

## EL CENSO DE POBLACIÓN Y VIVIENDAS DE 2021 BASADO EN REGISTROS ADMINISTRATIVOS: UN GRAN PASO ADELANTE EN EL CONOCIMIENTO ESTADÍSTICO DE LA POBLACIÓN EN ESPAÑA

ANTONIO ARGÜESO

*Subdirector General de Estadísticas Sociodemográficas  
INE*

[antonio.argueso.jimenez@ine.es](mailto:antonio.argueso.jimenez@ine.es)

ORCID iD: 0000-0003-4647-8432

**Cómo citar este artículo / Citation:** Juan López Doblas, Aina Faus Bertomeu y Rosa Gómez Redondo. 2021. "El censo de población y viviendas de 2021 basado en registros administrativos: un gran paso adelante en el conocimiento estadístico de la población en España". *Revista Internacional de Sociología* 79(1):e181. <https://doi.org/10.3989/ris.2021.79.1.19.xxx>

### RESUMEN

Tradicionalmente, el Instituto Nacional de Estadística (INE) ha elaborado los censos de población y viviendas basados en visitas a los hogares como la mejor fotografía posible de la población. Pero la información de los censos ya está en los registros administrativos, luego, ¿para qué preguntarla en los hogares? Recorrer todo el territorio para censar la población es una fuente de errores de todo tipo. Obtener los datos de registros administrativos es una labor muy compleja y un verdadero reto, pero estos se usan desde hace ya un par de décadas para producir estadísticas en muchos ámbitos y se observa que, actualmente, ofrecen mejor calidad que las encuestas. El censo de población no es una excepción. No estamos ante ningún salto al vacío, sino en un proceso que lleva madurando en el INE más de veinte años y que está suponiendo ya una mejora cualitativa en el conocimiento que tenemos de la población en España.

### PALABRAS CLAVE

Demografía, censo, registros administrativos, estadísticas de población.

## THE 2021 POPULATION AND HOUSING REGISTER-BASED CENSUS: A GREAT STEP FORWARD IN THE STATISTICAL KNOWLEDGE OF THE POPULATION IN SPAIN

**Copyright:** © 2021 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

### ABSTRACT

Traditionally, The National Statistics Institute (INE) has carried out the population and housing censuses, based on door-to-door enumeration, as the best feasible photograph of the population. But census information can be already found in the administrative registers, so why asking the households? Going through the entire territory to census the population is a source of errors of all kinds. Obtaining data from administrative records is a very complex task and a real challenge, but for a couple of decades they have been used to produce statistics in many areas and we have concluded that they offer today better quality than surveys. The population census is no exception. We are not facing any leap into the void but rather a process that has been maturing at INE for more than twenty years and that is already bringing a qualitative improvement in our knowledge of the population in Spain.

### KEYWORDS

Demography, Census, Administrative Registers, Population Statistics

## INTRODUCCIÓN

El principio de incertidumbre de Heisenberg establece que es imposible conocer simultáneamente y con precisión arbitraria dos variables que definen una partícula. Así, si queremos conocer con precisión cuál es la posición de un electrón en un momento dado, no podremos precisar su cantidad de movimiento y viceversa. Si diseñamos un experimento para conocer con mucha precisión una de ellas forzosamente tendremos mayor imprecisión en la otra. No es una cuestión técnica, sino un principio de la naturaleza y un pilar sobre el que se basa la física cuántica.

Con las encuestas sociales ocurre algo parecido. Si queremos precisar una información social, el método que diseñemos para medirlo interferirá en el resultado, por muy diversas razones, pero ninguno podrá tener precisión absoluta. Unos métodos podrán dar mejores resultados sobre unas variables, pero serán peores para otras. El trabajo de los estadísticos es buscar cuál es el mejor método, el que mayor información y menor error nos proporciona sobre la población, no el más barato, ni el más rápido. A eso nos venimos dedicando los productores de estadísticas sociales desde siempre. Nada ha cambiado hoy cuando consideramos que el mejor método para el censo es basarlo en registros administrativos. ¿Por qué el mejor método era la recopilación de cuestionarios puerta por puerta y ahora no lo es? Sobre todo porque la información que tenemos sobre las personas es infinitamente superior ahora que hace veinte años, aunque es difícil encajar las piezas. En segundo lugar también hay que considerar que la recogida directa de información de los hogares es más difícil ahora, tanto en España como en todo nuestro entorno.

El documento que suscita este debate es un resumen de respuestas de 30 personas (usuarios y productores) que contiene muchos comentarios donde quizá prime (desde luego, se recoge extensamente) una cierta percepción pesimista por parte de algunos usuarios. Es difícil, en el corto espacio de este artículo, rebatir todos los elementos que se mencionan; nos intentaremos centrar solamente en algunos de ellos que consideramos principales, pero sobre todo trataremos de dar argumentos para convencer del salto cualitativo hacia adelante que supone este nuevo sistema.

Priorizar el uso de registros no es algo que se esté dando solo en España, buscando “modernidad”, ni ahorros, ni mucho menos atendiendo a presiones de organismos internacionales. Todos los países del mundo se afanan por basar las estadísticas en registros administrativos simplemente porque es mejor, y lo es tanto más cuanto más y mejores registros se tienen. Pero no todos los países están en la misma situación. España se sitúa entre los países líderes en esta materia porque tiene buenos registros admi-

nistrativos, porque tiene un padrón que sirve de esqueleto y porque tiene leyes que permiten acceso a estos registros. Son tres condiciones fundamentales que solo unos pocos países consiguen cumplir.

## LOS “EXTREMOS DE LA COLA” EN UN CENSO CLÁSICO Y EN UN CENSO BASADO EN REGISTROS

Un primer elemento en defensa del nuevo sistema es el propio universo poblacional que nos brindan un censo clásico y uno basado en registros. Pensemos en la población extranjera. Podemos usar las evidencias de las propias encuestas del INE. Lo más cercano (de hecho, casi similar) a un cuestionario censal es el de la Encuesta Continua de Hogares (ECH), que el INE realiza desde 2013, con un tamaño muestral muy considerable (superior a 50 000 hogares, es decir, 120 000 adultos al año). Esta encuesta, como ocurre con otras del INE, se realiza mediante muestreo probabilístico seleccionando viviendas ocupadas, no personas. Si la población extranjera se sitúa a 1 de enero de 2019 en el 10,7 %, tomando datos de la ECH de los años 2018 y 2019, es decir, tomando las casi 246 000 personas entrevistadas en esos dos años, deberíamos esperar recoger cuestionarios de unos 26 300 extranjeros. Pero solo aparecen 16 200, es decir, si diéramos por buenos estos datos, la población extranjera en España sería un 61,5 % de la que figura empadronada. Se perderían casi dos millones de extranjeros. Para todas las nacionalidades extranjeras encuestamos menos población de la esperada, pero para algunas en concreto las diferencias son muy acusadas, con diferencias por sexos muy grandes. Así, solo encontramos en la ECH el 41 % de senegaleses o el 50 % de chinos de los que deberíamos encontrar, de acuerdo a los que figuran empadronados.

Este tipo de problemas ocurre en encuestas y también en los censos en todo el mundo, donde se dejan sin entrevistar porcentajes crecientes de población. Los datos observados por encuesta son mucho menos consistentes con otras fuentes que los que ofrece el Padrón. ¿Qué pasaría si el INE publicara un número estimado de niños para para ciertos grupos de edad y nacionalidades que fuera muy inferior al de matriculados en los colegios? Los datos administrativos nos están mostrando realidades insoslayables.

Uno de los principales recelos que se esgrimen contra el uso de registros administrativos es el temor de que la información administrativa se aleje de la realidad en los “extremos de la cola”, es decir, en los casos de personas en situaciones de marginalidad o de las muy ricas, porque supuestamente su información no aflora correctamente en los registros administrativos y sí lo hace en las entrevistas a los hogares. Pero precisamente uno de los elementos

que más claramente nos orienta al uso de registros es mejorar la información para esas colas, toda vez que la recogida directa ofrece claras desventajas. En el caso de encuestas sobre ingresos, el fenómeno está ya muy estudiado (Vega y Méndez 2014).

Aunque son muchos los ejemplos que se podrían dar, pensemos en un caso extremo. Quizá el mayor foco de marginalidad en España sea la Cañada Real, en la zona sudeste de Madrid, una barriada que se extiende por varios términos municipales. Un censo de población clásico logra pobres resultados en estos barrios de extrema marginalidad. Requiere que las visitas de los agentes estén apoyadas por servicios sociales para evitar recelos (a veces, los agentes son acompañados por la policía) pero aun así adolece de mucha falta de respuesta.

Una sección censal en concreto concentra la mayor cantidad de población y los peores indicadores. En el censo de 2011, se visitó íntegramente esa sección censal durante el “primer recorrido” (para censar todos los edificios y las viviendas) y apenas se censaron 320 viviendas y se estimó una población de 1.525 (si bien esto se hizo por muestreo). En cambio, en 2011, figuraban 4.016 personas empadronadas allí, en 993 grupos humanos (hogares, en terminología censal). De ellos sabemos que el 51 % son extranjeros, casi exclusivamente de Rumanía y Marruecos. Podemos observar las coordenadas geográficas tomadas por los agentes censales y superponer con ortofotografías, y se observa que la mayoría de las viviendas de esa zona no se lograron entrevistar.

Pero podemos cruzar esos residentes según padrón con datos tributarios. Esto ya se ha hecho dentro del proyecto de estadística experimental “Atlas de Distribución de Renta de los Hogares” (ADRH) (Instituto Nacional de Estadística 2020a). El ADRH es un producto “satélite” del censo de 2021 que ya se ha empezado a publicar en 2019 (¡el censo de 2021 ya está dando frutos!). Encontramos datos de ingresos para un 82 % de estos hogares. Así sabemos que su ingreso medio por hogar en 2017 fue inferior a 10 500 euros netos al año, frente a los 28 400 € de la media nacional; el 63,9 % de los hogares está por debajo del umbral de pobreza y solo un 52 % de los ingresos proviene de salarios o pensiones, muy lejos de la situación media del país.

¿Es esa información la que de verdad refleja la realidad social de ese lugar? Seguramente no, habrá múltiples limitaciones y sesgos en nuestro conocimiento sobre quiénes viven ahí o cuáles son sus condiciones de vida, pero, ¿es la *verdad administrativa* de inferior calidad a la que nos da un censo clásico? La respuesta parece clara: es abrumadoramente mejor que la que obtenemos por observación directa para una zona deprimida como esa, en la cola inferior de los ingresos.

Otro tanto ocurre en la cola superior de distribución de renta: los hogares ricos son tradicionalmente los más difíciles de encuestar. Los registros administrativos pueden, de nuevo, ofrecer información imperfecta, pero invitamos al lector a recorrer el ya mencionado ADRH y comprobar por sí mismo los resultados para las zonas ricas.

## LA REALIDAD MEDIDA POR CUESTIONARIOS Y LA REALIDAD ADMINISTRATIVA

Un segundo grupo de recelos, por parte de los defensores del censo clásico, consiste en dudar que los registros ofrezcan información válida sobre la población en general, dados los sesgos contenidos en ficheros no diseñados para fines de análisis social o demográfico y que, por tanto, lo que se va a ofrecer es una realidad “administrativa”.

Para discutir esa afirmación, conviene recordar qué tipo de información sobre las personas recoge un censo, qué variables concretas intervienen y cómo se combinan los registros para ofrecer información sobre estas variables.

Pensemos en las variables demográficas que describen los censos, como, por ejemplo, el estado civil. Ningún registro incluye el estado civil de toda la población adulta. Pero una decena de fuentes contienen esa variable en distintos momentos, para distintos grupos poblacionales y con distintos valores: datos tributarios (quienes realizan la declaración de la renta declaran su estado civil); datos de registro civil, porque los matrimonios o los divorcios (no todos) se registran datos de seguridad social, etc. Aquí es donde interviene la lógica, usada en este tipo de censos, de considerar fuentes ganadoras.

En la crítica al censo se hace a veces una interpretación de este término como si el INE tomara unas arriesgadas decisiones basadas en apriorismos o en consultar una bola de cristal para decidir que un registro es mejor que otro. Es algo mucho más sencillo: si de una persona tenemos un dato registral de un matrimonio y otro de un divorcio posterior, la fuente ganadora es la del divorcio, o viceversa cuando es posterior el matrimonio. Si lo primero que hay de ella es una declaración de la renta que dice que está casada, la consideramos así y gana como fuente la tributaria; pero si esa persona luego declara percibir prestación de viudedad pasará a ser viuda. Esa es la lógica del uso de fuentes ganadoras. La fuente tributaria no está por encima del Registro Civil a priori, ni viceversa, sino que son las situaciones concretas las que determinan la fuente, mediante algoritmos de decisión que, en general, consisten en aplicar el sentido común.

Otro tanto ocurre con el nivel educativo de la población, o con la situación laboral. Ningún fichero nos ofrece este dato para toda la población. Y, sin embar-

go, ya desde 2014 hemos sido capaces, combinando más de una decena de ellos, de asignar un nivel educativo a toda la población y poder así publicar, como ya ocurre a día de hoy, estadísticas de defunciones o de nacimientos por nivel educativo (de los padres, en el segundo caso) sin pedir esa información en el boletín de defunciones (INE 2020b). Esa es una ventaja indudable de este método censal: ya está dando frutos desde hace varios años.

Pero acometamos la variable que quizá sea la mayor fuente de críticas: la composición de hogares y las relaciones de parentesco entre sus miembros. Volviendo a tomar datos de la ECH para un año completo, podemos comparar qué composición de hogar se declara en el cuestionario y cómo figura ese mismo hogar en el padrón. Se ha cotejado el padrón con la ECH durante varios trimestres y se comprueba que, de los hogares que responden a la encuesta (que rondan el 70 % de los seleccionados originalmente), en el 83 % de casos aproximadamente la composición es exactamente la misma: son los mismos miembros que figuran empadronados. Hay un 17 % de hogares con situaciones de todo tipo, en los que el hogar “padronal” no coincide con el declarado en la entrevista. Todos podemos argumentar sobre casos que conocemos de personas empadronadas donde no viven. Pero no podemos afirmar (el *principio de incertidumbre*) que el cuestionario tradicional vaya a recoger esa información mejor. De hecho, en muchas encuestas se comprueba la tendencia a declarar menos miembros, quizá para terminar el cuestionario antes, lo que da lugar a tamaños medios de hogar a veces inverosímiles, por ejemplo, en algunas encuestas por Internet. No tenemos por qué suponer que el hogar declarado en un cuestionario es más representativo de la realidad. Tengamos siempre eso en cuenta: no estamos comparando el registro administrativo con la realidad (esa no la podemos conocer) sino con la que nos ofrece otro método como es el censo clásico, y puede estar tan alejado de ella como el anterior o más.

En cuanto a los parentescos, efectivamente los registros administrativos no van a poder determinar bien las relaciones de hecho en parejas; es esa una de las razones que nos lleva a complementar el censo con una encuesta dirigida al 1 % de la población, la ECEPOV-2021, que permitirá efectuar procedimientos mejores de imputación. Pero la imputación se aplica a muy pocos hogares.

## SE PUEDE CONSTRUIR INFORMACIÓN ORDENADA A PARTIR DE REGISTROS ADMINISTRATIVOS

Podemos hablar de un tercer tipo de recelo, el de sustituir la información ordenada que ofrece un censo clásico (cuyo instrumento, no lo olvidemos, es un

cuestionario) con la nebulosa difícil de analizar de registros administrativos “de su padre y de su madre”. Se critica que el censo solo investiga lo que está en los registros, olvidándose de lo demás. Pero la realidad no es así. La construcción de un censo, basado o no en registros administrativos, parte del mismo instrumento: un cuestionario, una lista de variables a responder. El proceso en 2021 consiste, como en el censo de 1991, el de 2001 o el de 2011, en intentar rellenar un cuestionario, pero sin pedirle a la población que lo haga, sino “rellenando el cuestionario por ellos” buscando en distintas fuentes.

Las variables en un censo basado en registros administrativos son, como mínimo, las mismas que en un censo tradicional, con la ventaja de que se pueden añadir otras; pero los registros no determinan la materia a investigar. Es lógico: en el ámbito de la UE, todos los países tenemos que generar el mismo producto censal (así lo dicta un Reglamento europeo), y unos lo haremos con unos métodos y otros con otros. Solo una docena de países (el más poblado de los cuales es España) hemos avanzado tecnológicamente para hacerlo mediante registros. Desde septiembre de 2019, se encuentra publicado ya un proyecto técnico bastante detallado (INE 2019). En él se describe el método que se seguirá variable por variable.

## EL USO DEL BIG DATA EN EL CENSO 2021

Por último, figura un cuarto tipo de recelo. El documento que da pie a este debate concentra una gran atención en el uso de *big data* en el censo, asumiendo que el proyecto censal va a apoyarse en *big data* de una forma muy distinta a lo que de verdad se plantea, y buena parte de las reticencias se basan en esta circunstancia.

Afortunadamente, en este terreno no tenemos que hablar del futuro, sino de algo ya realizado, porque toda la aportación del *big data* al proyecto censal se ha producido y se ha publicado ya (INE 2020c).

El INE contemplaba la posibilidad de utilizar datos de posicionamiento de teléfonos móviles para completar una parte de la información que no se recoge bien en los registros administrativos, la movilidad cotidiana. Por tanto, más allá del término demasiado general de *big data*, lo que el INE hizo fue exclusivamente estudiar una fuente (la telefonía móvil) para una variable (la movilidad cotidiana). Ese trabajo, tras muchas vicisitudes, por fin en 2019 encontró una solución basada en datos de los tres principales operadores. No se toma una información ya precocinada o sesgada con fines comerciales, sino que los operadores hacen el mismo trabajo para el INE basado en directrices metodológicas muy claras y detalladas. El

trabajo ya se ha publicado en 2020 con todo detalle. Dado que es una información bastante estructural, se ha adelantado al censo 2021 y no se prevé una publicación nueva en 2021. El censo de 2021 sigue dando frutos. Esta es la aportación de *big data* para el censo. Que juzgue el usuario sobre sus limitaciones y sus virtudes. Conviene advertir que nunca antes se ha tenido información sobre población cotidiana o estacional como se dispone aquí, a partir de matrices de origen-destino tan detalladas. El INE es la referencia a nivel europeo en esta materia, y se ha presentado el proyecto en multitud de organismos internacionales suscitando el interés (y los elogios) de los colegas de muchos institutos de estadística.

Se critica también el salto al vacío, al “abismo”, la falta de planificación y de reflexión. Pero no se tiene en cuenta que el INE viene cruzando y explotando registros administrativos desde hace más de 20 años y muchas estadísticas se basan ya en ellos, con abundantes descripciones metodológicas en su página web. Además, en concreto para el censo de 2021 ya se ha hecho un ensayo general casi completo, el fichero precensal de 2016, que nos llevó a la conclusión de que el mínimo exigido por el Reglamento Europeo, es decir, hacer un censo muy similar al de 2001, ya se cumplía entonces. No se trabaja sin red, sino sobre terreno seguro.

Desde entonces, las principales mejoras en la investigación han venido por el lado de la georreferenciación y el censo de viviendas. Esto nos permite asegurar que la cantidad y calidad de información que se va a publicar estará muy por encima de la

de los censos anteriores. El INE centrará todos los esfuerzos en hacer una difusión muy detallada de la información, que superará con creces la ofrecida para el censo de 2001, y en proporcionar toda la meta-información necesaria para interpretar mejor los resultados.

## CONCLUSIONES

Un censo así no es un punto final de unos viejos tiempos en los que teníamos mejor información de la sociedad, ni una forma de “salvar los muebles” en tiempos de crisis, sino el principio de un salto cualitativo hacia un nuevo modelo con mucha más información, longitudinal, más detallada y más oportuna.

La disponibilidad creciente de datos administrativos sitúa al sistema estadístico en un estado de ebullición. Hay mucho por hacer en muchos frentes. Este crecimiento de la cantidad y calidad de oferta de datos ya se ha iniciado. De hecho, ya se está viendo con las estadísticas de condiciones de vida, el atlas de renta, los datos de movilidad, de migraciones o de stock de población que se ofrecen ya con muchos detalles y casi en tiempo real, pero el censo va a permitir sistematizar y geolocalizar mucho mejor toda esta información.

Quizá sea pronto para congratularnos (estamos hablando de algo que aún no ha cristalizado), pero sí al menos invito a los usuarios a enfrentarse al nuevo censo con unas expectativas razonablemente optimistas. En cambio, no vemos razones de peso para lo contrario.

## REFERENCIAS

- Vega, Pilar y Méndez, José María. 2014. “Comparación de los ingresos del trabajo entre la Encuesta de Condiciones de Vida y las fuentes administrativas”. Documento de trabajo 02/2014. INE. [https://www.ine.es/ss/Satellite?L=es\\_ES&c=INEDocTrabajo\\_C&cid=1259944417750&p=1254735116586&pagename=ProductosYServicios%2FPYSLayout](https://www.ine.es/ss/Satellite?L=es_ES&c=INEDocTrabajo_C&cid=1259944417750&p=1254735116586&pagename=ProductosYServicios%2FPYSLayout)
- Instituto Nacional de Estadística 2020a. *Atlas de Distribución de Renta de los Hogares*. Proyecto técnico. Madrid. Instituto Nacional de Estadística. [https://www.ine.es/experimental/atlas/experimental\\_atlas.htm](https://www.ine.es/experimental/atlas/experimental_atlas.htm)
- Instituto Nacional de Estadística 2020b. “*Estadística de Desfunciones. Método de asignación de nivel educativo, relación con la actividad laboral y ocupación*”. Madrid. Instituto Nacional de Estadística. [https://www.ine.es/metodologia/t20/Nota\\_meto\\_MNP.pdf](https://www.ine.es/metodologia/t20/Nota_meto_MNP.pdf)
- Instituto Nacional de Estadística. 2019. *Censos de Población y viviendas 2021. Proyecto Técnico*. Madrid: Instituto Nacional de Estadística. Madrid. Instituto Nacional de Estadística. [https://www.ine.es/censos2021/censos2021\\_proyecto.pdf](https://www.ine.es/censos2021/censos2021_proyecto.pdf)
- Instituto Nacional de Estadística 2020c. *Estudios de movilidad a partir de la telefonía móvil*. Madrid. Instituto Nacional de Estadística. [https://www.ine.es/experimental/movilidad/experimental\\_em.htm](https://www.ine.es/experimental/movilidad/experimental_em.htm)

**ANTONIO ARGÜESO JIMÉNEZ** es licenciado en Ciencias Físicas por la Universidad Complutense de Madrid y Subdirector General de Estadísticas Sociodemográficas (INE). Fue el coordinador de los Censos de Población en España en 2011 y dirige el futuro censo de 2021, así como las estadísticas sobre la población y condiciones de vida que elabora el INE. Es miembro del grupo de Directores de Estadísticas Sociales en Eurostat (oficina estadística de la Unión Europea). Este grupo coordina las estadísticas demográficas y sociales en Europa.