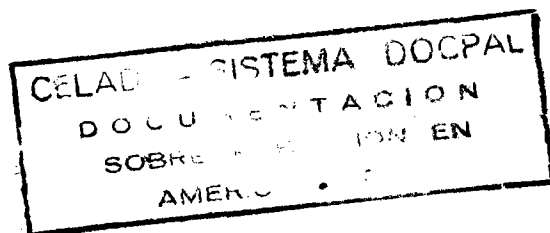


13994.02
(36744)
c2

CONSIDERACIONES SOBRE EL USO DEL MUESTREO
EN LOS CENSOS DE AMERICA LATINA

CELADE

Documento a ser presentado en el Seminario sobre uso del muestreo en las fases preparatorias y en la utilización de los datos censales, organizado por INDEC-Argentina, coauspiciado por la CEPAL y con la colaboración de Chile, CELADE y la Oficina de Estadísticas de las Naciones Unidas.
Buenos Aires, 18-21 abril, 1989.



En este documento se pretende ofrecer algunas observaciones sobre aspectos prácticos del uso del muestreo en los censos de población y vivienda de la región, tanto en cuanto a sus ventajas como a sus limitaciones. Se tratará esencialmente del uso del muestreo para la recolección de los datos. También se considerará la importancia del uso de los datos básicos del censo para la construcción de un marco muestral que apoye las encuestas por muestreo durante el período intercensal. Finalmente se hará referencia a un instrumento que podría ayudar con las tareas de análisis de los datos y el manejo del marco muestral, el paquete REDATAM.

MARCO GENERAL

Las razones para el levantamiento de los censos de población y vivienda son conocidas. Las administraciones nacionales y otros organismos necesitan información sobre las características de la población y sus condiciones de vida, para facilitar la elaboración y realización de sus planes y proyectos de desarrollo social y económico. Así como la información censal tiene varios usuarios, cada usuario tiene sus propias preocupaciones y necesidades de datos, las que no siempre son compatibles.

Los censos normalmente se realizan aproximadamente cada diez años. En países donde los cambios demográficos son rápidos éste es un período excesivamente largo; por lo tanto, es sumamente importante que los datos obtenidos sean de alta calidad, para que al problema de la interpretación de los cambios reales no se agregue un punto de partida equivocado.

Tradicionalmente, la información censal se ha recolectado a través de un recuento de toda la población residente en el territorio de un país. Es un ejercicio costoso, que implica un alto grado de coordinación para asegurar un resultado útil. Además, la mayoría de los censos en América Latina han sido de hecho, lo cual normalmente requiere que el censo se levante en un sólo día. En este caso, se necesita un gran número de personal, y su formación adecuada resulta ser difícil.

Cuando la información requerida se multiplica, estos esfuerzos y costos llegan a ser considerables, necesitando el abandono de muchos elementos de información, o el uso de métodos alternativos de obtención de los datos. Es así que en los últimos censos en varios países se ha aplicado el muestreo para la

recolección de información más amplia de una parte representativa de la población.

Para permitir la recolección de los datos de la población y de la vivienda en un censo, el muestreo puede aplicarse de distintas maneras. El método más aplicado hasta ahora ha sido el uso de dos boletas. La primera contiene las preguntas básicas para las cuales se estima que es preciso obtener la información de toda la población. La segunda boleta, con un mayor número de preguntas que pueden proporcionar información bastante útil aunque se obtenga de sólo una fracción de la población, se aplica a una muestra, la que se elige según algún método de muestreo probabilístico para que los resultados sean aplicables al total de la población.

Cualquiera sea la estrategia adoptada, tiene sus consecuencias, las que analizaremos más adelante. El uso de una sola boleta simplifica la operación censal pero tiene limitaciones estrictas en cuanto a su contenido. En cambio, el uso de una boleta general reducida con una (o varias) boleta(s) extensiva(s) para muestras de la población, evidentemente permite cubrir un número mayor de distintos elementos de información, pero tiene la desventaja de complicar los aspectos logísticos.

NECESIDAD DE BUENOS DATOS BASICOS

El censo en primer lugar sirve para el establecimiento de un inventario general de datos básicos sobre la población en un punto fijo en el tiempo, de la misma manera que los registros vitales sirven para proporcionar información sobre los flujos de la población a través del tiempo. Finalmente, las encuestas por muestreo sirven para estudios en profundidad sobre temas específicos. En países en desarrollo, cuyos registros a menudo son deficientes, tales encuestas a veces sirven también para la medición de los flujos.

La principal preocupación del CELADE por este tema es que los censos sirvan a los países como herramienta que proporcione datos útiles para la formulación de políticas y planes de desarrollo. El CELADE también es un usuario directo de los datos censales como insumos para trabajos de

investigación y apoyo a los países en materias relacionadas con políticas de población y desarrollo social, incluyendo asesoría en la elaboración de estimaciones y proyecciones de la población realizada por todos los países de América Latina y el Caribe.

Desde luego el CELADE tiene un interés especial en el componente demográfico, y en su comparabilidad entre países. Sus actividades en la ronda del 90 se concentran en la asesoría para el diseño de las boletas censales y para el procesamiento electrónico de los datos, debido a que cuenta con una amplia experiencia en estos campos. Sin embargo, su interés se extiende a todos los aspectos de la planeación, de la recolección, del procesamiento, de la evaluación y análisis y del uso de la información que pueda proporcionar un censo de población y vivienda, en particular en lo que refiere a la información demográfica propiamente tal. Su mayor preocupación entonces es la calidad de los datos básicos, sin dejar de desear ver información enriquecida.

En este sentido, ha parecido conveniente que la información básica en un censo se recolecte de toda la población. Es por esta razón que el CELADE ha favorecido el uso tradicional de una boleta única para el universo, salvo para aquellos países donde por razones de su tamaño u otras particularidades se justifique el muestreo. Sin embargo, en el contexto actual, aquellas razones han llegado a ser imperantes en un gran número de países de la región. De aquí en adelante, esta ponencia tratará del caso en que sí se piensa aplicar una u otra forma de muestreo en la etapa de la recolección de los datos básicos.

Cuando se aplica una boleta única al universo, se hace necesario que ésta sea razonablemente breve, incluyendo solamente la información más necesaria. En varios países se ha utilizado o se propone utilizar el muestreo para la recolección de los datos mediante dos boletas. En este caso, hay que decidir el contenido reducido de la boleta universal, y al mismo tiempo elaborar una segunda boleta más amplia. Teniendo en cuenta el objetivo de las boletas censales, parece razonable considerar la inclusión de preguntas en ellas a distintos niveles de factibilidad respecto a los datos que se pudieran recolectar en tal instrumento.

De primera prioridad serían los datos básicos, los que siempre tienen que

obtenerse de toda la población. Macció (1988) identificó este conjunto mínimo de datos como el inventario de tres tipos de unidades básicas: unidades físicas (la identificación geográfica hasta la vivienda), individuos (el número de personas, su sexo y edad, estado conyugal, y tal vez el lugar de residencia actual y en el pasado) y unidades sociológicas (hogares y familias, relación de parentesco con el jefe del hogar). Estos temas y la formulación de las correspondientes preguntas se analizaron recientemente en el documento presentado por el CELADE en el Seminario sobre contenido y diseño de la boleta censal (CELADE, marzo 1989). Estas preguntas de primera prioridad, se incluirían en la boleta básica. Siempre y cuando no haya demasiado preguntas prioritarias, se tomar la decisión de utilizar una sola boleta, no demasiado amplia, y omitir todos los demás temas del censo. De esta manera, no se aplicaría el muestreo para la recolección de los datos censales. Esta decisión normalmente se toma también en el caso de la mayoría de los países de relativamente menor población y/o extensión geográfica.

Sin embargo, el asunto no es tan simple, ya que en segundo lugar casi siempre viene un gran número de temas y preguntas bastante importantes. Son los que, siendo considerados para incluirse en la boleta básica, por distintas razones tienen que ceder su lugar ante otros más esenciales dentro del contenido necesariamente reducido de dicha boleta. De alguna manera, más de estas preguntas hay, menos se pueden incluir en la boleta básica, ya que para satisfacer el mayor número posible de preocupaciones de esta segunda categoría, se necesita una boleta muy amplia para la segunda fase, la de la muestra, con los costos que ello implica.

La tercera categoría incluye a todo tipo de tema o pregunta que nunca alcanzó a ser seriamente considerado para su inclusión en la boleta básica, ya sea por su aplicación reducida u otra razón. Estos temas deben ser tratados en otros momentos y por otras vías, tales como estudios regulares o especiales dentro de un programa de encuestas de hogares por muestreo. A fin de permitir la realización de tales estudios, es importante contar con excelentes datos básicos del censo mismo, para que se pueda elaborar y mantener un marco muestral adecuado durante los subsiguientes años.

En ningún caso se debe tratar de hacer todo el censo por muestreo. Es

decir, cualquier información que tenga la primera prioridad tiene que ser recolectada del cien por ciento de la población. Así la información básica se puede utilizar y analizar desde el nivel nacional hasta las unidades menores de la administración. El intento de obtener la información básica a partir de una muestra, respetando la necesidad de tener información fiable para áreas muy pequeñas, implicaría una fracción muestral tan alta que normalmente sería más barato empadronar a toda la población, y ciertamente más eficaz.

En cambio, si la etapa básica se realiza adecuadamente, la información básica que ella provee debe utilizarse para la elaboración de un marco muestral que permita llevar a cabo una serie de investigaciones sobre cualquier tema que necesite basarse en un conocimiento de las características esenciales de la población.

TRABAJO EN EL TERRENO

Existen varios prerequisites para que la operación censal se realice de forma exitosa. Debido al hecho que a menudo estas condiciones no se cumplen adecuadamente, son fuentes de problemas o errores.

En la mayoría de los países de la región no existe una cartografía censal adecuada, lo que constituye un problema grave. Aunque un esfuerzo importante se haya realizado al momento del censo del 80, en muchos países no se ha podido contar con los recursos necesarios para la mantención al día de los mapas. Este problema se agrava en aquellas zonas urbanas que han sido el objeto de inmigraciones importantes en los últimos años.

Dado que la etapa de recolección de datos es de muy corta duración (idealmente un solo día), es preciso que los límites de los áreas de empadronamiento sean claramente definidos, para evitar duplicaciones u omisiones. Si se trata de realizar una parte de la recolección de los datos censales mediante una muestra, es aún más importante que la cartografía sea buena. La selección de la muestra tiene que basarse en la misma jerarquía de unidades geográficas que constituyen las unidades administrativas del empadronamiento: no por razones de la teoría del muestreo, sino porque

representan la única estructura disponible en ese momento. Entonces, cualquier equivocación respecto de los límites de estas unidades no solamente perjudica la cobertura de la parte básica del censo, sino puede llevar a serios sesgos en la representatividad de la parte de la información obtenida por muestreo.

Otros elementos importantes que frecuentemente sufren de atención insuficiente son el reclutamiento y la formación del personal de terreno, tanto los empadronadores como sus supervisores de primera línea. En algunos países se suele usar alumnos de los colegios secundarios como empadronadores, y tal vez sus maestros como supervisores. Surge el problema de la falta de conocimiento de los requerimientos de datos y otros conocimientos necesarios para la buena realización de estas tareas. Puede existir además un problema de falta general de motivación, entonces es de temer que este problema sea aún más grave en lo que trata de una tarea esencialmente marginal desde el punto de vista de la persona encargada de tareas censales durante no más de unos pocos días.

El problema de la motivación a menudo dificulta también la planificación y realización de todos los componentes del censo. De este modo se ve más dudoso aún cualquier componente que complique las operaciones, tal como la necesidad de contar con dos o más boletas, o seleccionar una muestra según ciertos criterios bien específicos.

Más particularmente, la supervisión directa del trabajo de terreno es de primera importancia, si se aplica un plan de muestreo cualquiera. Es imprescindible que se realice el diseño (trabajo de oficina) de tal manera que el empadronador no desvíe del plan de muestreo, por ejemplo seleccionando los hogares que le parezcan más pequeños o de alguna otra manera más fáciles de entrevistar, lo que puede introducir sesgos tan grandes que la información así recolectada llegue a ser inútil. Por ejemplo, el sesgo en favor de los hogares pequeños puede implicar serias subestimaciones de la fecundidad y de la mortalidad infantil. Dado que este tipo de omisión fácilmente sería más fuerte en los sectores desfavorecidos donde existen condiciones de hacinamiento y algunas familias muy grandes, el sesgo en las estimaciones sería más exagerado justamente para las poblaciones donde la fecundidad y la mortalidad infantil son mayores, disfrazando así las diferencias respecto de otros sectores de la población.

Por todas las razones arriba mencionadas, es preciso dar la primera prioridad a la buena realización de las operaciones básicas del censo. Ello implica simplificar lo más posible el contenido de la boleta reducida y concentrar los esfuerzos en una buena formación del personal, acompañado por una estricta supervisión en el terreno. Solamente cuando estas condiciones se cumplen se puede pensar en operaciones más sofisticadas.

CONSIDERACIONES TECNICAS

A pesar de las dificultades prácticas y de las preocupaciones señaladas en la sección anterior, en varios países la presión por obtener datos sobre una amplia gama de sujetos hará necesario el uso del muestreo como complemento de la información universal. En este sentido cabe mencionar algunos aspectos que debieran ser tomados en cuenta en este trabajo, específicamente en el sentido de la aplicación, no del diseño propiamente tal.

Normalmente un diseño de muestra correctamente hecho no debe permitir importantes sesgos muestrales, siempre y cuando los estimadores aplicados a los datos obtenidos respetan el diseño probabilístico que se haya aplicado para la selección de la muestra.

Entre las fuentes de sesgos no muestrales se puede mencionar todo lo que a cobertura se refiere, ya sea una mala aplicación de las instrucciones para la selección de la muestra, simples omisiones o posibles duplicaciones de hogares. Además con una boleta muy recargada existe la posibilidad de omisiones accidentales o deliberadas dentro de los hogares. Estos problemas se pueden resolver si la supervisión es suficientemente completa.

Por otra parte, la varianza muestral está fuertemente afectada no solamente por el tamaño de la muestra (normalmente no es un problema serio excepto para áreas menores o características relativamente raras), sino también por efectos de la estructura del diseño respecto a la distribución en la población de las variables bajo estudio (clustering effects). Este factor puede ser muy importante cuando la unidad más pequeña de selección de la muestra

incluye un número no muy reducido de hogares (por ejemplo, selección de áreas de empadronamiento enteros para facilitar la distribución y supervisión de las tareas).

Cabe considerar la forma en que se hace la selección de la muestra, empezando con los elementos del diseño. En primer lugar, es necesario determinar la unidad más apropiada para ser seleccionada. Queda claro que la unidad "atómica" o indivisible para el muestreo es el hogar (preferentemente la vivienda entera en el caso de hogares múltiples): no puede ser la persona. Pero sí podría ser una unidad algo más grande constituido de varios hogares: vivienda, manzana o alguna otra unidad administrativa. Por razones prácticas a veces se ha recomendado elegir una de estas unidades más grandes como elemento de selección, o sea un conglomerado de hogares, entrevistando a todos los hogares contenidos en las unidades seleccionadas con la boleta expandida.

No obstante la justificación administrativa de esta opción, hay que reconocer el compromiso entre la eficiencia estadística y la eficiencia operativa, ya que la primera sugiere tomar la unidad más pequeña (el hogar) y la segunda consideración implica trabajar con unidades bastante más grandes, hasta el área que representa la tarea completa de un empadronador. Esto resulta ser peligroso debido a hecho que ciertas características socio-demográficas son fuertemente correlacionadas dentro de áreas pequeñas. De esta manera la selección de un conglomerado de hogares podría revelar características muy distintos de los conglomerados vecinos. Aunque al nivel nacional o regional este tipo de diseño puede no resultar sesgado, es ineficiente (desde el punto de vista de la varianza muestral) y puede proporcionar estimaciones para áreas desagregadas que no son representativas.

También entra en juego la cuestión de la fracción de muestreo más adecuada: en la Argentina en 1980, se aplicaron distintas fracciones muestrales según provincia y tamaño de area urbana; en otros países se ha aplicado, por ejemplo, una fracción única de 10% (u otro valor), lo que, si bien simplifica la operación, también puede implicar una pérdida importante de eficiencia (en el sentido costo/unidad útil de información obtenida), eso debido a que en cierta medida es el tamaño de la muestra que importa más que la fracción muestreada, sobre todo para estudios desagregados.

Con dos boletas, se complica la recolección de los datos. Sin embargo, como se discutió arriba, hay tantos requerimientos de datos que es imposible atender a todos aquellos con una boleta básica: por un lado, sería demasiado costosa la operación (e incluso se pondría en peligro hasta la digitación de la totalidad de la información); por otro lado, el hecho de omitir todo salvo la información de la boleta básica implicaría desaprovechar una ocasión ideal para juntar información bastante detallada, a una base de referencia única que es la información básica.

Sin embargo, el mismo hecho de usar una muestra para la parte ampliada de la información implica un incremento en el costo marginal por unidad de información. Es decir, dado que se necesita un esfuerzo mayor en cuanto a la formación, y que el ejercicio de muestreo por sí mismo implica un costo, cada entrevista por muestreo cuesta más que una entrevista básica, aunque el cuestionario fuera del mismo tamaño.

Por otro lado, si la muestra no se levanta correctamente, el costo radica más bien en el hecho que la información resulta ser mala.

Sin embargo, si el costo de recolectar elementos de información del 100% de la población no permite la inclusión de ciertas preguntas, con el posible resultado que algunos requerimientos de información no se llenan, el problema llega a ser, preguntar de una fracción de la población o no preguntar del todo. En tal caso, la teoría dice que normalmente es posible calcular una fracción muestral apropiada y elaborar un diseño de muestra que optimiza. Lamentablemente, aún la teoría no nos ayuda tanto cuando hay que optimizar respecto de muchas variables a la vez (por lo menos, la teoría que se pudiera implementar, aún en condiciones muy favorables). Cuanto más complejos sean los requerimientos, más difícil la implementación de una estrategia optima.

USO DEL MARCO MUESTRAL PARA ESTUDIOS EN PROFUNDIDAD

Aun cuando la muestra se realiza en forma correcta, quedan dudas, especialmente en lo que trata del uso de información desagregada. Hoy en día en los países de la región existe cada vez más preocupación por disponer de datos subnacionales, ya sea a nivel de estados o provincias, o para municipios u otras unidades administrativas pequeñas. El uso del muestreo para la recolección de determinadas variables implica severas limitaciones para su uso a niveles demasiado desagregados, sobre todo cuando la fracción muestral es demasiado baja (digamos menos del 10 por ciento). Quizás sea más importante intentar garantizar la calidad de la información universal, que luego sirva como base de referencia para estudios específicos, los cuales de todas maneras no se pueden prever ni manejar en su totalidad al momento del censo. Así es importante tener un buen marco básico para facilitar tales estudios.

Por un lado, la necesidad de disponer de datos básicos que se refieran directamente a las unidades menores hace aumentar la prioridad de recolectar esta información del 100% de la población. Adicionalmente, las distintas encuestas en profundidad que se hacen necesarias durante los años intercensales son muy diversas y su naturaleza no se puede predecir con certeza al momento del censo. Tampoco se puede predecir cuáles de estos estudios se harán en qué parte del país, tal vez en zonas relativamente pequeñas. Estos hechos hacen aún más necesario la creación y mantención de un marco muestral completo y permanente, el que solamente puede radicar en información censal de alta calidad para toda la población.

Un ejemplo reciente del uso de los datos censales para el diseño de una muestra nacional es la Encuesta Retrospectiva Demográfica de Guyana (GUYREDEM 1986). En esta encuesta se usó la cartografía y las listas del Censo de 1980 y la estructura administrativa/censal de Regiones y Areas de Empadronamiento (AE) para la elaboración de un diseño muestral que respondiera a la exigencia de obtener datos demográficos detallados y representativos para todas las Regiones del país, incluyendo las áreas remotas y poco pobladas. Se aplicó un diseño muestral bietápico, no autoponderado al nivel nacional. En la primera etapa se seleccionaron una cantidad de Areas de Empadronamiento (AE) por región con probabilidad proporcional a su tamaño (número de hogares) según los datos

censales. En segundo lugar se seleccionaron sistemáticamente los hogares según su posición en las listas disponibles. El hecho que se pudo realizar la puesta al día de las listas en el terreno pocos meses antes del comienzo de las entrevistas posibilitó el buen funcionamiento de esta estrategia, ya que el diseño se adaptó automáticamente al crecimiento o disminución del número de hogares en cada AE. De esta manera se levantó la encuesta con un diseño muestral nacional, insesgado y con excelente cobertura.

Dichas operaciones resultaron posibles por que la cartografía del Censo de 1980 fue bastante buena y completa , y la población había cambiado relativamente poco durante lo seis años intervinientes. Sin embargo, se encontraron errores importantes en la cobertura cartográfica del censo en varias de las AE seleccionadas (errores de identificación de límites de las AE, duplicaciones y omisiones). Estos errores, cuando son frecuentes, constituyen un problema serio, lo que indica la necesidad de dedicar un esfuerzo importante a la formación y supervisión de los empadronadores en el momento del levantamiento del censo para minimizarlos. En el contexto de los crecientes requerimientos para datos a niveles menores, éste es más importante aún, ya que un error respecto de la inclusión o exclusión del equivalente de un área de empadronamiento (o tal vez varios, ya que los errores tienden a ser correlacionados) quizás invisible al nivel nacional o provincial, llega a ser muy importante cuando se estudia un area pequeña.

USO DE UNA TECNOLOGIA APROPIADA

Mucho se ha dicho sobre las dificultades operacionales que imposibilitan el mejor aprovechamiento de la moderna tecnología, tal como es la computación electrónica. También se escuchan discusiones sobre el uso de muestras de adelanto por que el procesamiento de los datos es tan lento y costo. Cabe la impresión que estas inquietudes, muy válidas en las rondas censales del 70 y del 80, están llegando rápidamente a ser casi irrelevantes en muchos casos.

La disponibilidad del microcomputador ha cambiado totalmente las reglas del juego. A partir de la no existencia de este instrumento hace menos de diez años, a través de una primera aplicación para trabajos relativamente pequeños

hace sólo tres o cuatro años, hemos llegado al punto, hoy mismo, en que se están digitando, verificando, analizando y publicando censos nacionales enteros con un número reducido de microcomputadores cuyo costo total es una pequeña fracción del precio de uno de los computadores que se usaron hace menos de diez años. Es el caso de varios países de Africa occidental donde el censo fue levantado en 1987 o 1988, y esos países están usando equipos que son dos a tres veces más lentos, y tienen menos capacidad de almacenamiento, que los que en este momento se están vendiendo en el las ciudades principales de América Latina por precios cercanos a cinco mil dólares. También es cierto que en varios países de América Latina y el Caribe, todas las etapas del procesamiento de los datos del próximo censo se realizarán con microcomputadores.

Un caso específico del uso de microcomputadores para resolver algunos de los problemas de la difusión y análisis de los datos censales, que también podría servir para la construcción de los marcos muestrales, es el paquete REDATAM desarrollado en el CELADE. En su forma actual (Versión 3.1), almacena los microdatos de censos de población y vivienda (o de encuestas por muestreo) en microcomputadores y así permite al usuario obtener tabulaciones para áreas de interés suyo. Ya existen varias aplicaciones de REDATAM. Por ejemplo en Chile se ha creado una base de datos en REDATAM para la totalidad de la información del Censo de Población y Vivienda de 1982, y el INE está usando esta base de datos, conjuntamente con la información cartográfica correspondiente, para responder a demandas de una variedad de usuarios.

La facilidad que tiene REDATAM para manejar la identificación de las unidades de la jerarquía geográfica de un país, desde las unidades administrativas mayores (provincias o estados) hasta las unidades elementales tales como manzanas, lo hace un instrumento ideal para el almacenamiento de la información que corresponde al marco muestral. En este momento se está desarrollando un proyecto para agregar a REDATAM un sistema de información geográfica, más una selección de funciones estadísticas. Ya en la versión actual es posible seleccionar muestras aleatorias sistemáticas.

REFERENCIAS BIBLIOGRAFICAS

CELADE (1986) Algunas Reflexiones sobre los Censos del 90: Posibles Areas de Colaboración. IX Conferencia Interamericana de Estadística, 15 al 18 de septiembre de 1986, Río de Janeiro, Brasil.

CELADE (1989) El Contenido Demográfico de la Boleta de los Censos de Población de la Década del 90. Seminario sobre contenido y diseño de la boleta censal, 13 al 16 de marzo de 1989, Río de Janeiro, Brasil.

Chackiel, J. y Arretx, C. (1988) Recent Experiences in the Collection of Demographic Data in Latin American Population Censuses of the 1980s. Seminario sobre recolección y procesamiento de datos demográficos en América Latina, 23 al 27 de mayo de 1988, CELADE/IUSSP, Santiago, Chile.

INDEC-CELADE (1985) Los censos de población del 80. Taller de Análisis y Evaluación, mayo de 1985, Buenos Aires, Argentina.

Macció, G. A. (1988) Hacia la puesta al día del programa censal. En el umbral de los 90. Seminario sobre recolección y procesamiento de datos demográficos en América Latina, 23 al 27 de mayo de 1988, CELADE/IUSSP, Santiago, Chile.