



1210

01

15008.01	38056	90/06/01
Nº de acceso	Nº ISIS	Fecha
04 S M T <u>V</u> C	06 a <u>m</u> c s	94 <u>08</u>

Compendiador: Resumen informativo Resumen indicativo Indización Nada Barnaz

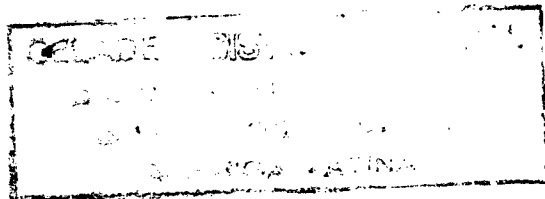
Documento en otro idioma: _____
Documento en otra fuente: _____
Documento complementario: _____

EL CENSO COMO MARCO MUESTRAL
Y EL POSIBLE USO DE REDATAM

CELADE

Documento a ser presentado en el Seminario Latinoamericano Sobre Evaluación Postcensal y Uso de Muestras Maestras, organizado por el DANE, Colombia, coauspiciado por la CEPAL y con la colaboración de CELADE y la Oficina de Estadísticas de las Naciones Unidas.

Cartagena de Indias, Colombia
4 al 8 de junio de 1990.



RESUMEN

En este documento se pretende presentar algunas consideraciones generales sobre el uso del censo de población y vivienda para la construcción de un marco muestral para encuestas intercensales. Se hará especial referencia a la posible utilidad del paquete REDATAM como herramienta para facilitar la mantención del marco y la selección de muestras.

INTRODUCCION

La intención de este documento es analizar el uso de la información que provee el censo nacional de población y vivienda, para facilitar la realización de las encuestas que se hagan necesarias a nivel nacional o sub-nacional durante el período intercensal. Específicamente, el enfoque principal será la utilización de los datos censales para la construcción de un marco muestral que pueda proporcionar tanto una muestra maestra nacional, como muestras adicionales o independientes para usos específicos a cualquier nivel de desagregación. Se estudiará algunos aspectos del posible uso del paquete "software" REDATAM (REcuperación de DATos censales para Areas pequeñas por Microcomputador), elaborado por el CELADE, en la preparación del marco muestral nacional. Además, se analizará su posible papel en la mantención del marco durante el decenio intercensal y en la selección y análisis de las muestras.

En los países que no cuentan con registros completos y actualizados de la población (y en esta situación se encuentran todas las países de la región), el censo de población y vivienda es la única fuente de datos razonablemente exhaustivos sobre la población y sus condiciones de vida en un momento fijo en el tiempo.

Como se señaló en el informe de la reunión de los Directores de Estadísticas de septiembre de 1987, pocos países de la región, al momento de preparar sus censos de población y vivienda, tienen presente el hecho que los censos constituyen el marco muestral más adecuado para otras investigaciones, especialmente encuestas de hogares por muestreo. Esto presenta un problema serio, ya que el censo es un ejercicio de muy alto costo y representa la única posibilidad de establecer un sistema integrado de muestreo de encuestas durante todo el período intercensal, normalmente diez años y frecuentemente algo más.

Es imprescindible que se aproveche de la mejor manera posible el esfuerzo que implica levantar el censo decenal, cuyos resultados deben servir hasta por lo menos el año 2000. Ello requiere de recursos regulares y un equipo estable de personal que permita una dedicación permanente para el proceso continuo de actualización de los datos del marco muestral, así como una atención por parte del personal para la selección de las muestras.

CARACTERISTICAS DE UN MARCO MUESTRAL NACIONAL

Para obtener una muestra, se requiere de cierta información sobre la población objeto de estudio. En el caso de una muestra nacional de hogares o personas, llámese "maestra" o no, esto implica disponer de un marco muestral basado en información adecuada en algún momento en el tiempo. La disponibilidad de tal marco muestral se hace particularmente importante si se prevé la necesidad de seleccionar no una, sino varias muestras a través del tiempo.

Además, se requiere de uno o más diseños muestrales aplicables a dicho marco muestral, así como evidentemente la capacidad de seleccionar y poner en uso las correspondientes muestras. Si no ha

de ser una sola muestra "una vez para siempre", rápidamente desactualizada y perdiendo su representatividad a todo nivel, también requiere que alguna parte de la información sobre la estructura del marco se mantenga lo más actualizada posible.

Según Kiregyera (1982), es de primera importancia disponer de un marco muestral adecuado, a saber una lista exacta, completa y actualizada de unidades muestrales en un censo o una encuesta. Un marco muestral nacional basado en datos censales es una herramienta para tener acceso a selecciones representativas de la población, en cualquier momento en el tiempo, sin gastar recursos excesivos. Por lo tanto, es necesario hacer uso óptimo de la información que provee un censo nacional de población y vivienda, el cual representa el único ejercicio de recolección de información con cobertura universal, con la capacidad de proporcionar la necesaria información estructural para la construcción de tal marco. Sin embargo, en la mayoría de los países en desarrollo, marcos muestrales nacionales adecuados y utilizables, especialmente de unidades últimas de muestreo, casi invariablemente son inexistentes (Kiregyera, 1982).

Ya al momento de la planificación del censo, hay que anticipar los requerimientos del marco muestral nacional, para incluir en el censo los elementos de información esenciales para la preparación del mismo. El censo es la única fuente de datos con cobertura universal y definiciones homogéneas.

Para responder a las necesidades del programa intercensal de encuestas por muestreo, es necesario establecer la capacidad de seleccionar muestras adecuadas para una variedad de eventuales encuestas de hogares, particularmente encuestas de mano de obra y de tipo demográfico-social. Además hay que prever algunos tipos de encuestas de índole no-tradicional que dicen relación con las emergentes preocupaciones para planificación y acción a nivel sub-nacional.

Esto apunta a la necesidad de confeccionar un instrumento que permita elegir muestras grandes y pequeñas, a nivel nacional o regional, y incluso a niveles más desagregados o geográficamente restringidos (y por tanto tal vez con fracciones de muestreo bastante elevadas, lo cual actúa en contra de la alternativa de mantener solamente una muestra maestra sin el marco completo).

A continuación se propone una aproximación al contenido necesario para estos fines:

Identificación geográfica: desde las regiones administrativas mayores hasta el nivel más desagregado, la vivienda o mínima agrupación de ellas. El grado de precisión del último nivel a incluir depende de alguna manera de la capacidad y estrategia previstas para la mantención del marco así como de la gama de posibles diseños muestrales que se podrían aplicar durante el período de vigencia del marco (es decir, hasta por lo menos algún tiempo después del levantamiento del siguiente censo nacional, momento en que incluso es muy posible que hayan cambiado de manera importante las definiciones de las unidades de la jerarquía geográfica).

Información demográfica: población (número de viviendas, hogares y personas de cada unidad elemental geográfica); número de personas por sexo y edad o grupos de edad. Algunas características de la vivienda, tal vez de la unidad superior (manzana o segmento), o sea conglomerado, la posible penúltima unidad de muestreo (y en la práctica, la última del muestreo "en la oficina": la selección de viviendas años después del censo o necesita una etapa de actualización en los conglomerados seleccionados, o una etapa de selección en el terreno; para los propósitos de la discusión actual, estos dos casos son equivalentes). Además, sería útil disponer de

algunas características socio-económicas para ayudar en la estratificación de la muestra de algunas encuestas sobre temas específicos.

De hecho pareciera que lo arriba señalado solamente repite algunos de los elementos de información que de todas maneras se tienen que incluir en el censo; sin embargo, es importante tener en mente este uso específico que se propone para ellos. Para los propósitos del marco muestral, algunos de estos datos pueden ser útiles a nivel de manzana, distrito u otro, aunque se recolectan a nivel del individuo en el censo; lo importante es que se tienen que mantener en el marco, que no necesariamente sería el contenido total del censo.

Todo esto nos lleva a considerar el marco muestral como el vehículo clave para proporcionar no una sola, sino muchas muestras, nacionales o sub-nacionales, durante un período de tiempo bastante extendido.

Un tipo de marco muestral se puede visualizar como un listado completo de la población y de ciertas características esenciales. Sin embargo, este tipo de lista, aunque fuera posible adecuarla plenamente en un momento dado en el tiempo, se desactualizaría de inmediato, debido a los efectos de mortalidad, fecundidad y migración, especialmente la migración interna. En todo caso, en la práctica, los micro-datos necesarios nunca han sido lo suficientemente completos, a los niveles más bajos de agregación, en el caso de censo alguno en el mundo, y aun menos en los países de la región. Por ello, es necesario considerar un marco estructurado, el cual sería bastante estable en el tiempo a los niveles más agregados, y factible de actualizar en algún grado a los niveles menores de su estructura.

Es altamente deseable que el marco muestral sea suficientemente completo y flexible para permitir una variedad de posibles diseños muestrales. Para estos propósitos, uno de los requerimientos claves es un sistema de estratificación geográfica consistente con las áreas administrativas mayores e intermedias del país, y que tome en cuenta las peculiaridades más importantes tales como los principales centros urbanos, áreas de rápida expansión o barreras topográficas. Para que la estratificación sea eficiente, se necesita aprovechar no solamente toda la información geográfica, sino también la información sustantiva sobre variables claves de la población y sus características socio-económicas.

Podría decirse que el marco muestral tal como se lo está visualizando aquí tiene componentes fijos (o relativamente fijos), y otros variables. De manera general, los elementos fijos son los más grandes, de los niveles más agregados de la jerarquía, y corresponden esencialmente a unidades geográficas administrativas tales como regiones, provincias, estados, distritos, o municipios. Los elementos variables, desde luego, son las mismas personas y hogares que conforman la población objeto de estudio, y con ellos sus viviendas. Entre estos dos extremos se observan los niveles más o menos estables, tales como manzanas y sectores de empadronamiento.

Es muy deseable que, en lo posible, la parte "fija" del marco se organice de tal manera que refleje el tipo de diseño muestral que se supone podría utilizarse en la mayoría de las aplicaciones. Esto implica la necesidad de dar suficiente prioridad a la estratificación y la definición de las unidades primarias de muestreo (UPM) y tal vez las unidades secundarias (UPS) del diseño, para garantizar una cierta estabilidad a los niveles mayores de la jerarquía, mientras se conserve suficiente flexibilidad a los niveles menores para permitir una diversidad de usos, y adecuarse al crecimiento desigualmente distribuido de la población y sus

cambiantes patrones migratorios (con especial atención a zonas de la periferia urbana y similar).

MUESTRAS MAESTRAS Y MUESTRAS ESPECIFICAS

El tema central de esta sesión se refiere a las muestras maestras nacionales. ¿Qué es una muestra maestra nacional, y para qué sirve? Podría responder a un amplio rango de definiciones, aproximadamente entre los dos siguientes casos extremos:

(i) Un marco muestral completo con indicadores geográficos precisos y plenamente actualizados por lo menos en cuanto al listado (o conteo) de sus elementos irreducibles (hogares o viviendas)

(ii) Una selección postcensal, fija para siempre, de un número determinado de unidades pequeñas (conglomerados), con o sin mecanismo para el muestreo de unidades elementales, y tal vez alguna posibilidad de actualización de las unidades así seleccionadas. Admitamos que el caso más reducido aún de una muestra permanente de unidades últimas, es decir hogares, es imposible e incluso no sería deseable bajo ningún punto de vista, por razones tanto de movilidad de las unidades poblacionales como de "response burden" en el largo plazo.

Queda claro que la primera definición corresponde a una situación idealizada y no realista, sobre todo en países que ni siquiera tienen registros confiables de eventos vitales. Por otra parte, la segunda definición puede ser realista, pero es obviamente inadecuada. Tratemos de ver cómo, en la práctica, se puede acercarse lo más posible de la realización del ideal, dado el actual contexto tecnológico y suponiendo la existencia de recursos razonables.

Una muestra maestra de algún tipo, históricamente ha sido necesaria, debido a la imposibilidad de volver directamente a los datos censales para diseñar independientemente la muestra de cada posible estudio de comienzo a fin. En el presente, aunque las herramientas de computación podrían obviar en parte esta restricción, sería un tremendo desperdicio de esfuerzos no aprovechar la estabilidad de la estructura geográfico-administrativa de la información y la posibilidad de mantener dicha información ordenada y disponible para usos repetidos. Adicionalmente, el uso de un diseño muestral que mantiene la selección de sus unidades mayores durante cierto tiempo ayuda en la tarea de reemplazar de manera sistemática las unidades menores y así evitar problemas de duplicación y la consiguiente molestia del público, que puede perjudicar la calidad de las respuestas.

Podríamos considerar como la muestra maestra esencialmente una muestra nacional de la parte "fija" del marco, por ejemplo de UPM y UPS, dentro de la cual se procedería a la selección de muestras específicas de unidades menores, hasta los hogares, para cada aplicación, ya sea las distintas repeticiones (trimestrales, ...) de una encuesta permanente de hogares, ya sea encuestas puntuales e independientes. De hecho, en algunos países se utilizan de esta manera los marcos muestrales censales, empleándose distintos métodos de reemplazo (rotación) o de actualización de las unidades ulteriores de la selección (conglomerados y hogares). En algunos casos, existe también la posibilidad de reemplazo de unidades de nivel "fijo" en el largo plazo, cuando se "agotan", hasta las UPM cuando sea necesario. Sin embargo, en la práctica en la mayoría de los países de la región, puede ser más razonable considerar las UPM y a veces UPS como selecciones permanentes. Las unidades menores, las que necesitan ser actualizadas cada vez que se utilizan, son entonces la parte variable de la selección.

No es la intención de este documento entrar en excesivos detalles de los posibles diseños muestrales, ya que ésto depende no solamente del caso de cada país, pero también en gran medida de los recursos técnicos disponibles. No obstante, se tratará de indicar algunos lineamientos que tal vez sean útiles en situaciones reales.

En cuanto a la cobertura y estructura del marco, es necesario garantizar su representatividad nacional. El marco, y su capacidad de muestreo, deben poder alcanzar toda o cualquier parte de la población del país. Esto implica que las unidades de menor nivel propiamente geográfico deben cubrir todo el territorio nacional, es decir que no se excluya ningún distrito, municipalidad o sector. También es necesario que sea posible seleccionar cualquier unidad que incluya individuo alguno de la población, aunque en la práctica ciertas unidades remotas o de muy escasa población puedan nunca estar incluidas en la muestra.

No hay que olvidar que a menudo se han presentado serios problemas de cobertura en el levantamiento de los censos en varios países de la región. Debería existir la posibilidad de incluir, potencialmente, la casi totalidad de la población entre las unidades "candidatos" para la selección (digamos un mínimo de 98% de la población, incluyendo áreas de crecimiento y remotas). Esto puede ser factible, aun cuando la cobertura del censo mismo sea algo menor, si la delineación de la estructura geográfica es completa y existe la posibilidad de actualizar las unidades (menores) a medida que se encuentren seleccionadas en alguna muestra. Cabe notar que la actualización de un marco muestral para incorporar nuevos elementos tales como edificios, nuevos asentamientos o casas es un ejercicio costoso. No obstante, este ejercicio vale la pena, al momento de la preparación del censo, cuando el marco se va a utilizar para una serie de encuestas

subsiguientes, especialmente si existe también una estrategia de actualización parcial como la ya señalado arriba.

Hay que prever que habrán múltiples usos y usuarios (o por lo menos demandas para utilizar la información), lo cual indica la necesidad de coordinación. La rápida difusión de los microcomputadores acompaña la creciente demanda para información, particularmente a niveles desagregados, apuntando la consiguiente necesidad de establecer la capacidad para su manejo y análisis. Visto la gran disponibilidad de microcomputadores poderosos, sería técnicamente factible que un número importante de usuarios dispongan de la base de datos (o una parte importante de ella) en sus propios equipos, sobre todo si se permite el uso de este marco muestral para estudios locales o descentralizados. Si esto fuera el caso, cada usuario podría estar haciendo uso independiente de la información, y posiblemente distintos usuarios podrían estar buscando información de los mismos hogares, creando una carga intolerable para los respondentes.

Por lo tanto (como también por razones tradicionales como la inquietud por garantizar la confidencialidad), el marco y su mantención debería quedar en custodia del organismo que realiza el censo, o por los menos éste debería poner en marcha algún sistema de garantía de coordinación, quizás una unidad de control a través de la cual tendrían que ser registrados todos los usos del marco. En este sentido REDATAM tiene una opción muy útil para la administración de la base de datos, que funciona a través del uso de claves de protección que se pueden utilizar para limitar el acceso a la base de datos a sólo ciertos niveles (los con unidades de mayor tamaño).

En resumen, se puede describir una posible estructura general del marco muestral, la muestra maestra y las muestras seleccionadas de ella a través del siguiente ejemplo:

Nivel jerárquico de la base de datos	Ejemplo	Nivel estructural del marco muestral
1	Región	
2	Provincia	Estrato
3	Comuna	
4	Distrito	UPM
5	Sector	
6	Segmento	USM
7	Manzana	
*	Vivienda	UTM

Nótese que la jerarquía tiene que ser estricta: es decir que una unidad puede contener una o más unidades del siguiente nivel inferior, pero cada unidad inferior debe pertenecer totalmente a una sola unidad superior.

En este ejemplo, las provincias (nivel 2) de todas las regiones son estratos naturales. El conjunto de los distritos (nivel 4) de las comunas de una provincia determinada son sus UPM, y algunas de estas unidades serán seleccionadas para la muestra maestra, los distritos así seleccionados representando entre sí a la provincia a la que pertenecen. (Sería igualmente posible tomar una muestra de distritos de cada comuna de la provincia, haciendo de las provincias un nivel inferior de estratos. Este tipo de estrategia facilitaría la selección de muestras desde la muestra maestra para estudios sub-nacionales, pero puede ser inconveniente para las muestras nacionales.)

Siguiendo con el ejemplo, dentro de cada distrito seleccionado y entre todos los segmentos de los distintos sectores de ello, se procedería a la selección de un determinado número de segmentos como unidades secundarias de muestreo (UPS). Según la estrategia adoptada, y de acuerdo con la frecuencia con la que se estaría

seleccionando muestras para una variedad de estudios, estas UPS seleccionadas podrían o no considerarse como propias de la "muestra maestra". Aun cuando lo sean, en el mediano plazo será necesario reemplazarlas, o correr el riesgo de repetir la selección de una fracción importante de los hogares que contienen.

Supongamos que se decide que la muestra maestra incluye las UPM y UPS seleccionadas, y se espera que no se tendrá que reemplazar las UPS durante un período razonablemente largo (unos años). En este caso, es perfectamente factible generar una sub-base del marco nacional con todos los micro-datos de la base de datos REDATAM para la muestra maestra, y utilizar esta sub-base como, de hecho, el marco muestral para las subsiguientes selecciones de muestras para cada estudio que se lleve a cabo. A medida que se obtenga información para actualizar partes de este sub-marco, se incorporaría en la base, hasta el momento que se decida que demasiadas UPS están agotadas, o que la desactualización de parte (no recientemente seleccionada) de la base es tan importante que no se puede dejar así como está. En ese momento, se volvería a la base nacional (marco muestral) para efectivamente seleccionar una nueva muestra maestra.

A continuación se estudiará más en detalle las posibles ventajas del uso del paquete REDATAM para apoyar las tareas anteriormente descritas.

EL USO DE REDATAM

Durante los últimos cinco años, como respuesta al problema de la no disponibilidad de información censal para áreas geográficas desagregadas, el CELADE ha producido y sigue desarrollando un poderoso instrumento que podría ayudar en el manejo del marco muestral, el paquete computacional REDATAM. REDATAM es un sistema

"software" para el uso de microcomputadores IBM o compatibles para el almacenamiento de todos los micro-datos de un censo completo de población y vivienda, o de una encuesta de hogares. Este sistema de fácil uso permite la elaboración de tablas con cualquier variable de la base de datos, para el área o grupo de áreas que se elija, hasta incluso resultados cruzados entre variables de vivienda y población (e.g. el número de niños por sexo según condición del techo de la vivienda)

Debido a su estructura de base de datos transpuesta y en forma comprimida, REDATAM permite un procesamiento bastante rápido, ya que no es necesario leer toda la base para analizar sólo una parte de ella. Sólo se necesita leer las variables involucradas en el análisis, así como las áreas geográficas de interés.

REDATAM tiene varias ventajas para la preparación y actualización del marco muestral. En primer lugar, su estructura geográfica corresponde, por definición, a la jerarquía utilizada en el censo, y puede incluir hasta el último nivel definido (manzana o menor). También incluye todos los micro-datos de las variables sustantivas del censo. Para cada unidad de cada uno de estos niveles, se puede acceder a los correspondientes micro-datos para la preparación de tablas analíticas.

Por lo tanto, el uso previsto de los datos para áreas desagregadas debe ser considerado al momento de la organización del censo, ya que una buena base de datos cartográficos, que puede ser actualizada durante el período intercensal, es imprescindible para la explotación de los datos para áreas pequeñas. Además, la definición del nivel mínimo de identificación geográfico es muy importante, dado que este nivel determina los límites del censo para proporcionar información desagregada.

REDATAM permite desplegar todos o algunos de los niveles geográficos en la base de datos activa. También permite desplegar, durante la definición del área geográfica de interés, el número de registros (viviendas y personas) seleccionados.

Se puede agregar una o más variables a la base de datos. Específicamente, se puede generar una variable indicadora que señala que determinadas unidades, a cualquier nivel de la jerarquía geográfica, han sido seleccionadas para alguna muestra, así facilitando la selección de sucesivas muestras a través del tiempo sin enfrentar problemas de repetición, permitiendo así el uso de un solo sistema de selección para todas las posibles encuestas, aun cuando la explotación del marco no es uniforme a través de las regiones del país.

También se puede actualizar los parámetros de las unidades geográficas (a cualquier nivel) tales como el número de hogares o personas, de acuerdo a la nueva información recogida durante los trabajos de terreno de las encuestas, siempre y cuando no se cambien los límites geográficos de las unidades involucrados. Esta posibilidad, conjuntamente con una estrategia apropiada de modificación del sistema de estimación, facilitaría en medida importante la mantención al día del marco y la validez de las estimaciones de las subsiguientes encuestas. Este factor puede llegar a ser de suma importancia en las zonas de alto crecimiento demográfico en la periferia de los grandes centros urbanos, u otras zonas de crecimiento rápido tales como nuevas áreas de explotación agrícola.

En el caso que hayan cambios en la definición de zonas, por ejemplo la división de un sector de alto crecimiento en distintas sub-unidades, si bien no se prevé por lo pronto la posibilidad de redefinir automáticamente dichas unidades sin repetir toda la generación de la base de datos REDATAM, en todo caso será posible

generar una variable indicadora ("flag") que indique que estas unidades son especiales, permitiendo aislarlas para un tratamiento manual o con otra herramienta computacional; en todo caso es fácil programar chequeos tales como, "si el nuevo tamaño (luego de una puesta al día) excede por xx (o yy%) el tamaño original, genere un indicador o haga alguna otra cosa", simplificando de esta manera no sólo la identificación de unidades aberrantes, pero también evitando los errores más groseros de mala entrada de datos.

REDATAM también puede ser útil para la selección de muestras. Ya en la versión actual, REDATAM 3.1, existe una función SAMPLE (Muestra) que permite la selección de una muestra sistemática de los casos en el área geográfica de interés, para análisis.

Actualmente se está desarrollando una nueva versión del sistema, REDATAM+, la cual será disponible hacia fines de 1990, con el objetivo de extender las actuales capacidades del paquete. Específicamente, permitirá relacionar la información sobre las variables sustantivas con datos cartográficos, a través de la incorporación del Sistema de Información Geográfica (SIG) PC-ARC INFO.

Con la nueva versión, REDATAM+, será posible seleccionar unidades geográficas de acuerdo a criterios cuantitativos, como por ejemplo el número de personas o hogares en cada una de ellas. De esta manera se podrá simplificar la operación de estratificación según rangos de tamaño de las unidades, o valores de otros indicadores tal vez sintéticas como tipo de barrio, nivel educacional, etc.

También sería posible hacer la selección de acuerdo a cualquier variable de la base de datos, ya sea geográfica o sustantiva, original o agregada.

De acuerdo con estas posibilidades, y teniendo en cuenta que la distintas encuestas casi siempre tienen algunos puntos de común interés, podría estudiarse la posibilidad, no sólo de analizar cada nueva encuesta con REDATAM, pero también de agregar a la base de datos que constituye el marco muestral nacional, la información obtenida de las mismas encuestas. Se podría incluir tanto la actualización de las variables ya comprendidas en la base, como los datos sobre nuevas variables que son el objeto sustantivo de la encuesta.

De todas maneras, aunque REDATAM no necesariamente permite realizar todas las etapas de la selección de una muestra en un diseño complejo, sí permite ejecutar las correspondientes estimaciones usando los correspondientes factores de ponderación. Por ejemplo, el diseño muestral de la Encuesta Demográfica Retrospectiva de Guyana (GUYREDEM 1986) no involucró uso alguno de REDATAM, ya que este paquete no existía en ese momento. No obstante, los datos que provinieron de la encuesta fueron ingresados a una base de datos REDATAM en 1987, y una parte del análisis de los resultados fue hecho con REDATAM. El diseño fue autoponderado a nivel regional pero no nacional (el rango de valores de los pesos fue de 4 a 20), y entonces fue necesario utilizar la ponderación para todas las tablas nacionales, lo cual se hizo sin problemas.

Tal vez uno de los usos más interesantes de REDATAM se encuentre en la posibilidad de hacer múltiples pruebas con relativamente poco esfuerzo. Por ejemplo, se puede visualizar la selección de una serie de muestras alternativas, utilizando distintas modificaciones del diseño o de algún parámetro de la selección, y inmediatamente calcular estimaciones de la representatividad efectiva de cada una de dichas muestras para el área de la que proviene, respecto de todas las variables existentes en la base de datos. De esta manera puede ayudar en la formación

de estratos o conglomerados para su posterior uso en muestra de una encuesta específica.

REDATAM 3.1 produce directamente distribuciones marginales y tablas de múltiple entrada (de hasta cinco dimensiones), y tiene una facilidad para calcular el promedio de una variable para cada celda de una tabla cruzada de hasta cuatro dimensiones. Además permite grabar un archivo conteniendo los datos del área geográfica de interés, en un formato apropiado para su ulterior análisis con SPSS o SL-MICRO, o en formato DOS (ASCII) para luego alimentar cualquier paquete sofisticado de análisis estadístico.

Otro aspecto de la nueva versión, REDATAM+, dice relación con la incorporación de un sistema de información geográfica (SIG), el que ayudará de manera importante en la visualización de la ubicación o dispersión física/geográfica de las posibles selecciones, facilitando de esta manera la toma de decisiones motivadas tanto por consideraciones metodológicas como presupuestarias. En ciertas circunstancias, como por ejemplo el diseño de una muestra especial sub-nacional, esta facilidad, conjuntamente con la posibilidad de hacer múltiples pruebas señalada en el párrafo anterior, también podría ser útil en la formación de nuevos estratos u otras unidades (tipo conglomerados) que corresponden a agrupaciones de unidades básicas del marco nacional.

También la combinación de REDATAM+ con el GIS puede ayudar a generar la cartografía para ser utilizada directamente en terreno, identificando los sectores, segmentos y manzanas seleccionados.

Un posible problema, pero que no es propiamente del REDATAM, concierne el control del uso de la base de datos. Desde un punto de vista, es muy conveniente que la información se difunda lo más posible (como ejemplo mínimo, tener la base correspondiente a cada

de estratos o conglomerados para su posterior uso en muestra de una encuesta específica.

REDATAM 3.1 produce directamente distribuciones marginales y tablas de múltiple entrada (de hasta cuatro dimensiones), y tiene una facilidad para calcular el promedio de una variable para cada celda de una tabla cruzada de hasta cuatro dimensiones. Además permite grabar un archivo conteniendo los datos del área geográfica de interés, en un formato apropiado para su ulterior análisis con SPSS o SL-MICRO, o en formato DOS (ASCII) para luego alimentar cualquier paquete sofisticado de análisis estadístico.

Otro aspecto de la nueva versión, REDATAM+, dice relación con la incorporación de un sistema de información geográfica (SIG), el que ayudará de manera importante en la visualización de la ubicación o dispersión física/geográfica de las posibles selecciones, facilitando de esta manera la toma de decisiones motivadas tanto por consideraciones metodológicas como presupuestarias. En ciertas circunstancias, como por ejemplo el diseño de una muestra especial sub-nacional, esta facilidad, conjuntamente con la posibilidad de hacer múltiples pruebas señalada en el párrafo anterior, también podría ser útil en la formación de nuevos estratos u otras unidades (tipo conglomerados) que corresponden a agrupaciones de unidades básicas del marco nacional.

También la combinación de REDATAM+ con el GIS puede ayudar a generar la cartografía para ser utilizada directamente en terreno, identificando los sectores, segmentos y manzanas seleccionados.

Un posible problema, pero que no es propiamente del REDATAM, concierne el control del uso de la base de datos. Desde un punto de vista, es muy conveniente que la información se difunda lo más posible (como ejemplo mínimo, tener la base correspondiente a cada

región o provincia en la oficina regional correspondiente del Instituto Nacional de Estadística y Censos). Sin embargo, si no hubiera un sistema centralizado de control del uso de esta información para la selección de muestras, es probable que habrían conflictos debido al uso independiente del marco y la consiguiente repetición de unidades seleccionadas.

CONCLUSIONES Y RECOMENDACIONES

Se ha analizado el uso de la información básica del censo nacional de población y vivienda, en la construcción de un marco muestral que proporcione muestras representativas de la población para el programa de encuestas intercensales. Hay que anticipar los requerimientos del marco muestral para incluir en el censo los elementos esenciales para su preparación. Estos incluirían la identificación geográfica completa, información demográfica sobre la población, y algunas características socio-económicas y de la vivienda .

Es deseable que el marco sea lo suficientemente completo y flexible para reflejar los posibles diseños muestrales a utilizarse durante el decenio. Para ello es necesario dar prioridad a la estratificación y la definición de las UPM y UPS, para garantizar estabilidad a los niveles mayores de la jerarquía, mientras se conserve flexibilidad a los niveles menores. Es necesario garantizar la cobertura y representatividad nacional del marco.

El análisis previsto para áreas desagregadas debe considerarse durante la organización del censo, ya que buenos datos cartográficos son imprescindibles para la explotación de los datos sustantivos para áreas pequeñas. La definición del nivel mínimo de identificación geográfico es muy importante, dado que determina los límites del censo para proporcionar información desagregada.

Una muestra maestra de algún tipo es necesaria, debido a la imposibilidad de volver a los datos censales para diseñar independientemente la muestra de cada estudio, pero una selección fija de determinadas unidades pequeñas es inútil. No obstante, un diseño muestral que mantiene la selección de sus unidades mayores durante cierto tiempo ayuda en la tarea de actualizar y reemplazar las unidades menores cuando sea necesario.

Podría considerarse como muestra maestra una muestra nacional de la parte "fija" del marco, dentro de la cual se procedería a la selección de muestras específicas de unidades menores, hasta los hogares, para cada aplicación. En algunos países se utilizan así los marcos muestrales censales, empleándose distintos métodos de reemplazo o de actualización de las unidades menores a medida que se encuentren seleccionadas. Existe también la posibilidad del eventual reemplazo de unidades de nivel mayor.

Es posible que un número importante de usuarios dispongan de la base de datos, si se permite el uso del marco muestral para estudios descentralizados. Parece conveniente que el marco muestral se difunda lo más posible, pero el organismo responsable del censo debe poner en marcha un sistema de garantía de coordinación. En este sentido REDATAM tiene una opción muy útil para la administración de la base de datos, que funciona a través del uso de claves de protección para limitar el acceso a la base de datos a sólo ciertos niveles.

Se han señalado otras posibles ventajas del uso del paquete REDATAM, para la preparación del marco muestral y en su mantención durante el decenio intercensal, así como en la selección y análisis de las muestras. REDATAM almacena todos los micro-datos de un censo de población y vivienda, o de una encuesta de hogares, en forma de una base de datos transpuesta y comprimida. Su estructura geográfica es la utilizada en el censo, y puede incluir hasta el

último nivel definido. Esto permite la elaboración rápida de tablas con cualquier variable de la base de datos, ya que sólo se necesita leer las variables involucradas en el análisis, así como las áreas geográficas de interés. REDATAM permite grabar un archivo con los datos del área de interés, para su análisis con otro paquete de análisis estadístico.

Se pueden agregar variables a la base de datos, específicamente una variable indicadora que señala que ciertas unidades han sido seleccionadas para alguna muestra, facilitando la selección de sucesivas muestras sin enfrentar problemas de repetición, permitiendo así el uso de un solo sistema de selección para todas las encuestas. También se puede actualizar los parámetros de las unidades geográficas de acuerdo a la información recogida en las encuestas, siempre y cuando no se cambien los límites geográficos de las unidades involucrados. Esto, conjuntamente con una estrategia apropiada de estimación, facilitaría la mantención al día del marco y garantizaría la validez de las estimaciones.

Cuando se hagan cambios en la definición de zonas, si bien no existe aún la posibilidad de redefinir automáticamente las unidades sin repetir la generación de la base de datos, será posible generar una variable indicadora que permite aislar estas unidades para un tratamiento especial.

REDATAM también puede ser útil para la selección de muestras. En REDATAM 3.1 existe una función que permite la selección de una muestra sistemática de los casos en el área geográfica de interés, para análisis. Aunque REDATAM no permite realizar todas las etapas de la selección de una muestra en un diseño complejo, sí permite ejecutar las correspondientes estimaciones usando factores de ponderación. Ofrece la posibilidad de probar la selección de una serie de muestras alternativas, utilizando distintas modificaciones

del diseño o de algún parámetro de la selección, y calcular estimaciones de la representatividad efectiva de dichas muestras para el área de interés respecto de todas las variables en la base de datos. Así ayudaría en la formación de estratos o conglomerados para uso en muestra de una encuesta específica.

Actualmente se está desarrollando REDATAM+, con el objetivo de extender las actuales capacidades del paquete. Permitirá relacionar las variables sustantivas con datos cartográficos, a través de un Sistema de Información Geográfica, el que ayudará en la visualización de las posibles selecciones. Con REDATAM+ será posible seleccionar unidades geográficas de acuerdo a criterios cuantitativos, simplificando la operación de estratificación según tamaño de las unidades, o valores de otros indicadores. También sería posible seleccionar según cualquier variable de la base de datos. En el diseño de una muestra sub-nacional, esta facilidad podría ser útil en la formación de nuevos estratos o conglomerados que corresponden a agrupaciones de unidades básicas del marco nacional. REDATAM+ con el GIS también puede ayudar a generar la cartografía para ser utilizada en terreno.

Debería estudiarse la posibilidad, no sólo de analizar cada encuesta con REDATAM, sino también de agregar al marco muestral la información obtenida de las encuestas, tanto la actualización de las variables ya comprendidas en la base, como los datos sobre nuevas variables que son el objeto de la encuesta.

REFERENCIAS BIBLIOGRAFICAS

CELADE (1987) REDATAM: Relevancia para los censos de 1990". CELADE, Santiago, Chile. LC/DEM/R.48, Serie A, N° 174, septiembre 1987. También en inglés: "The Relevance of the REDATAM System for the 1990 censuses".

CELADE (1988a) "REDATAM: Una Herramienta para Obtener Información Desagregada Geográficamente para la Planificación". CELADE, Santiago, Chile. LC/DEM/R.56, Series A, N° 183, junio 1988.

CELADE (1988b) Small-Area Census Data Services by Microcomputer: Applications of the REDATAM System in Latin America and the Caribbean". CELADE, Santiago, Chile. LC/DEM/R.57, Serie A, N° 185, agosto 1988.

CELADE (1989) Censos de Población de 1990: Selección de Documentos del CELADE, LC/DEM/G.77, Serie A, N° 193:

- El Contenido Demográfico de la Boleta de los Censos de Población de la Década del 90, pp. 7-30.
- El Rol de la Cartografía en la Utilización de los Datos Censales con los Sistemas de Información Geográfica Conectados a REDATAM, pp. 31-42.
- Anotaciones Sobre Factores Condicionantes al Uso del Muestreo en los Censos de America Latina, pp. 43-56.
- Consideraciones Sobre el Uso del Muestreo en los Censos de America Latina, pp. 57-69.

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.

Kiregyera, B. (1982) "Some Sources of Nonsampling Errors in African Census and Survey Work: A Review". International Statistical Review, 50, pp. 305-316.

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.