

EL MUESTREO PARA LA VALIDACION

Y EL CONTROL DE LA CALIDAD

DE LOS DATOS EN EL CENSO



- 5 JUN. 1999

**TRABAJO PREPARADO POR EL GRUPO DE
EVALUACION DEL CEN90, INTERGRADO POR:**

Coordinadora del grupo

Lic. Norma Pizarro

Subcoordinador

Lic. Gerardo Mitas

Lic. Omar Cordoba

Lic. Margarita Dialeva

Lic. Vera Tabakovic

Lic. Clara Grinblat

Lic. Carlos Guevel

Lic. Graciela Minardi

Lic. Daniel Griggio

INDICE

INTRODUCCION	1
1. LA IMPORTANCIA DE LA CALIDAD DE LOS DATOS CENSALES	1
2. TIPOS DE ERRORES EN EL CENSO	2
3. PROGRAMAS PARA MEJORAR LA CALIDAD DE LOS DATOS CENSALES	3
3.1 Programa de control de calidad	3
3.1.1 Control estadístico de la calidad	4
3.2 Programas de evaluación	5
3.2.1 Errores de contenido	7
3.2.1.1 Encuestas de reentrevistas	7
3.2.1.2 Encuestas permanentes de hogares	8
3.2.1.3 Registros administrativos	9
3.2.1.4 Componente correlacionada en el censo	9
3.2.2 Cobertura	10
3.2.2.1 Evaluación de las operaciones específicas	11
3.2.2.2 Estimación de la magnitud de la cobertura	12
3.2.3 Apareamiento	14
4. PLANEAMIENTO DE LA EVALUACION DEL CENSO	15
4.1 Objetivos	15
4.2 Usos	16
REFERENCIAS	18

INTRODUCCION

Es bien sabido que un censo "perfecto" es imposible de alcanzar. Por consiguiente, los Institutos Nacionales de Estadística (INE) se ven con la necesidad de contar con métodos que evalúen y examinen procedimientos, operaciones y resultados para asegurar la calidad de los datos censales.

Existen un gran número de métodos que pueden ser aplicados y estos difieren ampliamente en términos del nivel técnico de sofisticación y de la calidad de los resultados.

En los apartados 1,2 y 3 se describen sintéticamente estos métodos de control y evaluación; finalmente en el punto 4 se subrayan y se discuten algunas consideraciones y factores importantes envueltos en el planeamiento e implementación de un programa de evaluación censal por parte de los INE.

1. LA IMPORTANCIA DE LA CALIDAD DE LOS DATOS CENSALES

Los INE son grandes recolectores y diseminadores de información referente a población: su número, distribución, características, actividades, etc.; o sea, aquellos datos que provee, entre otras fuentes, el Censo de Población.

Los datos producidos por el censo son usados para una gran cantidad de propósitos, que van desde la representación política de los distintos estados y dentro de ellos, la distribución de fondos públicos, la planificación de los organismos oficiales, privados y compañías hasta las decisiones básicas que realizan los particulares a partir de dichos datos. Estas dos condiciones:

- Los INE son los únicos productores de los datos del censo
- Las numerosas decisiones y sus correlatos, las consecuencias, que se toman en base a los datos que provee el censo

llevan a los INE a dar suma importancia a la Calidad de los datos del Censo.

La "calidad" se mide en términos de la proximidad al "valor exacto" de las distintas variables investigadas en el censo. El concepto de exactitud significará la obtención del "valor verdadero" de la variable investigada en el censo, y preservarlo durante el procesamiento.

2. TIPOS DE ERRORES EN EL CENSO.-

Las respuestas dadas por los individuos u hogares a la pregunta perteneciente a la cedula censal permitira conocer la distribucion de una variable.

La pregunta Donde nacio? ofrece las alternativas de respuesta: 1. En esta localidad; 2. En otra localidad dentro del pais; 3. En otro pais; 4. Ignorado. Ellas constituyen las categorias de la variable: Lugar de Nacimiento. Supongamos que un individuo cualquiera responde: En otra localidad y sin embargo, por una distraccion involuntaria el censista o el codificador, consigna "En esta localidad". Vemos que:

respuesta dada por el individuo u hogar j	=	valor verdadero del j-esimo individuo u hogar	+	error proveniente del:
				- censista,
				- respondente,
				- cuestionario,
				- codificacion,
				- procesamiento, etc.

$$Y_j = \mu_j + e_j \quad (1)$$

Estos errores, pueden ser introducidos tanto en la faz de relevamiento como en la faz del procesamiento de los datos. Pueden deberse a: el diseno del cuestionario, el metodo de indagacion, las caracteristicas de la pregunta, las actitudes del respondente u otras causas. Pueden ser accidentales o introducidos intencionalmente, el respondente puede malinterpretar una pregunta y dar una respuesta equivocada sin intencionalidad o responder incorrectamente por tener la conviccion de que ciertas respuestas son "mas aceptables" que otras.

De la misma forma el censista puede afectar la exactitud de la respuesta sugiriendola o transcribiendola incorrectamente al cuestionario. El codificador puede leer o interpretar mal la respuesta o codificarla incorrectamente con caracter voluntario o involuntario.

Estos errores conocidos como errores no muestrales, pueden clasificarse en dos grandes grupos: **sistematicos** (consistentes) y **variables** (inconsistentes). Los errores sistematicos son aquellos que ocurren en una misma direccion, creando un sesgo en las estimaciones. Por ejemplo existe la tendencia a declarar la edad, a partir de ciertos tramos, siguiendo la atraccion representada por los digitos 0 y 5, en estos casos producira una subestimacion del valor verdadero: otro ejemplo es aquel en que un codificador sistematicamente coloca un codigo erroneo produciendo en este caso una sub(sobre)estimacion en la variable. Los errores variables son los que no se producen en una direccion dada. Por ejemplo en la pregunta -cual es el nivel que curso ? - o -el mas alto que curso ? -

puede haber sido omitido por el censista, al advertirlo coloca **primario** en el casillero correspondiente. Supongamos que el "valor verdadero" que corresponda al individuo fuera **secundario** en este caso el valor de la respuesta se vera subestimada. En cambio si el "valor verdadero" que deberia consignarse fuera **preprimario** el valor de la respuesta se vera sobreestimado.

3. PROGRAMAS PARA MEJORAR LA CALIDAD DE LOS DATOS CENSALES

Los **INE** disponen de dos programas para mejorar la calidad de los datos censales, o sea contribuir a conocer las fuentes de los errores, su medida y control. Estos son:

- Un programa de control de calidad que mediante el planeo de un conjunto de procedimientos intenta alcanzar una calidad fijada.
- Un programa de evaluacion cuyo objetivo sera estimar una medida de variacion (rango de variacion) y la deteccion de las fuentes que provocan errores.

Naturalmente el censo requerira de los dos procedimientos si lo que se quiere es obtener una buena calidad en los datos a publicar. En el censo un estudio de evaluacion no puede ser sustituto de un programa de control de errores, es decir, de control de calidad, sino que resultara un medio para determinar cuando un programa de control de calidad produce los resultados deseados y cuando los controles adicionales son requeridos.

3.1. PROGRAMA DE CONTROL DE CALIDAD

Al control de calidad se lo puede tipificar, basicamente, en aquellos controles que incluyen elementos del muestreo estadistico y en los que no hacen uso de esta tecnica. Es asi que el control no estadistico de calidad de un censo incluye el ciclo de actividades por medio de las cuales se supervisa una etapa del operativo con el objeto de asegurarse que se respeten las reglas de trabajo prefijadas y mantener un nivel aceptable de calidad en los procedimientos; por ejemplo:

- En el trazado de mapas se realiza un control revisando la claridad de los límites marcados y la cobertura de divisiones administrativas para evitar duplicaciones u omisiones.
- Controlar la calidad de las operaciones preparatorias tales como la impresión de formularios y la compaginación de materiales para la enumeración en campo.
- En un censo de hecho se deberá cuidar la selección del personal para el trabajo, brindar un entrenamiento adecuado a censistas y supervisores y revisar que el trabajo se haya completado antes de abandonar el campo
- El procesamiento, el control de recepción de los cuestionarios, la preparación de las hojas de conteo, la correcta ubicación geográfica son también algunas etapas a mencionar para este tipo de control.

Por el contrario el control estadístico de calidad para censos permite a través de la aplicación de un plan de inspección por muestreo controlar y medir la calidad de trabajo en algunas etapas del procesamiento de la información, facilitando la adopción de una acción correctiva. Por ejemplo:

- Puede realizarse control estadístico de calidad en el trabajo de campo y en la supervisión de un censo de derecho .
- En la impresión de formularios y compaginación del material pueden implementarse planes de muestreo continuo.
- La codificación y el análisis de consistencia manual pueden controlarse estadísticamente a través de un plan de inspección por muestreo.

3.1.1 CONTROL ESTADISTICO DE LA CALIDAD

Con el objeto de controlar la calidad de una determinada producción se aplican ciertas técnicas de inspección que permiten mantenerla dentro de standards prefijados. Pueden considerarse dos alternativas: la inspección de la totalidad de los ítems, o la inspección de una muestra de ellos. La inspección 100% es muy costosa y además la experiencia indica que no permite detectar todos los errores como se creería. La inspección por muestreo estadístico ofrece resultados iguales o mejores que la anterior y es más práctica en el sentido de tiempo y costo. Algunas ventajas de esta última son:

- Se conoce la probabilidad o riesgo de tomar una mala decisión (rechazar un "lote" de calidad aceptable y aceptar uno de calidad inaceptable).

- El diseño permite que el tamaño de la muestra provea al plan la capacidad de distinguir entre pobre y buena calidad.

En el caso de una inspección por muestreo se deben tener presentes dos tipos de muestreos comúnmente usados en operaciones censales que son el continuo y el de aceptación; cada uno de ellos presenta distintos planes.

El muestreo continuo se aplica a unidades de naturaleza y producción homogénea, se usa generalmente en operaciones de tipo mecánicas tales como impresión de cuestionarios, compaginación de materiales, colocación de etiquetas. Las ventajas de este tipo de muestreo son las siguientes:

- el proceso es constante y no se deteriora con el tiempo (a pesar de que existe algún tipo de intervención humana).
- no hay lotes de trabajo como en el muestreo de aceptación.
- la inspección y corrección se lleva a cabo a medida que se produce el trabajo.

El muestreo de aceptación se aplica a tareas que básicamente sufren la intervención humana. Utiliza lotes de trabajo (grupos de variables) que se inspeccionan por muestreo. La decisión de aceptar o de rechazar un lote se basa en el número de defectuosos encontrados en la inspección de una muestra.

La ventaja de este tipo de muestreo se refiere a la practicidad de aplicación cuando la producción puede agruparse en lotes de trabajo, es decir, no es una producción continua.

3.2 PROGRAMAS DE EVALUACION

La evaluación estadística de la calidad de un censo requiere la definición de una medida del "error total". Esta medida está basada en la elección de un modelo simple (1) para describir las respuestas a las variables que se investigan.

La suma de los valores registrados para una variable a través de todos los individuos nos dará el total correspondiente a una variable. Por ejemplo: "total de niños que concurren a la escuela primaria", "total de mujeres entre 50 y 60 años", "total de viviendas con antigüedad mayor de 10 años", etc.

Una medida del error total del censo para la variable en cuestión es el llamado **error cuadrático medio** que básicamente, puede describirse como la suma de los siguientes tres elementos:

- El cuadrado del sesgo
- La varianza simple de respuesta
- La componente correlacionada de la varianza

No solo es necesario a través de un programa de evaluación conocer la magnitud de dichas componentes sino que su estudio permite detectar grupos de variables o de personas que presentan comportamientos anómalos.

El **sesgo** es la diferencia entre el promedio de los totales correspondiente a una variable a través de infinitas repeticiones y el total verdadero para la población en estudio. Dos son las fuentes que dan origen al sesgo en el censo:

- Los errores de respuesta (errores sistemáticos)
- Los errores de no respuesta (errores de cobertura)

La **varianza simple de respuesta** es una medida de la consistencia de las respuestas obtenidas para una variable, considerada a través de las innumerables e hipotéticas repeticiones del censo. Refleja el efecto de todos aquellos factores no sistemáticos que introducen cambios en las respuestas.

La **componente correlacionada de la varianza** es un indicador de la variabilidad o diferencias en la tendencia de los censistas a modificar las respuestas de los informantes. Se ha comprobado a través de numerosos estudios que los censistas alteran las respuestas de los individuos, ya sea por la forma de preguntar, la formulación de aclaraciones, la interpretación de las respuestas, etc.

Cuando se quieren estimaciones para el total de la población, la mayor

influencia sobre el error cuadrático medio la ejerce el sesgo. A su vez, cuando dichas estimaciones son requeridas para poblaciones de menor tamaño y/o dominios pequeños, aumenta el efecto de la varianza simple de respuesta. En cuanto al impacto de la componente correlacionada, este dependerá del número de censistas o de la carga que posean, siendo mayor cuanto mayor sea dicha carga. El caso óptimo es aquel en el que cada censista entrevista a un único individuo lo que ha dado origen, en países desarrollados a la autoenumeración.

El sesgo provocado por los errores de respuesta sistemáticos, la varianza simple de respuesta y la componente correlacionada se incluyen como errores de contenido mientras que la medición del sesgo introducida por el error de no respuesta corresponde a errores de cobertura.

3.2.1 ERRORES DE CONTENIDO

Los errores de contenido pueden medirse a partir de la comparación entre la información del censo y la información proveniente de otra fuente. Una vez hechas las comparaciones, se construyen indicadores apropiados para la medición de estos errores, así como estimadores de las componentes del error total. Las fuentes alternativas de información son:

Encuestas de reentrevistas.

Encuestas permanentes de hogares.

Registros administrativos.

por otra parte los errores de contenido pueden medirse usando los datos censales "diseñando" adecuadamente las cargas de los censistas:

Estudios para medir la componente correlacionada en el censo.

3.2.1.1. ENCUESTAS DE REENTREVISTA

Es una encuesta diseñada específicamente para evaluar el contenido. De acuerdo a los objetivos de medición pueden estructurarse dos tipos de encuestas: a) aquellas que tienen por objetivo medir la varianza simple de respuesta y b) aquellas donde el objetivo es medir el sesgo.

Estas diferencias de objetivos conducen a una diferencia en los supuestos teoricos y metodologicos. Es de destacar entre los supuestos teoricos la diametral oposicion de las condiciones en que idealmente se realizan las reentrevistas. Se asume que en una reentrevista del tipo a) las condiciones generales deben ser las mismas que las del censo (cuestionario, encuestadores, supervision, publicidad). En cambio para las reentrevistas del tipo b) se supone un conjunto de condiciones generales que no solo sean distintas, sino que conlleven a la obtencion de respuestas de mayor calidad que las obtenidas en el censo. En esta ultima, habra que minimizar el riesgo de obtener valores no verdaderos para cada unidad (persona/vivienda), de lo que se deduce que el entrenamiento de los encuestadores debiera ser intensivo, llegando en algunos casos a recurrir a una tercera entrevista para reconciliar las diferencias encontradas en las respuestas.

A pesar de estas diferencias, poseen puntos en comun como el disenio muestral, apareamiento y tiempo de realizacion. Por lo general se empleara un disenio estratificado multi-etapico donde ademas, las unidades primarias se eligen con probabilidad proporcional al tamano. Una vez extraida la muestra y realizada la reentrevista, deben aparearse las unidades del censo y de la reentrevista. Solo se utilizan los casos apareados de la muestra de lo que se deduce que es primordial que el criterio de apareamiento no introduzca sesgo; es decir, se debe cuidar que la proporcion de apareamientos erroneos sea minima no olvidandose el efecto opuesto, un criterio de apareamiento muy estricto podria descartar unidades con diferencias relativamente pequenas.

En cuanto al tiempo de realizacion cuando la reentrevista se realiza muy proxima al censo existe el riesgo de condicionar la respuesta. Es decir los sucesos no seran independientes produciendo subestimaciones en las estimaciones que tengan cualquiera de los objetivos mencionados (a o b). Por otra parte, si el comienzo de la reentrevista es tardio, encierra una potencial perdida de exactitud, especialmente en aquellas variables afectadas por el tiempo. Los tiempos recomendados son entre 1 y 12 semanas a partir del operativo censal.

3.2.1.2 ENCUESTAS PERMANENTES DE HOGARES

Es una encuesta que no se ha disenado para evaluar al censo pero presenta la caracteristica de ser una fuente mas independiente que la reentrevista aunque debe mencionarse que:

- solo se pueden evaluar variables que figuran en la EPH y en el censo simultaneamente
- Las definiciones y categorizacion de variables pueden diferir.
- puede haber una diferencia notable con el tiempo del censo.

Las soluciones a estos problemas suelen ser costosas y se pueden abordar

a través de procedimientos, preguntas especiales y aun reentrevistas tanto para posibilitar el apareamiento como para conciliar diferencias. En un grado limitado es posible agregar o modificar preguntas del cuestionario con propósitos de evaluación. Ante la necesidad de mantener la comparabilidad en el tiempo, se podrán reformular preguntas referentes a la edad, el tiempo de residencia, o la ocupación y la actividad económica.

3.2.1.3 REGISTROS ADMINISTRATIVOS

Para utilizar este procedimiento, es necesario disponer de un conjunto de registros para la población censada. Los sistemas de registros adecuados son, por ejemplo:

Listas de personas enumeradas en el censo previo

Registros de nacimientos durante el periodo intercensal

Lista de estudiantes primarios, secundarios y universitarios

Padron electoral

Registros de viviendas

Registros de inmigración

Como en el caso de la EPH, los registros existen con otro propósito que el de evaluar el censo, con lo cual disminuye el costo de obtener la información. Pero en consecuencia acarrea las mismas desventajas señaladas anteriormente en cuanto a una variable en estudio y a comparabilidad en el tiempo. A esto se agrega un apareo tal vez más dificultoso, por estar los registros ordenados en forma distinta que los registros del censo.

La independencia y exactitud de la información hacen que esta sea una fuente apta para la medición del sesgo en ciertas variables (por ejemplo: fecha de nacimiento). Si además se puede combinar con un estudio de reentrevista con el objetivo de medir la varianza simple de respuesta, entonces se podrá dar estimaciones del sesgo y de la varianza simple de respuesta.

3.2.1.4 COMPONENTE CORRELACIONADA EN EL CENSO

El estudio de la componente correlacionada no requiere la realizacion de una encuesta de reentrevista pero si es necesario "diseñar" la forma en que se asigna la carga a los censistas. Este diseño es el de "muestras interpenetrantes", nombre con el que se designa a un tipo especial de aleatorizacion: el conjunto de unidades que constituyen la carga de cada censista se selecciona aleatoriamente del total de unidades.

Es necesario planificar el entrenamiento, el trabajo de campo y el procesamiento de forma tal que los errores registrados en las respuestas proveniente de un censista no esten relacionados con los del otro censista. El desarrollo del trabajo de campo exige un mayor control a fin de garantizar que cada censista complete especificamente su carga.

En los censos, la aplicacion de este metodo en forma global seria casi impracticable ya que el trabajo de cada censista podria abarcar areas muy separadas. Sin embargo, puede limitarse el estudio a algunas areas seleccionando aleatoriamente los censistas que se destinaran a ellas.

Una extension mas apropiada para este caso es la llamada "muestras interpenetrantes a dos etapas" que proporciona la aleatorizacion de los censistas y de sus cargas y disminuye la dispersion de los encuestadores. En los estudios de la componente correlacionada es posible brindar medidas de la precision de su estimacion.

Un metodo para el estudio simultaneo de la varianza simple de respuesta y correlacionada surge como una combinacion de las muestras interpenetrantes y la reentrevista. Se obtiene asi la ventaja de poder estudiar simultaneamente las variancias simples y correlacionadas y, ademas, otras causas que afectan al error medio cuadratico, como la correlacion entre las respuestas obtenidas en ambos relevamientos para el mismo individuo. Como contrapartida, presenta las siguientes desventajas:

- Trabajo de campo complejo que requerira mayor supervision y costo.
- Necesidad de controlar el periodo de tiempo que transcurre entre ambas visitas.

3.2.2 COBERTURA

Uno de los principales factores que introducen sesgo en los procedimientos censales es que los relevamientos suelen cubrir solo a un conjunto de la poblacion y, por otra parte, a ciertos individuos los alcanza en mas de una oportunidad.

Estas deficiencias se conocen como **errores de cobertura**.

Algunas de sus causas son: errores en las definiciones censales, aplicacion incorrecta de definiciones correctas, tendencia de los censistas a omitir ciertos tipos de hogares o de individuos (viviendas alejadas, minorias raciales), personas sin viviendas fijas, individuos que no desean ser censados, etc.

El grado del error de cobertura puede variar de un grupo de unidades a otro, y esto puede provocar serias distorsiones en la distribucion de la poblacion por grupos. Tampoco resultara suficiente estudiar el error de cobertura en toda la poblacion sino que sera adecuado hacerlo por regiones o por grupos de individuos.

Las severas consecuencias que los errores de cobertura pueden acarrear sobre las estimaciones y la alta frecuencia con que ellos se presentan, han provocado que una serie de paises desarrollen y apliquen diversos metodos tendientes a paliar este problema.

Los objetivos fundamentales que se persiguen al considerar los errores de cobertura son:

- evaluar las operaciones especificas que se efectuan para mejorar la cobertura censal.
- estimar la magnitud de la cobertura.

Estos objetivos no deben ser considerados sustitutivos sino complementarios, por cuanto una metodologia optima para la evaluacion y analisis de los errores de cobertura requerira el cumplimiento de ambos por igual.

3.2.2.1 EVALUACION DE LAS OPERACIONES ESPECIFICAS

Se puede enumerar una serie de actividades, cada una apuntando a incrementar la cobertura de grupos de personas cuyo alcance es dificultoso. Entre las actividades de este tipo que se pueden realizar, dentro del marco de un censo de derecho, merecen mencionarse: verificacion de hogares deshabitados, campana "Fue Ud. censado?", centros de informacion, censo casual, etc.

Una de las principales actividades que realizan varios paises, tanto aquellos que ejecutan un censo de derecho como los que efectuan uno de hecho, consiste de un conteo de la poblacion previo al censo (prelistado) que servira como guia del relevamiento propiamente dicho, como asi tambien para la correcta construccion de los segmentos y cargas del trabajo para los censistas. Este prelistado no necesariamente debe construirse a partir de un conteo en campo, sino que puede ser obtenido de diversos tipos de registros administrativos.

En muchos paises existen regiones cuyos habitantes no utilizan el idioma nacional sino uno diferente. Sera adecuado utilizar en esas zonas cuestionarios escritos en el idioma en uso, o enviar censistas que dominen esa lengua, tratando

de incrementar así la cobertura de esos sectores.

3.2.2.2 ESTIMACION DE LA MAGNITUD DE LA COBERTURA

El segundo conjunto de métodos, aquellos que buscan estimar la magnitud de la cobertura, necesitan el empleo de una o varias fuentes de información independientes del relevo censal.

- Tal es el caso del análisis demográfico que puede obtener una estimación a través de un modelo demográfico del total poblacional que luego es contrastado con el número de personas censadas. Una desventaja de este análisis es que requiere la existencia de registros actualizados y en general solo se obtienen estimaciones a nivel agregado, no pudiéndose obtener con frecuencia valores para regiones o zonas pequeñas. Como principal ventaja puede decirse que se desarrolla a partir de registros o listados independientes del relevamiento censal. Por otra parte el costo de su implementación es menor que el de otros métodos de estimación de cobertura.

- El método de evaluación por reentrevista está muy difundido, es en general satisfactorio, y su diseño es más sencillo, sin embargo puede resultar inapropiado si la reentrevista es muy posterior al censo, además existe cierta tendencia a que los individuos no alcanzados por el censo tampoco sean cubiertos por la reentrevista, lo que origina una subestimación en la razón de cobertura.

- Otros métodos que buscan estimar la cobertura censal son aquellos que utiliza la Encuesta de Hogares y que luego, mediante un apareamiento o comparación de los cuestionarios obtenidos en cada uno de los relevamientos, obtienen un total poblacional estimado y con él una estimación de la cobertura. El empleo de una o varias ondas de la Encuesta de Hogares reduce costos, por cuanto un mismo relevamiento servirá para diversos objetivos y tiene la ventaja de emplear personal con experiencia en estas tareas; sin embargo su diseño e implementación será mucho más complejo y además se corre el riesgo de que un objetivo interfiera en el otro, disminuyendo la calidad de la información recopilada.

- Una alternativa a los métodos de reenumeración es el control por registros, que consiste en tomar una lista de personas como muestra, de una fuente distinta al censo, y controlar sus cédulas censales para determinar si fueron enumeradas.

Las omisiones de las listas para el control de registros tienen menos tendencia a estar correlacionadas con las del censo que las omisiones de las reentrevistas; por lo tanto es menos probable la subestimación de la falta de cobertura. Sin embargo, es difícil encontrar listas actualizadas que abarquen a

toda la poblacion y que sirvan de referencia para la extraccion de la muestra.

Personas migrantes

Otro elemento a tener presente en la implementacion de una cobertura por reentrevista es que en muchos casos los habitantes de un hogar no seran los mismos en ambos relevamientos. Habra que decidir entonces como tratar a estos individuos, por cuanto su omision dara como resultado un subestimacion de la razon de cobertura. Pueden mencionarse tres metodos para tratar a este tipo de personas:

- **Metodo A:** trata de reconstruir la composicion del hogar en la fecha censal obteniendose de familiares o vecinos, datos basicos de las personas que se mudaron a otro domicilio. Este metodo tiene la ventaja de simplificar el proceso de apareamiento para los migrantes, pero se estima que los encuestadores relevan menos de la mitad de este tipo de individuos.

- **Metodo B:** releva a todas las personas que viven en el domicilio al momento de la reentrevista, preguntandoles su direccion en la fecha censal, poniendo especial cuidado en los migrantes. Dado que cada persona contesta por si misma este metodo produce una lista muy completa de quienes cambiaron sus direcciones, pero la informacion relativa a sus domicilios en la fecha censal suele ser vaga, y a menudo es practicamente imposible encontrar la correspondiente cedula censal.

Fue realizado un estudio comparando ambos metodos, que dio como resultado que la proporcion de migrantes enumerados en la reentrevista para los cuales se encontro la correspondiente cedula censal fue practicamente la misma para los dos criterios. Basandose en esta circunstancia fue construida una alternativa que intenta asociar las ventajas de los dos metodos descriptos.

- **Metodo C** estima la proporcion de "migrantes apareados" como el metodo A, y la multiplica por el total de migrantes estimado por el metodo B. Para utilizar este criterio es necesario relevar en la reentrevista a todos los habitantes del hogar mas aquellos que hubieran en el en la fecha censal. Este metodo combina la relativamente simple operacion de apareamiento del metodo A, con la enumeracion mas completa y directa del B.

Se espera que su sesgo sea inferior a los de los otros metodos. Tiene como desventaja que su costo es mayor que en los otros casos y se incrementa su dificultad de procesamiento debido a que utiliza un mayor volumen de informacion.

Metodos de estimacion: simple - dual

Sistema simple: algunos procesos de reenumeracion son disenados especialmente para obtener estimaciones por sistema simple. La idea es obtener un estimador a partir de la reentrevista que se cotejara con el resultado censal, en lugar de construir un estimador a partir de ambos relevamientos.

El punto fundamental es obtener en la reentrevista informacion mas exacta que en el censo. Asi se pone especial cuidado en la formulacion de las preguntas y en la capacitacion de los encuestadores para relevar a todas las personas que debieron ser enumeradas el dia del censo. Este metodo tiene como desventaja que sus resultados dependen en gran medida de los errores de respuesta que se espera que tengan una importante magnitud en este tipo de relevamiento.

Sistema dual: Es el metodo mas utilizado para obtener estimaciones de la cobertura censal. Este criterio requiere independencia entre ambos relevamientos, en forma tal que habra que prestar atencion a este punto.

Otro aspecto de cuidado es que este sistema requiere el correcto apareamiento de cada cuestionario de reentrevista con el correspondiente formulario censal, tarea que podria introducir un grave nivel de error.

El desarrollo de un disenio muestral para el estudio de los errores de cobertura dependera fundamentalmente de las caracteristicas propias de cada pais y de la precision deseada para las estimaciones.

3.2.3 APAREAMIENTO

Los tres metodos que acaban de ser descriptos (reentrevista, encuesta de hogares y registros administrativos) tienen en comun la necesidad de cotejar las cedulas censales con las del nuevo relevamiento o con los registros. Quizas sea este el elemento con mayor peso en el momento de elegir que camino tomar para evaluar los errores de cobertura. La dificultad principal del apareamiento radica en que al realizarse dos mediciones "independientes" sobre los mismos individuos habra diferencias entre las respuestas obtenidas que dificultaran el apareamiento de los cuestionarios o registros correspondientes a cada persona. Por lo tanto habra que disenar un proceso que ataque este problema.

Este proceso puede tener dos modalidades: apareamiento automatico o manual.

El apareamiento automatico tiene la ventaja de eliminar la subjetividad pero es muy complejo de programar y requiere que la informacion de ambos relevamientos sea ingresada a los equipos de computos.

Por otra parte el apareamiento manual sera mas lento y estara sujeto a los errores propios del trabajo humano.

Pero independientemente del sistema elegido se deberan formular, en la

forma mas clara posible; las tolerancias con que se trataran las diferencias entre las respuestas obtenidas. Estas tolerancias dependeran de varios factores: variables evaluadas, precision deseada, habilidad de los censistas y encuestadores, etc.

El grado de dificultad del procedimiento de apareamiento es diferente para los metodos de evaluacion descriptos. En el caso de utilizar registros administrativos los problemas de cotejo pueden ser una fuente de error mas importante que para los otros metodos, dado que la estructura de la informacion suele ser totalmente diferente y ademas los registros pueden contener datos con varios anos de antigüedad.

Para las Encuestas Permanentes de Hogares el apareamiento es algo mas sencillo, pero subsisten algunos problemas de estructura; ademas mientras los censos son en general de hecho, las encuestas son de derecho, incrementandose asi las diferencias entre los individuos relevados en cada vivienda.

La reentrevista ad hoc es el metodo que en apariencia deberia tener los menores problemas de apareamiento por cuanto la toma de la informacion estaria destinada a ello. Sin embargo esto no debe interpretarse como que el apareamiento en este caso es simple, por cuanto las dificultades que existen para aparear los cuestionarios correspondientes a cada persona continuan siendo muy importantes

4. PLANEAMIENTO DE LA EVALUACION DEL CENSO

4.1 OBJETIVOS

Para planear las tareas de evaluacion del censo sera necesario determinar que se quiere de la evaluacion. Esto podra hacerse juntamente con la o las personas encargadas del planeamiento del censo.

Al fijarse los objetivos de la evaluacion quedaran determinados el alcance, factibilidad, costos y recursos.

alcance

El alcance de la evaluacion del censo dependera del pais y las experiencias de los censos anteriores y el uso fundamental a que se destinan los datos del censo.

Asi por ejemplo, un pais que fijo como objetivo principal el obtener la mejor estimacion del total de la poblacion establecera que la evaluacion del censo tendra como alcance realizar encuestas de cobertura a nivel nacional. Es un objetivo simple pero no facil de implementar ya que esta decision debera ser acompanada de un estudio de factibilidad.

Es mas probable que un programa de evaluacion tenga objetivos multiples, asi en el ejemplo citado se podria realizar tal vez encuestas de cobertura de alcance restringido (conglomerado urbano o de ciertas regiones) juntamente con una encuesta tendiente a medir calidad (encuesta de reentrevista) de algunas variables investigadas en el censo.

factibilidad

En la planificacion de evaluacion de la calidad de un censo debera considerarse la posibilidad de llevar a cabo efectivamente las tareas que se proponen. Asi, en el ejemplo citado anteriormente, el estudio de factibilidad requerira conocer si se cuenta con una muestra capaz de dar estimaciones a nivel pais. Esta muy relacionado con las condiciones, recursos y costos. Para un pais que nunca ha conducido evaluacion del censo no es factible planear un complejo programa de evaluacion que requerira enormes recursos tecnicos y presupuestos.

costos

El costo es un factor determinante. Si se considera que un programa de evaluacion aceptable puede asumir el 5 o el 10% del presupuesto censal se debera hacer cierta eleccion de alternativas a emplear en el programa de evaluacion.

Generalmente en la mayoria de los paises se piensa que la evaluacion del censo es una parte integral del mismo y por lo tanto resulta necesario asignarle una cierta cantidad de fondos.

recursos

Existen otros recursos que tambien afectan el planeamiento de la evaluacion del censo. Ejemplo: incremento en el numero del personal tecnico con manejo de muestreo, diseno experimental, estadistico-matematico y metodologia de encuesta, asi tambien como programadores con entrenamiento matematico. Por otra parte sera necesario contar con equipamiento de microcomputadoras para algunas etapas de evaluacion.

4.2 USOS

Lo producido por un programa de evaluacion del censo puede ser utilizado de diversas maneras:

- Guia para realizacion de futuros censos o encuestas:

el objetivo de cualquier censo sera alcanzar para un costo minimo un deseado nivel de precision. La contribucion de los errores no muestrales en uno u otro metodo alternativo permitira la eleccion de procedimientos adecuados de obtencion de datos y procesamiento.

Un programa de evaluacion da a los productores y usuarios informacion valiosa para la planificacion de futuros censos y encuestas.

- Asistir a los usuarios en la interpretacion de los datos:

la publicacion de los resultados de un programa de evaluacion del censo sirve para los siguientes propositos:

- a) alertar al usuario con respecto al error que presentan los datos
- b) dar informacion sobre la magnitud relativa de los errores de una variable en particular y sobre la importancia relativa de las distintas fuentes de error.

Esta informacion puede usarla el usuario para determinar el grado de confiabilidad con que puede extraer sus conclusiones o extender sus inferencias.

- Programas de ajuste:

los INE pueden emprender un "programa de ajuste" de los datos censales a partir de la informacion proveniente de los estudios de evaluacion. Esta decision, no obstante, debera ser hecha mucho antes de que los resultados de dichos estudios se hayan procesado y no a posteriori, ya que los problemas que rodean al proceso de ajuste son variados y complejos. La misma decision de ajustar los datos debera pasar no solo por los institutos de estadistica sino tambien por los niveles gubernamentales. Como esta decision es muy delicada tampoco debera recaer sobre los analistas, estadisticos y tecnicos. Estos solo tendran responsabilidad, basada en sus conocimientos tecnicos, en dar recomendaciones apropiadas a aquellas personas u organismos que esten por tomar una decision.

Es importante tener presente que si se efectua un proceso de ajuste en los datos censales se dispondra de un valor alternativo (la estimacion proveniente del ajuste) al valor computado por el censo creando una situacion conflictiva en la eleccion, por ejemplo, en el valor definitivo a publicar. La controversia se acentua, ya que se sabe que no hay ningun camino para alcanzar el total verdadero de la poblacion y usar una estimacion, a partir del ajuste, no asegura disponer de un numero mas proximo a ese total; tambien el uso de un valor ajustado puede crear incertidumbre o dar la apariencia de que los datos censales tienden a alcanzar un resultado "politico" deseado. Otro problema es decidir donde y como hacer el ajuste, o en que niveles geograficos y/o de la poblacion se debera aplicar ya que las distribuciones demograficas, por ejemplo, se veran afectadas.

REFERENCIAS

U.S BUREAU OF THE CENSUS (1979). POPSTAN. Statistical Training Document, ISP - TR 4A-4B-4C.

U.S BUREAU OF THE CENSUS (1988). The Coverage of Population in The 1980 Census.

U.S BUREAU OF THE CENSUS (1987). Programs to Improve Coverage in The 1980 Census.

U.S BUREAU OF THE CENSUS (1985). Evaluating Censuses of Population and Housing, Statistical Training Document, ISP - TR - 5.

FELLEGI, I. (1964). Response Variance and its Estimation. JASA. 59, 1016-1041.