

# **Mejoras en la declaración de edad a partir de la incorporación de fecha de nacimiento en el Censo 2017.**

## **Resumen (v. larga)**

Como ya concluía Carmen Miró en el Seminario de las Naciones Unidas sobre Evaluación y aprovechamientos de datos censales en América Latina 1959 la evaluación de los resultados de los censos de población es indispensable por un lado para una mejor utilización de los mismos, y por otro para dar una apreciación crítica de los métodos y procedimientos empleados en las etapas de empadronamiento y elaboración de resultados (Miró, 1963), es por esto que existen reglas de consistencia que se aplican en la fase postcensal, que permiten dar cuenta de posibles errores, inconsistencias, o de información faltante.

Siendo la edad una variable demográfica clave, unos de los procedimientos demográficos básicos conocidos al momento de evaluar indirectamente la información censal es el de la declaración de la edad. Estos análisis suelen hacerse por sexo y compararse dentro de la historia censal. En general, la evaluación de la estructura por edad y sexo de la población así como las prácticas de corrección de datos es un tema que se ha abordado extensamente en la literatura demográfica, claro ejemplo lo constituyen el de Chackiel & Macció (1978).

Dependiendo el tipo de análisis que se vaya a realizar y de la información disponible puede utilizarse la edad exacta en años, meses y días, o la edad cumplida como el número exacto de años que una persona tendría en su último cumpleaños. Generalmente para recoger información sobre la edad en los censos de población y vivienda suele preguntarse por la edad cumplida en años al momento del censo, pero en el tiempo se ha ido incorporando la pregunta por fecha de nacimiento.

La incorporación de la fecha de nacimiento plantea distintas formas de procesar los datos, ya que puede ocurrir que coincida o no con la edad declarada, puede que tenga o no mayor omisión de respuestas y puede que tenga o mayor concentración en ciertos dígitos, en fin múltiples de combinatorias que permitan mejorar la calidad de los datos de edad que se ponen disposición en la base de datos definitiva.

El objetivo de este trabajo es exponer el proceso técnico de validación de la edad previo a la entrega de base de datos definitiva y presentar la utilidad de la incorporación de la fecha de nacimiento en el censo de población y vivienda de Chile 2017.

Posterior al levantamiento en terreno, se encuentra la etapa de procesamiento, la cual comienza después de la reversa de material a la empresa encargada de la lectura de los cuestionarios. Se compone de dos etapas, procesamiento primario y secundario, dentro de las cuales se contemplan bases de datos entregables. El procesamiento secundario corresponde a la fase donde se aplican las reglas de consistencia e imputación para luego

tener una base de datos validada, en esta fase trabajan en conjunto 3 unidades del INE: Unida Técnica y Método (UTM), Tecnologías de Información Censo (TI) y el Subdepartamento de Demografía y Estadísticas Vitales.

Como parte del rol del subdepartamento de demografía, correspondía el análisis demográfico de las bases, dar lineamientos técnicos respecto del diseño y aplicación de las reglas y la revisión de programación hecha por TI.

Utilizando el Software Redatam 7, la unidad de demografía programó los tipos de edad que se pueden construir al tener pregunta por edad declarada y fecha de nacimiento, las reglas de validación de la edad, y los indicadores para la evolución: estructura por edad y sexo para las pirámides, índices de masculinidad por edad y el índice de Myers. Todo esto a nivel nacional, regional y comunal (y si se quisiera en posteriores análisis se podrían llegar a niveles menores)

Con la celeridad requerida ante el cronograma de publicación, el uso de Redatam, permitía hacer consultas rápidas a la base de datos, y modificaciones a las programaciones. Esta posibilidad de contar con esta información y este producto, anterior a la base validada, permitió identificar si había errores que venían del terreno, del procesamiento primario o, en vez de errores situaciones y circunstancias que son propias de determinado lugar geográfico.

Se utilizaron 4 tipos de edad: La *edad declarada* proveniente de la pregunta 9 ¿Cuál es su edad en años cumplidos? La *edad calculada* proveniente de la pregunta 9.a mes y año de nacimiento. La *edad validada 1* usa la estructura válida de la edad calculada y completa los ignorados e inconsistentes con la información de la edad declarada. La *edad validada 2* a la inversa de la validada 1, usa la estructura válida de la edad declarada y completa las inconsistencias con la información válida de la edad calculada. En los casos donde no había información en ni en la edad declarada ni en mes y año de nacimiento, se completó con imputación dinámica (hot-deck) considerando la información del parentesco, sexo, edad, información en residencia anterior (categoría aun no nacía), educación, fecundidad y en última instancia la moda comunal.

Con una base de viviendas particulares que representaba el 92% respecto del total de la población proyectada a 30 de junio, se observa que con cualquier validación de la edad los porcentajes de inconsistencias se encuentran dentro de la recomendación internacional (menos del 2%), además se ve que el menor porcentaje (0,58%) de inconsistencias se da en las edades validada 1 y 2. Cuando se observa a nivel regional 3 de las 15 regiones del país tienen un porcentaje levemente superior al 2%, pero dichas regiones se caracterizan por tener un mayor peso de viviendas colectivas propias en lugares donde hay mayor actividad económica minera. Al bajar a nivel comunal, utilizando la edad validada 1, 342 de las 346 comunas tiene un porcentaje de inconsistencias dentro del recomendado.

Desde 1982 Chile posee una “buena declaración de la edad” con un índice de Myers nacional menor a 5. En el censo del Censo 2017, al analizar los índices de Myers para cada validación de edad se observa que si se mantiene la edad declarada se obtiene un valor cercano a 2,5 (muy similar a 1992-2002) en cambio con la incorporación de la fecha de nacimiento, al observar la edad validada 1 el índice de Myers total baja a 0,7. Lo mismo al

observar el nivel regional, todas las regiones utilizando la edad validada 1 tienen la menor concentración de dígitos, lo mismo al bajar al nivel comunal, ya que con la edad validada 1 311 de las 346 comunas tienen un índice de Myers menor 5 es decir una buena calidad. En este nivel se observó que las comunas con los Myers más altos, asociados generalmente a una “mala declaración” se concentran comunas del país que tienen menos de 1000 habitantes y que por sus características (regimiento militar) si es posible que se concentren hombres de una determinada edad (ej. 18 años)

Finalmente, se decide usar la edad validada 1, que además de tener el mejor índice de Myers, recoge mejor los menores de 5 años, especialmente los menores de 1 año, típicamente sub-enumerados. Se observa una pirámide de población acorde con los cambios desde el censo anterior 2002, que al igual como lo vienen revelando las estadísticas vitales muestra una reducción importante de los nacimientos junto con un evidente y acelerado envejecimiento de la población.

**Palabras clave:**

Procesamiento, Validación, Evaluación, Edad.