPO DE VIVIENDA SEGÚN USO

El método clásico para la determinación de la clase de vivienda convencional según su uso (principal, secundaria, vacía) se basa en la cumplimentación de esta información por parte de un agente censal. Este método es altamente subjetivo. Por ejemplo, muchas veces, aunque el agente se encuentre enfrente de la vivienda no hay evidencias para discernir si se trata de una vivienda secundaria o de una vacía.

4.3.1 Viviendas principales y no principales

La distinción fundamental, y aquella a la que obliga el actual Reglamento comunitario para las viviendas convencionales es la de vivienda principal y no principal.

Determinar las viviendas principales de forma consistente con el proyecto censal en su conjunto pasa por tener un identificador que permita asociar el fichero de personas con el fichero de viviendas, de forma que si una vivienda contiene al menos un registro del fichero de personas, la vivienda pasa a considerarse principal. Así, el número de viviendas principales en el FCV coincide por definición con el número de hogares en el fichero censal de personas.

Este paso está prácticamente resuelto de forma que hay (salvo problemas menores) enlace perfecto entre identificación del hogar en el fichero de personas tal como está a día de hoy y en el marco MDG, del cual se construirá el FCV.

4.3.2 Clasificación de viviendas no principales según consumo eléctrico

Para el censo 2021 se pretende establecer una nueva clasificación de las viviendas no principales basada en hechos más objetivos y abandonar la desagregación clásica de secundarias, vacías (y otras) y sustituirla por una nueva clasificación basada en la cantidad de consumo eléctrico (kwh): se pasaría de hablar de viviendas secundarias y vacías a poder distinguir viviendas no principales con bajo consumo eléctrico o sin consumo. Nada impediría, en principio, establecer rangos de consumo para clasificar en un mayor número de categorías a estas viviendas

Se dispone de información de consumo eléctrico, a partir de fuentes tributarias, individualizada de más de 40 millones de contadores eléctricos de todo el territorio nacional, correspondientes no solo a viviendas sino también a locales, comunidades de vecinos o empresas de todo tipo.

La primera dificultad a la hora de cruzar esos datos de consumo eléctrico con el FCV reside precisamente en determinar cuáles de esos registros corresponden a viviendas.

Una segunda dificultad se encuentra en el enlace entre ficheros. En una situación ideal, en la que el fichero de consumo eléctrico enlazara perfectamente con el de viviendas, se podría además proporcionar otras características de las viviendas no principales (todas aquellas características que estén en el marco o se puedan asociar a él), pero en la actualidad se está lejos de esa situación.

El identificador para enlazar FCV y consumo eléctrico podría ser la referencia catastral pero su uso todavía no está generalizado de manera que el fichero de consumo eléctrico apenas contiene ese campo relleno en un 51% de los registros. A falta de un identificador único de vivienda los cruces se realizan a partir de la dirección postal. Estos procesos de cruce a partir de direcciones postales siempre son más complicados y dan peores resultados que los de personas a partir de nombres y apellidos. Actualmente se logra enlazar en torno a un 70% de las viviendas lo cual no permite a día de hoy poder determinar, para cada vivienda del fichero censal, su consumo eléctrico.

Se está trabajando en una mejora de los cruces utilizando como apoyo el NIF de los titulares de los contratos del que se dispone en el fichero de consumo eléctrico y el NIF de los residentes según padrón del que se dispone en el marco, así como del NIF de los titulares de los bienes inmuebles de que se dispone en catastro. No obstante hay que tomar con mucha precaución esta información y garantizar que el enlace se produce no solo a través del NIF sino que simultáneamente se produce o bien por referencia catastral o bien por dirección postal.

En definitiva, no es posible todavía cruzar el fichero de consumo eléctrico satisfactoriamente con el MDG. Es probable que se logren niveles de cruce muchos más altos en los próximos meses pero sin duda subsistirá un número de secciones donde el cruce no arroje porcentajes suficientemente altos.

Eso no invalida la construcción de indicadores para el censo a partir del consumo eléctrico. Un primer objetivo que parece alcanzable es determinar la sección censal en la que se encuentran cada uno de los registros que en el fichero de consumo eléctrico se han identificado como viviendas.

Si se dispone de tal clasificación, una aproximación realista aunque quizá demasiado prudente consistiría en proporcionar resultados renunciando al enlace de ficheros y utilizar las cifras agregadas que se obtienen a nivel de sección censal y de municipio para, a partir de ellas, construir datos para la situación de viviendas no principales según consumo. No podríamos dar información sobre variables que no estén en el fichero de consumo eléctrico a menos que logremos mejorar el enlace con el MDG.

Es decir, a partir del MDG y su cruce con el fichero de población se podría proporcionar información para cada sección censal del total de viviendas y de las viviendas principales; por otro lado, a partir del fichero de consumo eléctrico, sin cruzar con ninguna otra fuente, se podrían proporcionar indicadores para cada sección censal sobre número de viviendas por rango de consumo. Obviamente esto puede arrojar información inconsistente, como secciones censales en las que hay muchas más o menos viviendas principales de las que tienen un consumo eléctrico que podríamos considerar compatible con estar ocupadas.

Así, el objetivo consiste en tener, para cada sección censal:

- El nº total de viviendas, y de ellas, el número de viviendas principales, se obtendría a partir del FCV, donde estaría integrado el padrón, catastros y otras fuentes. Por tanto, para estas viviendas tendríamos coordenadas geográficas y otras variables. Es decir, estamos considerando que toda vivienda en la que figura población empadronada es una vivienda principal.

- De las viviendas no principales podríamos tener el total, como recuento a partir del marco, pero no podríamos dar detalles para una subclasificación al no estar enlazada con el marco la información de consumo eléctrico, que permite discriminar entre ellas.

En resumen, para una sección o municipio dados tendríamos:

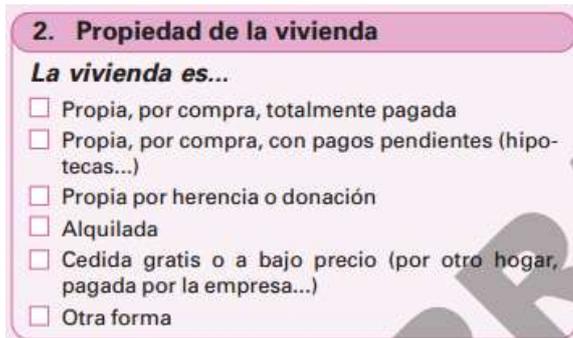
Información disponible según tipo de vivienda

| Tipo de vivienda: | Número | Coordenadas geográficas | Superficie, equipamiento, características |
|---|--------|-------------------------|---|
| Total | Si | Si | Si |
| Principales | Si | Si | Si |
| No principales | Si | Si | Parcialmente |
| con bajo consumo (entre X_1 y X_2 Kwh) | Si | No | No |
| Con muy bajo consumo ($0-X_2$ Kwh) o sin consumo | Si | No | No |

Otro tipo de información agregada que podría darse para la sección censal usando solo datos del fichero de consumo eléctrico podrían ser estadísticos sobre consumo medio o mediano, restringido solo a viviendas que tienen consumo o no, o datos sobre distribución de las viviendas según consumo (percentiles u otros) que, con seguridad, ayudan a dar una imagen agregada de las secciones censales en cuanto a su consumo eléctrico, e indirectamente, a su ocupación y nivel de vida.

4.4 RÉGIMEN DE TENENCIA

En el cuestionario del censo de 2011 se realiza la siguiente pregunta:



2. Propiedad de la vivienda

La vivienda es...

- Propia, por compra, totalmente pagada
- Propia, por compra, con pagos pendientes (hipotecas...)
- Propia por herencia o donación
- Alquilada
- Cedida gratis o a bajo precio (por otro hogar, pagada por la empresa...)
- Otra forma

Esta información se proporciona solo de las viviendas principales. Obtener esta clasificación a partir de fuentes tributarias de nuevo parece una labor difícil.

En el peor de los casos esa información sobre régimen de tenencia se tendría que construir a partir de encuestas existentes. Si tomamos la actual Encuesta continua de hogares, que tiene como muestra efectiva unos 56.000 hogares al año y acumulamos dos años, tendríamos unos 112.000 hogares distintos. Incluso se pueden sumar muestras de otras encuestas (La Encuesta de Condiciones de Vida podría aportar otros 20.000 hogares al año). En definitiva se podría tener, a partir de encuestas y acumulando dos años, un porcentaje de hogares cercano al 1% lo cual permitiría ofrecer buenas cifras para ámbitos de población como las provincias o para ciudades a partir de un cierto umbral, que con seguridad estará por encima de los 50.000 habitantes. Además, los diseños de estas encuestas no garantizan disponer de muestra suficiente en todos esos municipios aunque sí a nivel provincial.

Pero existe cierta información en fuentes administrativas con la que se puede trabajar. La relación entre propiedad de la vivienda y personas que residen en ella se puede determinar a partir de cruce entre padrón y catastros; por tanto se puede estudiar en cuántos casos el propietario de la vivienda reside en ella.

También se ha empezado a trabajar con fuentes tributarias. En principio, aquellas personas que estén obligadas a realizar la declaración (modelo 100 o similar en Haciendas Forales) tienen que indicar el régimen de tenencia de la vivienda en la que residen (propiedad, alquiler u otro tipo). Igualmente, si una

persona dispone de propiedades en alquiler que le proporcionan una rentabilidad, también tiene que incluirlas en la propia declaración.

A día de hoy no se dispone todavía de resultados que permitan evaluar la calidad de las fuentes administrativas. El objetivo es conseguir información a nivel individual tanto de vivienda en propiedad como de vivienda en régimen de alquiler; pero si se consiguen unos porcentajes superiores al 80% de las viviendas a partir de datos administrativos, el resto se podría imputar a partir de las encuestas.

Hay una perspectiva razonable de que al menos se logre alcanzar ese umbral. La combinación entre estos datos administrativos e imputación podría, en el peor de los casos, ser suficiente para publicar datos agregados para secciones censales o municipios al menos a partir de umbrales de tamaño bastante bajos (del orden de 50 o 100 viviendas).

4.5 OTRAS VARIABLES SOBRE LAS VIVIENDAS PRINCIPALES

La construcción del FCV y del FCVP a partir del marco de direcciones pasa por incorporar las variables que se pueda obtener de distintas fuentes. Además del régimen de tenencia, en el censo de 2011 se obtenían ciertos datos de las viviendas principales:

3. ¿Cuáles de las siguientes instalaciones tiene la vivienda?

Calefacción

- Colectiva o central
- Individual
- NO tiene instalación de calefacción pero sí algún aparato que permite calentar alguna habitación (ejemplo: radiadores eléctricos)
- NO tiene calefacción

Cuarto de aseo con inodoro (WC, retrete)

SÍ NO

Ducha o bañera

SÍ NO

4. ¿Tiene la vivienda contratado servicio de acceso a Internet?

SÍ NO

5. ¿Cuál es el sistema de suministro de agua?

Agua corriente por abastecimiento público

Agua corriente por abastecimiento privado o particular del edificio

No tiene agua corriente

6. ¿Cuál es aproximadamente la superficie útil de la vivienda?

No incluya espacios que no sean habitables como terrazas abiertas o jardines; tampoco sótanos, desvanes, trasteros...

m²

7. ¿Cuántas habitaciones tiene la vivienda?

Incluya la cocina, los dormitorios y todas las habitaciones que tengan 4 metros cuadrados o más. NO incluya cuartos de baño, vestíbulos, pasillos, terrazas abiertas...

habitaciones

También, en el censo de edificios 2011, el agente censal recogió ciertos datos del edificio como la antigüedad estimada, el número de plantas sobre rasante, etc.

De nuevo nos encontramos con información difícil de encontrar en registros. No obstante, algunas de las variables relativas a viviendas o edificios sí que sería posible obtenerlas a partir de la información contenida en los ficheros de los distintos catastros.

El catastro incorpora también información sobre el año de construcción/reforma del edificio y sobre la superficie de los bienes inmuebles que se podría incorporar al FCV en aquellas viviendas en las que el enlace sea posible.

Otras variables como el número de plantas sobre o bajo rasante pueden intentar deducirse analizando todos los registros de una misma dirección postal.

Lamentablemente, existen variables que no están disponibles en Catastro ni en otro tipo de fuentes alternativas. En función de la naturaleza de cada una de estas variables podemos adoptar diferentes soluciones. Por ejemplo:

- Para las variables cuarto de aseo y baño o ducha se puede asumir que están presentes en el 100% de las viviendas o edificios y no parece grave dejar de investigar con carácter general esta variable, pues los porcentajes recogidos en el Censo de 2011 fueron cercanos al 99,5% en casi todos los ámbitos territoriales. Quizá tenga sentido mantener una investigación de campo específica solamente en aquellos lugares donde pudieran darse problemas o asociada al estudio de alojamientos no convencionales.

- Para la variable accesibilidad podemos utilizar la información del Censo 2011 y, adicionalmente, asumir que esta característica es obligatoria para los edificios

de nueva construcción en virtud del Real Decreto 173/2010 de 19 de febrero por la que se modificó el código técnico de edificación.

-Para las variables calefacción, acceso a Internet, sistema de suministro de agua, número de habitaciones, ascensor en el edificio, garaje, número de plazas del garaje, gas, tendido telefónico, agua caliente central y evacuación de aguas residuales aparentemente no disponemos de una fuente de información que nos proporcione información con carácter exhaustivo.

Algunas de estas variables (como por ejemplo la disponibilidad de garaje) se pueden encontrar a partir de la información de la inspección técnica de edificios. De todos modos, según el artículo 21 del RD-Ley 8/2011 de 1 julio la inspección técnica de edificios se debe realizar a "todos los edificios destinados preferentemente a uso residencial" situados en municipios de más de 25.000 habitantes (300 aproximadamente) y que cuenten con más de 50 años de antigüedad. El tamaño del municipio y la antigüedad pueden variar según normativa de las CCAA e incluso ordenanzas municipales. Por tanto, no se tienen garantías de disponer de esta información en todos los casos.

4.6 CARACTERÍSTICAS DE LOS EDIFICIOS

Algunas de las variables que se han mencionado antes son propias de la vivienda, otras son propias del edificio. El fichero censal de viviendas está construido en torno a la vivienda como unidad mínima de información. No se dispone en la actualidad de un fichero actualizado de edificios si bien se tiene un censo de edificios de 2011; en principio no es un objetivo prioritario la actualización del censo de edificios o la construcción de un fichero censal de edificios para 2021 puesto que la información se puede referir a las viviendas.

No obstante no se renuncia a la construcción de tal fichero censal de edificios. La solución ideal consistiría en disponer de un fichero censal de edificios con un identificador que permitiera asociar a cada vivienda su edificio, de tal manera que las características del edificio se pudieran asociar a todas las viviendas contenidas en el tales como la accesibilidad, existencia de ascensor, plazas de garaje, etcétera.

4.7 VIVIENDAS, LOCALES Y HUECOS EN EL CENSO DE 2021

Tradicionalmente en los censos se utiliza **el concepto de "hueco" (también llamado "inmueble") para englobar tanto a las viviendas como a los locales.** Desde el punto de vista de un censo de población el concepto de vivienda es distinto del concepto utilizado con carácter general.

Un inmueble que se ha construido con fines residenciales (es decir, como vivienda) puede ser a efectos de un censo un local. Por ejemplo imaginemos una oficina o una consulta médica situada en una planta intermedia de un edificio de viviendas. Desde el punto de vista del Censo de Población y Viviendas ese "hueco" o inmueble no se considera una vivienda sino un local.

La clasificación de los inmuebles o huecos en viviendas y locales requiere por tanto conocer cuál es el uso que se hace de él en un momento dado. Con los registros administrativos disponibles actualmente podemos determinar el número de inmuebles que originalmente fueron construidos como viviendas pero no cuántos de ellos se están utilizando como locales. Por esa razón los datos de viviendas del censo de 2021 no serían estrictamente comparables con las cifras de viviendas de censos anteriores sino con los inmuebles calculados en censos anteriores y también publicados

4.8 ALOJAMIENTOS

Es uno de los elementos que actualmente es más difícil de extraer de los registros administrativos.

De acuerdo con el actual Reglamento censal hay que aportar información tanto del número de alojamientos existentes a nivel municipal como de las características de las personas que viven en ellos.

En el padrón es difícil identificar los alojamientos, salvo tratamientos específicos en algunos ayuntamientos, que identifican chabolas, por ejemplo, pero es una tarea muy manual y poco fiable.

Todavía se encuentra en fase de estudio la posible incorporación de fuentes locales o de otros ámbitos y la posibilidad de complementar la información con trabajos de campo previos al censo.

Se dispone de información de muchas fuentes, por ejemplo de los alojamientos recogidos durante el recorrido del Censo de 2011 para analizar los municipios susceptibles de mayor investigación.

5 Establecimientos Colectivos

5.1 LA POBLACIÓN RESIDENTE EN ESTABLECIMIENTOS COLECTIVOS EN EL CENSO DE 2011

No hay que olvidar que si bien la mayor parte de la población (el 99% en el último censo) reside en viviendas familiares, hay otra parte de la población que reside en las denominadas viviendas colectivas. Y también es preciso cuantificarlas y obtener las características sociodemográficas de estas personas.

Debido a las características específicas de los establecimientos colectivos (hoteles, residencias de personas mayores, cuarteles, prisiones, conventos...) éstos se investigaron en el censo de 2011 mediante una operación de campo independiente.

La Encuesta de Colectivos formó parte del conjunto de operaciones que constituyeron los Censos de Población y Viviendas 2011. Así como los datos de edificios e inmuebles se obtuvieron del recorrido del Censo de Edificios y la información correspondiente a las viviendas familiares se obtuvo a partir de la encuesta a una muestra de hogares, la información correspondiente a los colectivos se obtuvo a partir de la Encuesta de Colectivos.

La Encuesta de Colectivos utilizó un marco obtenido de distintas fuentes, aprovechando información disponible en el INE, tanto del censo anterior, como de diversas encuestas (por ejemplo la Encuesta de Alojamientos Turísticos o la Encuesta de Discapacidad EDAD-08) y de directorios facilitados por organismos externos, como el Ministerio de Defensa.

Sobre el directorio previo de establecimientos colectivos se hizo una investigación exhaustiva para todos los tipos de establecimiento excepto los hoteles, pensiones y similares, que se investigaron mediante muestra. Dentro de cada colectivo se entrevistaba a una muestra de personas. Se utilizó un factor de elevación para los centros para dar tratamiento a la no respuesta y un factor de elevación para las personas.

Los cuestionarios fueron enviados por correo ordinario y se ofrecieron múltiples canales para su devolución: Internet, correo postal, teléfono y fax. Constaban de una parte a cumplimentar por el propio establecimiento y de los cuestionarios individuales para las personas seleccionadas.

La encuesta permitió determinar el número de establecimientos colectivos en la fecha de referencia censal y su tipo, así como el número y las características demográficas básicas de las personas que residían en ellos. Estas variables y su nivel de desagregación, menor que en el caso de la población en viviendas familiares, están recogidas en la reglamentación desarrollada por la Unión Europea a excepción del tipo de establecimiento, que se incluyó atendiendo exclusivamente a necesidades de usuarios nacionales.

Hay que mencionar que en el censo de 2011 se introdujo una importante novedad a la hora de censar la población residente en establecimientos colectivos. Se consideró residente en un establecimiento colectivo aquella persona que lleva residiendo o espera residir un año en él, independientemente

de su situación de empadronamiento, pero también se les preguntó a las personas por esta circunstancia. Así, se obtuvo que solo el 54% de las personas residentes en colectivos estaban empadronadas en ellos, destacando las prisiones en las que apenas un 10% de la población residente está empadronada.

Dado que el censo de población se construyó a partir de un tratamiento de ficheros que a su vez derivaba del padrón, las personas residentes en viviendas familiares se contaron donde figuraban empadronadas. Así, las personas que residiendo en establecimientos colectivos no estaban empadronadas en ellos, se contaron como residentes en las viviendas familiares donde figuran empadronados para evitar contarlos doblemente.

5.2 PLANTEAMIENTO PARA EL CENSO 2021

Se puede partir, como planteamiento inicial, de considerar como población residente a la población registrada. Esto vale para la población residente en viviendas familiares y también para la población institucionalizada de tal manera que en el censo 2021 se renunciaría a la doble clasificación de población residente y población empadronada en establecimientos colectivos. Se tomaría como población residente a la población empadronada en establecimientos colectivos.

La principal dificultad reside en conocer cuáles de las direcciones postales en el fichero censal de viviendas o de personas son establecimientos colectivos. Esta información no está del todo depurada a día de hoy por lo que la línea principal de trabajo para los próximos meses consiste en cruzar registros de población institucionalizada y fichero censal de personas. Estos registros pueden ser prisiones, conventos, cuarteles, residencias, pisos tutelados, hoteles y cualesquiera otros tipos de establecimientos donde pueda residir población.

6 Incorporación de nuevas variables no investigadas en censos anteriores

Un Censo de Población y Viviendas clásico se basa en el cuestionario como instrumento de recogida de información. El hecho de que los cuestionarios se envíen a todos los hogares y que por tanto deban ser cumplimentados por los informantes sin ayuda de un entrevistador limita enormemente el contenido de la investigación.

La principal limitación a la hora de diseñar cualquier cuestionario censal consiste en que debe ser cuanto más corto y más fácil de entender mejor, porque con ello se conseguirá una mejor calidad de respuesta.

En el caso de un censo que utilice en gran medida registros administrativos esta limitación desaparece; por tanto, la investigación puede abarcar conceptos difíciles de incluir en un cuestionario censal y en muchos casos muy demandados. Por ejemplo, nos estamos refiriendo a la información relativa a la discapacidad, si bien es cierto que la discapacidad, tal como se recogería en un cuestionario ofrece datos distintos a los que se obtienen de registros administrativos. Se pretende incorporar información sobre discapacidad y dependencia en el censo de 2021 si bien todavía no se ha iniciado el análisis de la integración de esta información.

Otros elementos que pueden incorporarse un censo como éste son datos de salud o datos más detallados sobre educación o trabajo.

Para el censo 2021 se contempla también la posible incorporación de información, no a nivel individual, sino agregada por secciones censales. El ejemplo más sobresaliente es el de nivel de ingresos. No se puede obtener para el censo información individualizada del nivel de ingresos a partir de fuentes tributarias pero sí es posible incorporar información sobre el nivel de ingresos medios u otros indicadores de ingresos (como pueden ser datos de desigualdad) a nivel de sección censal. Esta información de contexto sobre la población puede ser extremadamente útil como aproximación a la condición socioeconómica, algo muy demandado por los usuarios de censos anteriores

Otro elemento muy novedoso que se está analizando de cara al censo 2021 es la posible incorporación de información sobre movilidad a través de la posición de los teléfonos móviles, como se menciona en el apartado 3.3.6.

7 Evaluación de la calidad. Operaciones de campo

7.1-¿POR QUÉ TIENE MAYOR CALIDAD UN CENSO QUE UTILIZA MASIVAMENTE REGISTROS?

Existe unanimidad entre la comunidad de productores de información censal en considerar a los países nórdicos como los más avanzados en este ámbito. Todos ellos realizan desde ya varias ediciones un censo basado en registros y cada vez son más los países que dan el salto a este modelo.

En España, la opción de un censo que se apoya en registros parece a día de hoy mejor que cualquiera de las alternativas basadas en la recogida directa por diversos motivos.

Hay que tener en cuenta que, a partir de los distintos registros, se dispone de información de casi todas las variables que se estudiaron en el Censo de 2011. Además, la gran mayoría de los registros disponen de identificadores de calidad que permiten la posibilidad de realizar enlaces entre las distintas fuentes con poco margen de error.

Para algunas variables se obtiene un nivel de detalle muy grande con un coste muy pequeño si lo comparamos con una metodología tradicional.

El censo que utiliza en todo lo posible registros elimina, o al menos atenúa, algunas de las desventajas más importantes de un censo basado en la toma directa de datos por parte de los ciudadanos, como son:

- El informante del cuestionario no siempre conoce todas las respuestas. Muchas veces es otra persona la que responde por él, esto se conoce como un *proxy* o informante indirecto. Es posible que esta persona no conozca de forma detallada algunas preguntas (como por ejemplo las relativas a variables económicas).
- Los resultados basados en auto-clasificación (ocupación, relación con la actividad) muchas veces no reflejan del todo bien la realidad.
- El cuestionario resulta difícil de rellenar para muchos informantes, lo que propicia errores de todo tipo o información incompleta.

Además, la realización de un censo tradicional o basado en encuestas se enfrenta a otros problemas que no están presentes en un censo como éste, o lo están en menor medida entre los que cabe destacar la dificultad sistemática para encontrar ciertos tipos de personas. Por ejemplo, la población extranjera siempre es más difícil de censar.

El censo basado fundamentalmente en registros pasa a ser una labor transparente para la sociedad. El INE frecuentemente recibe quejas de los ciudadanos seleccionados para realizar encuestas en las que se le reclama que no se debería preguntar por cosas **que "la Administración ya conoce"**. El censo así diseñado es también, la plasmación del esfuerzo por no molestar a los

ciudadanos con información que la Administración ya tiene, aunque no resulte nada fácil de tratar en muchos casos por dificultades técnicas de todo tipo.

Por último, entre las ventajas que proporciona esta estrategia destaca la posibilidad de obtener información con una periodicidad mayor que la decenal pudiendo llegar a ser anual en determinadas situaciones. El esfuerzo de disponer de datos administrativos puede dar frutos casi con la frecuencia que se desee, y casi con mismo coste, mientras que un censo tradicional apenas el cabe plantearlo cada diez años.

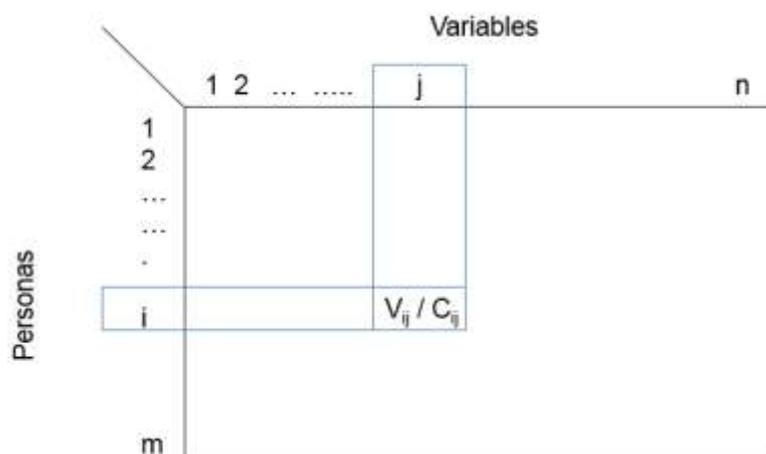
En fin, la conclusión a la que se llega por parte del INE es que el censo permitirá encontrar casi toda la información que figura en los cuestionarios, que lo hará con un nivel de calidad que a día de hoy es superior y que abrirá un enorme campo a la estadística pública. No debemos olvidar que, además de la información que se busca recoger para 2021, los registros pueden proporcionar en el futuro mucha información difícil de recoger en cuestionarios (salud, ingresos, discapacidad, ocio). El censo 2021 va en línea con lo que la sociedad espera y demanda mayoritariamente de una oficina de estadística en la actualidad.

7.2 UN ANÁLISIS DE LA CALIDAD DEL FICHERO CENSAL A PARTIR DE LAS FUENTES

El fichero censal se construye asignando para cada individuo y cada variable un valor que se obtiene a partir de las distintas fuentes e implicadas. Así, para el individuo i , el valor que toma la variable j lo denominamos V_{ij} . Pero para cada individuo y cada variable podemos también asignar unos valores C_{ij} (que denominamos *indicador de calidad*) con la que se ha construido el valor V_{ij} .

El fichero censal, por tanto, está compuesto por una dupla (V_{ij}, C_{ij}) para cada individuo y cada variable, como figura en el esquema siguiente.

Figura 5. El fichero censal de personas. Duplas de valores.



La forma en la que se construye este indicador de calidad tiene que ver con la fiabilidad que se asigna a cada fuente para cada variable. Por ejemplo, la variable “*nivel educativo alcanzado*” se le asigna a cada individuo a partir de un algoritmo en el que intervienen distintas fuentes.

Si para la persona i el nivel educativo alcanzado se ha construido a partir de una fuente directa del sistema educativo (por ejemplo, porque dispone de un título universitario) podemos asignar a esa variable para esa persona en concreto un valor $C_{ij}=1$

En caso de que para una persona el valor se haya asignado a partir de una fuente indirecta o de menor calidad, como por ejemplo, porque la única información que disponemos para esa persona es que refiere un determinado nivel educativo en el censo de 2001 o en el formulario de empadronamiento, podríamos asignar en ese caso $C_{ij}=2$.

Si no se dispone de ninguna referencia en ninguna fuente, el dato se habrá construido a partir de imputación, y en ese caso el indicador de calidad sería $C_{ij}=3$.

En resumen, para cada individuo y para cada variable, podemos tener una dupla formada por el valor que toma la variable en concreto y el indicador de calidad correspondiente, que podría tomar los valores 1,2 y 3 en una primera versión.

Un análisis de la matriz de indicadores de calidad nos da una información muy útil sobre la calidad del fichero censal en su conjunto. Sí analizamos por filas (personas) esta matriz podemos ver, para cada individuo, cuál es la calidad de la información construida para él; si analizamos la información por columnas podemos ver, para cada variable, con qué calidad (y a partir de qué fuentes) se ha construido.

7.3 UNA ENCUESTA DIRIGIDA A LA POBLACIÓN Y OTRAS POSIBLES OPERACIONES DE CAMPO

Como se ha mencionado en el apartado anterior, el análisis por filas, es decir, por personas, nos permite identificar aquellos individuos para los cuales la información contenida en registros administrativos pueda no tener suficiente calidad y se esté basando fundamentalmente en imputación.

Se pretende realizar este análisis a lo largo de los próximos meses de forma que se puedan identificar cuáles son los perfiles de los individuos para los cuales el fichero el método censal no daría resultados suficientemente buenos.

Este análisis puede permitir focalizar una encuesta dirigida a completar la información solamente para aquellas personas cuya información en los registros es defectuosa. Un elemento fundamental es el análisis geográfico. En caso de que se identifiquen áreas geográficas que concentran población con esas características (por ejemplo, población extranjera, población en situación de exclusión social) también se podría plantear que la encuesta de población se dedicara específicamente a proporcionar la información censal para esa

población, incluso sustituyendo por completo, para determinadas secciones censales, la información de registros administrativos por los datos recogidos en los cuestionarios.

Otro tanto puede decirse de las viviendas. El fichero censal de viviendas también será objeto de un análisis de calidad similar. Se puede plantear la realización de una operación de campo previa a 2021 para mejorar la información del censo de viviendas en aquellas zonas donde se observe que la calidad de la información basada en datos administrativos no es suficientemente buena o donde se observe una incoherencia entre la información que proporcionan distintas fuentes cómo pueden ser las catastrales, el marco del INE y los datos de consumo eléctrico.

8 Calendario de tareas

Las tareas a acometer están condicionadas por el grado de éxito en la fase actual de cruce de registros. De forma tentativa se pueden establecer los siguientes hitos en el proyecto:

- A lo largo de 2017 se deben realizar los cruces con todas las fuentes disponibles con el fin de generar a primeros de 2018 un primer fichero precensal de población.
- 2018 – 2019:
 - o En el primer semestre de 2018 se debería disponer de una primera versión del fichero precensal de viviendas.
 - o En el primer semestre de 2018 se decidirá sobre los tipos de operaciones de campo para completar la información censal sobre la población.
 - o Se deben construir nuevas versiones mejoradas de los ficheros precensales de viviendas y personas completos. Se deben identificar las posibles fuentes adicionales y decidir el ámbito final del censo en cuanto a variables a estudiar.
 - o Primer semestre de 2019: Realización de encuestas piloto dirigidas a la población
 - o Primer semestre de 2019: Se deben decidir las operaciones de campo a realizar para completar la información censal sobre viviendas
 - o Finales de 2019: aprobación del proyecto censal definitivo
- Año 2020: se realizarían operaciones de campo relativas al territorio para la mejora del marco y por tanto del censo de viviendas y se pondrían en marcha los procedimientos de contratación administrativa necesaria para las operaciones de campo y encuesta de población.
- Año 2021:
 - o Primer trimestre: encuesta dirigida a la población
 - o A lo largo de todo el año 2021 se realizará la última edición de los distintos cruces con registros administrativos con el fin de proporcionar a finales de 2021 resultados preliminares del Censo
- Año 2022: Publicación los resultados completos el censo tanto para fines nacionales como para satisfacer las demandas de información internacional.

Su opinión nos interesa...

Por favor, si tiene cualquier duda, sugerencia o comentario aclaratorio respecto al contenido del próximo Censo, envíenos un correo a la dirección censos2021@ine.es.