

Ari N. Silva



**PROCESAMIENTO
DE DATOS CENSALES:
LECCIONES DE LA EXPERIENCIA
LATINOAMERICANA**

CENTRO LATINOAMERICANO DE DEMOGRAFIA

Mathematical Induction

Principle of Mathematical Induction

Let $P(n)$ be a statement involving the natural number n . If $P(1)$ is true and $P(k) \Rightarrow P(k+1)$ for all $k \in \mathbb{N}$, then $P(n)$ is true for all $n \in \mathbb{N}$.

Example: Prove that $1 + 2 + 3 + \dots + n = \frac{n(n+1)}{2}$ for all $n \in \mathbb{N}$.

Ari N. Silva

**PROCESAMIENTO
DE DATOS CENSALES:
LECCIONES DE LA EXPERIENCIA
LATINOAMERICANA**



CENTRO LATINOAMERICANO DE DEMOGRAFIA

Serie A, Nº 170

Santiago de Chile

Junio de 1985

LC/DEM/G.29

Junio de 1985

Procesamiento de Datos Censales:
Lecciones de la Experiencia Latinoamericana

I N D I C E

	<u>Página</u>
I. Introducción	1
II. Etapas del procesamiento censal	2
III. Experiencias en el Procesamiento	3
IV. Recomendaciones adicionales	31
1. Utilización de paquetes	31
2. Microcomputadores	32
3. Descentralización	33
V. Conclusión	35
VI. Bibliografía	36

I- INTRODUCCION

Este documento tiene como objetivo principal concentrar metódicamente, en un ordenamiento lineal de tareas, diversas experiencias recogidas por el CELADE en el procesamiento de algunos censos recientes de América Latina, de tal modo que pueda servir como apoyo a la ejecución de los censos por realizar en otros países.

La intención de los comentarios de uno u otro punto en especial no es la crítica en sí; más bien se pretende aportar observaciones constructivas para mejorar las condiciones del procesamiento, con el objeto de producir resultados censales con el mayor grado de calidad y rapidez posible.

El procesamiento de un censo es una tarea lo suficientemente compleja como para no despreciar experiencias anteriores. A pesar de las distintas peculiaridades que presentan los censos de cada país, es muy significativo el grado de similitudes que los une, sobre todo en la fase de procesamiento. Por eso es importante que se examinen las buenas y malas experiencias, intentando aprender de ellas, y así mejorar nuestro propio proceso.

El capítulo II, "Etapas del Procesamiento Censal", presenta una lista de todas las operaciones que tienen una relación, pasiva o activa, con el procesamiento de datos. No se trata de un inventario de todas las actividades censales, sino solamente de aquéllas que, de alguna manera usan, o deberían usar, el procesamiento de datos para el mejor logro de su objetivo.

El capítulo III, "Experiencias en el Procesamiento", utiliza la lista anterior como índice y describe, a través de ejemplos, cada una de sus etapas, haciendo recomendaciones específicas cuando sea el caso.

El capítulo IV presenta algunas "Recomendaciones Adicionales" que no fueron mencionadas anteriormente, o que tienen una importancia tal que justifica que se destaquen. También se busca integrar en el procesamiento censal las nuevas técnicas y/o nuevos paquetes en desarrollo.

Finalmente, en la Conclusión, se hace un resumen de los puntos más importantes y recomendaciones sobre los aspectos que más afectan la calidad de los datos censales.

II- ETAPAS DEL PROCESAMIENTO CENSAL

A continuación se encuentra una lista de las etapas u operaciones que, de una manera u otra, tienen influencia o son parte del procesamiento de datos. Esta lista de etapas, sin ser exhaustiva, deberá servir de guía en la discusión del capítulo siguiente.

A- Planificación

- A.1- Cronograma
- A.2- Presupuesto
- A.3- Definición de Responsabilidades
- A.4- Definición del Equipo

B- Preparación

- B.1- Cartografía
- B.2- Diseño de la Boleta
- B.3- Censo Piloto
- B.4- Capacitación del Personal
- B.5- Muestra de Adelanto

C- Empadronamiento

D- Procesamiento

- D.1- Organización del Material
- D.2- Codificación Manual
- D.3- Grabación
- D.4- Control Cuantitativo
- D.5- Consistencia de Datos
- D.6- Análisis de los Resultados
- D.7- Tabulación y Publicación Básica

E- Explotación de la Información

- E.1- Almacenamiento en Base de Datos
- E.2- Creación de Archivos Derivados
- E.3- Divulgación de la Información

El orden de las etapas dentro de los grandes grupos no significa necesariamente el orden de ejecución ni tampoco una prioridad o importancia, a pesar de que, en gran medida, ésa es la secuencia habitual. La división y separación se ha identificado estrictamente con el propósito de organizar los ejemplos y experiencias a ser presentados, como si formaran parte de un censo hipotético a ser ejecutado.

III- EXPERIENCIAS EN EL PROCESAMIENTO

En este capítulo se presentan ejemplos reales y sugerencias o recomendaciones para evitar o minimizar los errores y sus influencias en el resultado final.

A- Planificación

A.1- Cronograma

El cronograma es una lista exhaustiva con TODAS las actividades censales, comprendiendo sus fechas límite de inicio y término, el tiempo máximo de duración y, si es posible, el detalle de las actividades que la preceden. Esta lista de actividades es importantísima durante todo el proceso censal por varias razones:

1- El hecho mismo de hacer una lista sirve para ordenar las tareas y no olvidar ninguna de ellas, por pequeña que sea.

2- La lista de actividades constituye un requisito indispensable para dos de las etapas subsecuentes, Diseño del Presupuesto y Definición de Responsabilidades. En rigor, si se desconocieran las actividades que forman parte del proceso resultaría imposible establecer líneas presupuestarias o asignar tareas específicas al personal. Por el contrario, la identificación de labores es un auxiliar fundamental para determinar los tiempos y los costos involucrados por cada una de ellas. Indudablemente, resulta mucho más fácil estimar los requerimientos de cada una de las tareas y luego sumarlos para obtener los totales de tiempo y costo que intentar estimaciones globales.

3- Más adelante el cronograma sirve para el seguimiento de las actividades por los responsables de la ejecución. A través de él es posible verificar el retraso total, así como cuáles son las actividades que influyen en el camino crítico de tiempo.

4- El inicio de las tareas de Planeamiento y Preparación censales exigen un mínimo de 2 años de anticipación a la fecha del empadronamiento. Esto es particularmente válido para la Cartografía y los Censos Pilotos o Experimentales. Si a esto se suma el tiempo de procesamiento, que requiere entre uno y medio a dos y medio

años como promedio, se tiene que el proyecto censal demanda unos cuatro años.

Todos aquellos países que no hicieron su propio cronograma censal pudieron verificar, de una manera dolorosa, esta necesidad en determinados momentos en los cuales se hizo imperiosa una reprogramación de las actividades en virtud del retraso de alguna de ellas.

La elaboración de un cronograma es una tarea que exige varios intentos interactivos antes de que se obtenga el resultado final. Con tal fin es preciso verificar todas las limitaciones de recursos de cada actividad (personal, equipos, presupuesto) y lo referente al cumplimiento de algunas "fechas-clave" (casi siempre fuera del control de los organizadores) como son la fecha del empadronamiento, y/o la fecha de presentación de los resultados para efectos, por ejemplo, de la elección de parlamentarios o representantes. Es decir, en la práctica se presenta una situación en la que se debe hacer un cronograma "de atrás para adelante", en donde se tienen las fechas de término y se buscan las de inicio, o la duración de cada actividad.

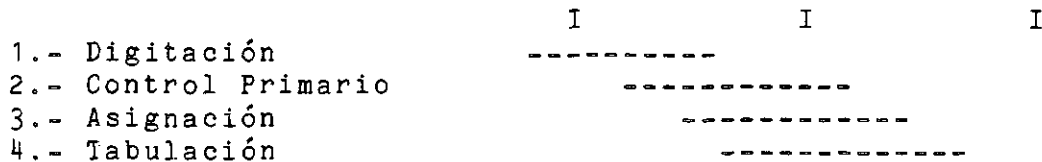
Algunos países se basaron en el cronograma para la construcción de una red de actividades PERT/CPM ("Program Evaluation Review Technique/Critical Path Method"). Esta realización, bastante exitosa, permitió juzgar, en el caso de retraso en la ejecución de una de las etapas, si ello afectaría o no las actividades subsecuentes, debido a que existen etapas en las cuales un retraso es irrecuperable, mientras que en otras (con holgura en sus tiempos de ejecución), se puede postergar su inicio si es necesario.

En el proceso de definición de un cronograma, es sumamente importante que desde un comienzo esté involucrado un experto en procesamiento de datos con conocimiento de los equipos a ser utilizados, de modo que pueda estimar con fidelidad los tiempos necesarios para todas las etapas del procesamiento, y así garantizar que no se omita ninguna fase. Una etapa que fue olvidada en más de un caso fue la del Control Primario, que garantiza la cantidad de boletas procesadas por área geográfica. En otro ítem se tratará en detalle este punto, pero aquí se quiere resaltar que quizás se pudiera evitar este problema si se involucrara un especialista en procesamiento censal en la confección del cronograma.

Una segunda observación en el tópico de Cronogramas se refiere a la elección de la fecha de empadronamiento. Hay una serie de factores a ser considerados al seleccionar una fecha, o un período, para hacer un censo, como por ejemplo, condiciones climáticas de la región, vacaciones escolares, migraciones internas debidas a eventos cíclicos locales (ferias ganaderas o épocas de cosecha), festividades

regionales o internacionales (fiestas patrias, navidad), etc.

Por último, pero no por eso menos importante, es necesario examinar la secuencia y superposición de actividades en un cronograma. Supóngase un cronograma ficticio en donde se destaca la parte del procesamiento a través de sus grandes etapas:



El desfase existente entre el inicio de dos etapas subsecuentes, como digitación y control primario, es necesario para que se tenga un tiempo suficiente de digitación de un lote de procesamiento. Sin embargo, este desfase debe ser el mínimo posible entre una etapa y otra, de manera que se pueda usar la propia ejecución de la etapa siguiente para probar la anterior. Si hay un error en la ejecución de una etapa, cuanto más temprano se lo detecte y corrija, tanto mejor porque en la mayoría de los casos habrá necesidad de reprocesamientos para reejecutar los lotes iniciales.

La situación más frecuente es aquella en que se tiene una fase lista para ejecución y la siguiente aún no ha terminado su desarrollo. En el intento de ganar tiempo, habitualmente motivado por presión de las autoridades, se es impulsado a empezar la producción de la fase desarrollada, pero sin el respaldo de la etapa siguiente como "control de calidad", y en este caso el riesgo es grande. En la mayoría de los casos hay que reejecutar gran parte de la fase, o bien hacer "parches" para corregir el problema.

A.2- Presupuesto

Muchos presupuestos se hacen casi como si se pensara que el Censo termina con la recolección de los datos. Como los recursos disponibles son siempre escasos, los responsables los concentran en las tareas inmediatas, olvidándose que la parte post-recolección ocupa alrededor del 30 por ciento del total del presupuesto, aumentando significativamente este porcentaje si hay compra de equipo.

Un rubro muchas veces subdimensionado es la utilización de cintas magnéticas. En más de un país la falta de cintas disponibles obligó a una reutilización de cintas de etapas

anteriores, borrándose archivos importantes, o a la eliminación de etapas de respaldo. En ambos casos las consecuencias son previsiblemente negativas, siendo que en el primer caso la pérdida es irreparable, como fue lo que sucedió cuando se borraron los archivos de ANTES de la asignación automática, y luego se descubrió un error en ese proceso.

Otra situación que pudo ser evitada fue la de un retraso de seis meses en la liberación de recursos para la contratación de codificadores, provocándose un retraso irrecuperable. La contratación de los digitadores también sufrió de la misma enfermedad, pero en este caso, además del retraso, hubo otra consecuencia, la cual fue el inicio de la codificación en forma intensiva, sin el proceso de digitación para confirmarlo. Cuando empezó la digitación y se detectó una incompatibilidad entre los dos procesos, fue necesario recodificar gran parte de los cuestionarios.

A.3- Definición de Responsabilidades

Es importante que exista un organigrama funcional para establecer las responsabilidades de cada participante, principalmente cuando hay tareas interdisciplinarias que involucran a personas de varias áreas. Un ejemplo de eso ocurrió en un censo durante la etapa de definición del plan de consistencia y asignación automática, de la cual estaban encargados un programador y un demógrafo. Como eran funcionarios de departamentos distintos, y como no existía un responsable específico para esta tarea, transcurrió largo tiempo antes de que se terminara el plan.

En otra oportunidad, se pudo verificar la existencia de programadores responsables para cada programa del sistema, pero no había nadie responsable del sistema como un todo, comprometiendo la integración y comunicación de los programas entre sí.

A.4- Definición del Equipo

Este tema tiene una relación directa con el de Presupuesto, en la medida en que las Direcciones de Estadística responsables del Censo casi siempre poseen equipos anticuados y/o sobrecargados, que no soportarían la carga de trabajo adicional. Por eso, históricamente la época censal es justamente cuando se cambia de equipo para un modelo nuevo de más capacidad.

Debido a las restricciones presupuestarias siempre existentes, los responsables se inclinan a dimensionar el equipo con un criterio optimista, disminuyendo la necesidad real. En otros casos, la decisión de usar uno u otro equipo

está basada en informaciones de catálogo del propio fabricante, ya sea por falta de tiempo, o por imposibilidad de probar en la práctica el equipo propuesto. Para complicar aún más la situación, algunos países reciben el equipo a través de donaciones de fondos internacionales, lo que impide que los responsables del censo tengan cualquier influencia en la decisión.

Para la definición del equipo hay que analizar con cuidado todos los criterios y requerimientos necesarios para su utilización, tales como:

1- Capacidad de multiprocesamiento, principalmente si la digitación va a ser ejecutada en el mismo equipo.

2- Capacidad de memoria para programas de relativa complejidad.

3- Capacidad de almacenamiento de datos a nivel censal.

4- Existencia de un representante del proveedor en el país, para mantenimiento del equipo, así como para capacitación de programadores y operadores.

5- Disponibilidad de versiones de los paquetes de programas a ser utilizados.

Resulta que por lo menos en dos casos, el equipo definido se mostró insuficiente para soportar la digitación y ejecución de programas al mismo tiempo. La prioridad interna necesaria dada a la digitación para que se tuviera un rendimiento razonable de los operadores, impedía, o disminuía drásticamente la ejecución de otros procesos. El número máximo de pantallas disponibles también era un factor limitante porque, o bien se usaban todas para entrada de datos, o se reservaban algunas para los programadores. La consecuencia fue obvia: retrasos en el cronograma y falta de equipos para los programadores. Se llegó al extremo de pensar en la eliminación total del proceso de verificación para recuperar el tiempo perdido.

Existe un equipo de apoyo a la computación cuya falta se hace sentir, principalmente en los países con irregularidad en los servicios de distribución de energía eléctrica. Es el UPS (Uninterruptible Power System), el cual garantiza la continuidad de funcionamiento de los computadores durante los "peaks" de luz, variaciones instantáneas de voltaje, o cuando hay paradas momentáneas en el abastecimiento. En determinadas épocas del año, a veces por problemas de lluvias, las máquinas paran a cada rato. Al tiempo perdido en la reejecución de las tareas en proceso, hay que sumar el tiempo gastado con la reentrada en funcionamiento (15 a 30 minutos). Además, hay que agregar

también el aumento de la frecuencia de los procedimientos de respaldo (para salvar lo máximo posible), y el desgaste emocional de los digitadores y programadores.

E- Preparación

E.1- Cartografía

La cartografía censal es una de las primeras tareas operativas a iniciarse, por su complejidad y cantidad de informaciones disponibles. La confección (o actualización) de una buena base geográfica sirve "...principalmente para asegurar que todos los sectores del país sean cubiertos en el empadronamiento, y que los datos sean asignados a las divisiones administrativas correspondientes". 1/

Es decir, la mejor o peor calidad de la operación cartográfica va a reflejarse directamente en los problemas de cobertura censal.

El sistema tradicional de levantamiento cartográfico se apoya en la recolección de mapas de varias fuentes, que son evaluados, examinados con relación a la cobertura, superposición de áreas, escalas, etc., hasta que se obtengan los llamados mapas censales, que son sectorizados de acuerdo a la carga de trabajo de un empadronador. Es una labor exhaustiva que involucra la dedicación integral de numerosos técnicos, y exige equipos de dibujo, equipos fotográficos y equipos de copia. Es una labor netamente manual, y, por lo tanto abierta a todos los tipos de errores humanos. Finalmente, es una labor de carácter conyuntural, ya que para el próximo censo será necesaria la repetición de casi todas sus etapas para "actualizar" la base geográfica existente.

La definición de un sistema computacional es la solución para que se minimicen los problemas del proceso manual, con la utilización de equipos "digitizers" para la captura de los contornos de cada área, y de equipos graficadores para diseñar los mapas necesarios. Este sistema debe ser capaz de verificar las inconsistencias entre los límites de las áreas contiguas.

1/ Bureau of the Census, POPSTAN - Un estudio de caso para los Censos de Población y Habitación de 1980, Washington, 1979, parte A, pgs. 79-96.

El proceso de captura de los contornos debe ser ejecutado en niveles descendentes de jerarquía; es decir, primero se ingresa el contorno del país y sus departamentos, luego el de los municipios de cada Departamento, y así sucesivamente, hasta que se alcance los sectores urbanos para ingresar las calles, accidentes geográficos (ríos, cerros) y puntos de referencia (escuelas, iglesias, estadios).

Una vez que se tengan los contornos almacenados en un medio magnético, la "impresión" de los mapas de las áreas geográficas específicas, en escalas compatibles con el nivel jerárquico, es obtenida con la utilización de los graficadores, garantizando la precisión y rapidez en la reproducción de los mapas. Por ejemplo, estos archivos deben ser usados para entregar a los supervisores y empadronadores, los croquis, por manzana, de sus respectivas áreas de trabajo.

Este archivo con la "imagen" de la base geográfica puede ser combinado con los archivos de datos censales, produciéndose resultados censales en forma de mapas (densidad de población por áreas, crecimiento de población entre censos, etc.).

Finalmente, la ventaja adicional del sistema computarizado se presenta en el momento de la actualización de los mapas censales para los próximos censos o encuestas, disminuyéndose la necesidad de rehacer todo el trabajo de levantamiento cartográfico. Además, con la base geográfica almacenada en un computador, es recomendable actualizarla gradualmente en los períodos intercensales, manteniéndola al día.

B.2- Diseño de la Boleta

"Un cuestionario bien diseñado no debe desagradar a los que lo usan - ni a las personas que proporcionan las respuestas, ni a los encargados de registrar y procesar la información". 2/

La boleta censal puede ser dividida lógicamente en dos partes distintas, cuestionario y formulario. Por cuestionario se entiende la redacción de las preguntas y el establecimiento de las alternativas para sus respectivas respuestas. El formulario es la forma con la cual se hacen las preguntas, la disposición de las mismas en el papel, tipo de letra, tamaño, color, etc.

2/ Bureau of the Census, Op. cit., parte B, pgs. 115-141.

La elaboración de un formulario es una labor conjunta de varios especialistas, entre ellos los expertos en procesamiento de datos, "que tienen que ver con la colocación de casillas de codificación, complejidad de los requerimientos de codificación y crítica, manejo de los formularios por los operadores de teclado y la procesabilidad total del formulario". 3/

Para el procesamiento de los datos censales, el formulario es mucho más importante que el cuestionario propiamente tal. El responsable por la parte computacional debe estar directamente involucrado en el diseño de la boleta, disminuyéndose así la posibilidad de futuros problemas durante las etapas de procesamiento.

a) Diagramación de las preguntas

Hay veces en que un pequeño detalle perjudica toda una pregunta, como por ejemplo en la Figura 1. La pregunta 6a tuvo un porcentaje de no respuestas de cerca del 4 por ciento, mientras que la 6b tuvo un 19 por ciento. Aparentemente, lo que pasó fue que algunos empadronadores preguntaban la parte a y seguían a otro bloque, olvidándose de preguntar la parte b.

6 SERVICIO SANITARIO	
a) ¿De qué tipo de servicio sanitario dispone?	
W.C. conectado a red pública	<input type="checkbox"/> 1
W.C. con otros sistemas de desagüe (pozo ciego)	<input type="checkbox"/> 2
Excusado tipo municipal	<input type="checkbox"/> 3
Letrina común, etc.	<input type="checkbox"/> 4
No tiene	<input type="checkbox"/> 5
b) ¿Este servicio es de uso ...?	
Exclusivo de esta vivienda	<input type="checkbox"/> 1
Compartido con otra vivienda	<input type="checkbox"/> 2

Figura 1

Una solución es separar las dos preguntas con una raya horizontal y poner números distintos en ellas, como se ve en la Figura 2.

3/ Bureau of the Census, Op. cit.

6	¿De qué tipo de servicio sanitario dispone?	
	W.C. conectado a red pública	<input type="checkbox"/> 1
	W.C. con otros sistemas de desague (pozo ciego)	<input type="checkbox"/> 2
	Excusado tipo municipal	<input type="checkbox"/> 3
	Letrina común, etc.	<input type="checkbox"/> 4
No tiene	<input type="checkbox"/> 5	
7	¿Este servicio es de uso . . . ?	
	Exclusivo de esta vivienda	<input type="checkbox"/> 1
	Compartido con otra vivienda	<input type="checkbox"/> 2

Figura 2

Es evidente que algo podría obtenerse con una mejor capacitación a los encuestadores, enseñándoles a preguntar también los subítems, pero ello no otorga suficientes garantías de éxito.

b) Ubicación de las preguntas

La ubicación y ordenamiento de las preguntas puede ser causa de no respuestas, no solamente porque no se pregunta en el terreno, sino también porque no se ingresa la información en el computador. Un ejemplo de eso es la pregunta sobre Estado Civil, que en más de un caso fue ubicada a continuación de las preguntas sobre fecundidad y mortalidad. De esto resultó que muchos empadronadores se olvidaron de hacerla a las personas a las cuales la parte de fecundidad no se aplicaba, o bien los digitadores no la ingresaron, por el mismo motivo.

Lo más indicado es seguir el orden estructural de contestación de acuerdo a las edades, y localizar la pregunta de estado civil ANTES del bloque socioeconómico, obligándose así a que todas las personas mayores de una determinada edad (como 12 años) la contesten. Al ponerla después de las preguntas económicas se corre el riesgo de no contestación (o no grabación) para las personas no económicamente activas.

c) Formato de las preguntas

Para facilitar la labor a las personas que usan el cuestionario, todas las preguntas deben tener sus respuestas precodificadas, siempre y cuando no se perjudique la recolección de la información. Hay casos donde la precodificación es imposible como en las preguntas de ocupación y rama de actividad. En otros casos la

precodificación no es utilizada por falta de espacio en la boleta, por ejemplo en la pregunta sobre educación, en que se juntan las preguntas de nivel y grado. En el ejemplo de la Figura 3, por algún problema en el entrenamiento de los empadronadores, hubo muchos casos de contestación con una "x" al lado del nivel, en vez de escribir el número correspondiente al grado, como sería la forma correcta.

B...Características educacionales... PARA TODAS LAS PE			
ESCRIBIR ? (Si sólo lee o sólo escribe marque la casilla "NO")	9	SI <input type="checkbox"/> 1	NO <input type="checkbox"/> 2
ESTABLECIMIENTO DE ENSEÑANZA REGULAR: ALFABETIZACION, PRIMARIA, SECUNDARIA O SUPERIOR ?	10	Si asiste <input type="checkbox"/> 1 No asiste <input type="checkbox"/> 2	Se ignora <input type="checkbox"/> 9
¿CUAL GRADO, CURSO O AÑO QUE APROBO EN EL CENTRO EDUCACION, PRIMARIA, SECUNDARIA O SUPERIOR ? Escriba correspondiente el número del último grado, curso o año	11	Ninguno <input type="checkbox"/> 00 Centro de Alfabetización 1 --- Primaria 2 ---	Secundaria 3 --- Superior 4 --- Se ignora <input type="checkbox"/> 99

Figura 3

En estos casos se pudo conocer el nivel escolar, pero el último grado tuvo que ser asignado para ignorado (el porcentaje era muy elevado como para permitir una asignación automática). Para eliminar este tipo de error, la solución es precodificar el grado para cada nivel, aumentando el espacio necesario, o bien separar nivel y grado en dos preguntas distintas, ambas precodificadas, como en los ejemplos en las Figuras 4 y 5.

PARTE B. CARACTERISTICAS EDUCACIONALES (Para personas de	
<p>¿CUAL ES EL GRADO O AÑO DE ESTUDIO MAS ALTO APROBADO?</p> <p>(MARQUE LA CASILLA DE GRADO O AÑO APROBADO EN EL NIVEL EDUCACIONAL CORRESPONDIENTE).</p> <p>PARA LA EDUCACION DE ADULTOS BUSQUE LA TABLA DE CONVERSIONES EN SU MANUAL DE INSTRUCCIONES.</p>	<p>NINGUN GRADO APROBADO <input type="checkbox"/> 00 (Pase a la 11)</p>
	<p>1 PRIMARIA</p> <p><input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p> <p>1 2 3 4 5 6</p>
	<p>2 SECUNDARIA BASICA</p> <p><input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p> <p>7 8 9 10</p>
	<p>3 OBRERO CALIFICADO</p> <p><input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p> <p>1 2 3 4</p>
	<p>4 PRE-UNIVERSITARIO</p> <p><input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p> <p>10 11 12 13</p>
	<p>5 TECNICO MEDIO</p> <p><input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p> <p>1 2 3 4 5</p>
	<p>6 MAGISTERIO</p> <p><input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p> <p>1 2 3 4 5</p>
<p>7 SUPERIOR O UNIVERSITARIO</p> <p><input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p> <p>1 2 3 4</p> <p><input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p> <p>5 6 7</p>	

Figura 4

8. ¿Asiste o asistió a algún establecimiento educacional?	
Asiste <input type="checkbox"/>	Nunca asistió <input type="checkbox"/>
No asiste pero asistió <input type="checkbox"/>	Ignorado <input type="checkbox"/>
9. Si contestó "Asiste" o "No asiste pero asistió":	
a) ¿Cuál es el nivel que cursa o el más alto que cursó?	
Preescolar <input type="checkbox"/>	Técnica o industrial <input type="checkbox"/>
Primario <input type="checkbox"/>	Otra enseñanza media <input type="checkbox"/>
Bachillerato <input type="checkbox"/>	Superior no universitaria <input type="checkbox"/>
Comercial <input type="checkbox"/>	Universitaria <input type="checkbox"/>
Normal <input type="checkbox"/>	Ignorado <input type="checkbox"/>
b) ¿Completó ese nivel?	
Sí <input type="checkbox"/>	No <input type="checkbox"/>
Ignorado <input type="checkbox"/>	
c) ¿Cuál es el último año o grado aprobado en ese nivel?	
<i>Para las personas que aún no hayan aprobado el primer año o grado, marque "Ninguno".</i>	
Ninguno <input type="checkbox"/>	5º <input type="checkbox"/>
1º <input type="checkbox"/>	6º <input type="checkbox"/>
2º <input type="checkbox"/>	7º <input type="checkbox"/>
3º <input type="checkbox"/>	8º o mayor <input type="checkbox"/>
4º <input type="checkbox"/>	Ignorado <input type="checkbox"/>

Figura 5

d) Formato de la boleta

Las preguntas para una misma persona deben estar en la misma página. Sin embargo, también se usa el formato de varias personas en una misma página (formato columnar). Lo que no es aconsejable es tener varias personas en una misma página, con continuación en páginas posteriores. Eso porque la tendencia a olvidar las preguntas de las demás páginas es grande, toda vez que lo normal sería contestar las preguntas de la primera página para todas las personas.

Algo similar a eso sucedió con un censo que tenía el bloque de fecundidad separado al final del cuestionario. Para cada mujer de quince años y más se debería llenar una columna con las informaciones pertinentes y el número de la mujer en la parte inicial de la boleta, para que se pudiera hacer la relación. Este diseño aumentó mucho el nivel de no-respuesta en la fecundidad, y para los que contestaron, los problemas de pareo entre el bloque principal y el bloque de fecundidad fueron bastante frecuentes.

e) Color de la boleta

Tanto las experiencias como las pruebas reales probaron el aumento de la eficiencia de utilizar un cuestionario con un color básico, en comparación a uno blanco y negro. El color suave, sepia, azul o verde claro, sirve para descansar la vista de los operadores, y el efecto de sombreado sobre algunas partes resalta la atención sobre los asuntos importantes.

B.3- Censo Piloto

El Censo Piloto es una prueba extremadamente valiosa para todos los aspectos nuevos del censo, antes de ejecutar el censo definitivo. En el Censo Piloto se deben probar los procedimientos de recorrido y cobertura, los mapas cartográficos, el formato del cuestionario, comportamiento de nuevas preguntas, entrenamiento de los encuestadores, tiempos para la entrevista, en fin, TODAS las etapas de un censo real.

Varios países usaron esa herramienta con éxito, probando todas las etapas de un censo, hasta el empadronamiento. Sin embargo, el Censo Piloto no termina con el empadronamiento, como lamentablemente parece que sucedió con estos censos, porque las etapas siguientes a la recolección no fueron probadas (con raras excepciones). Los cuestionarios deben pasar por TODOS los procesos, manuales y automatizados, incluyendo la tabulación de los cuadros y análisis de los resultados.

Es necesario verificar los procedimientos de recepción de los cuestionarios, agrupación en carpetas, numeración, organización en las estanterías, codificación, captura de datos (digitación o lectura óptica), y todo el proceso de computación. El archivo del Censo Piloto debe ser usado para probar los programas de computador, desde la entrada de datos hasta la tabulación, no sólo con el objeto de determinar la exactitud de los programas, sino también para verificar la adecuación del equipo, los tiempos involucrados, la cantidad de operadores requeridos, etc.

Son muchas las razones por las cuales no se usa el Censo Piloto en toda su extensión, empezando por la escasez de recursos de personal. En otros casos, también frecuentes, es la inexistencia del equipo computacional, que en esta época todavía no está instalado, lo que impide completar la operación. Por último, si hay personal y equipo, no hay tiempo para definir y hacer los programas. En fin, siempre se puede justificar la no utilización del Censo Piloto en las etapas de computación, pero debe destacarse que ésta es una oportunidad única de probar el sistema con información real, y que ella no se debería

perder.

B.4- Capacitación de Personal

Sin lugar a dudas, este es el punto clave para la obtención de un buen resultado censal. Todo lo que se pueda hacer en este rubro siempre será poco, si se considera la magnitud de la operación debida a la cantidad de personal involucrado. El entrenamiento de los empadronadores es fundamental, pero no se debe olvidar a los codificadores, digitadores, programadores, y expertos investigadores.

Con relación a los programadores, el método y tipo de capacitación varía un poco, porque no son muchos, ni tampoco van a trabajar masivamente como en una línea de producción. Hay que enseñarles con mucha profundidad todos los paquetes con los cuales van a trabajar (ver capítulo IV), y que tengan un dominio total del equipo a ser utilizado, principalmente en el sistema operativo.

A los analistas investigadores es importante entregarles una capacitación en los paquetes de análisis de datos, tipo SPSS o SAS, para que ellos puedan obtener sus propios resultados, independientemente de la actuación de los programadores, casi siempre escasos o totalmente ocupados durante el desarrollo de los programas para el censo. Sin embargo, estos paquetes no son la panacea universal para todos los problemas, principalmente cuando se trata de archivos censales con varios millones de registros.

Los técnicos involucrados en la corrección automática también deben ser entrenados en la metodología y lenguaje de programación a ser utilizados, para que puedan dialogar con facilidad con los programadores.

Los empadronadores, codificadores y digitadores tienen como denominador común el gran volumen de material a ser manipulado, y lo repetitivo de sus tareas. Cualquier problema en estas etapas tiene muchas posibilidades de transformarse en un error sistemático, basta para ello que uno de estos operadores no entienda exactamente qué hacer en un determinado caso. En secciones anteriores ya se mencionaron ejemplos en donde un buen entrenamiento a los empadronadores podría haber solucionado los problemas. Quizás el ejemplo clásico sea el problema de no-respuesta en la parte de fecundidad y mortalidad, que en general es un alto porcentaje, sobre todo para las edades menores y solteras o estudiantes. Eso suele ocurrir por la tendencia existente en los empadronadores - quiénes a veces imbuidos por prejuicios socio-culturales, que afectan su propio comportamiento o que son causales de temor a herir los sentimientos de otras personas - no efectúan las preguntas a tales individuos. Al no efectuar las preguntas, los

entrevistadores tampoco registran el cero para el número de hijos tenidos; consecuentemente se eleva el número de ignorados.

Para los digitadores, como para los empadronadores, un buen entrenamiento también resulta importante. Por ejemplo, un cuestionario de un censo tenía como último ítem, luego de las preguntas sobre características socioeconómicas, una pregunta para todas las personas, sobre deficiencias físicas. Como eran muy pocos los casos, los digitadores terminaron por no ingresarla, sobre todo para las personas no económicamente activas, para las cuales ya habían saltado el bloque anterior. En este caso, un buen diseño del cuestionario habría solucionado el problema, pero de todas maneras, el entrenamiento también debería haberlo corregido.

En los casos de la codificación y digitación, los errores sistemáticos (siempre y cuando no sean de gran volumen) pueden ser superados si se detectan a tiempo, reejecutándose el grupo de cuestionarios con error. En cambio, en la etapa de empadronamiento el error es irreversible y no hay como corregirlo en las etapas posteriores. Por este motivo se debe poner gran énfasis en la capacitación de los encuestadores.

B.5- Muestra de Adelanto

Una muestra de adelanto es una manera de asegurar resultados oportunos, y por supuesto, sirve para adelantar muchos análisis y pruebas.

La conveniencia de una muestra para resultados adelantados, del punto de vista computacional, está relacionada con el mismo propósito del Censo Piloto, es decir, probar todas las etapas de procesamiento con un archivo real.

C- Empadronamiento

Este es el punto más importante de un Censo, en donde se deben concentrar todos los esfuerzos para hacerlo de una manera óptima. Cualquier error en este proceso es irreparable, y dependiendo de su extensión, puede comprometer gravemente la calidad de los resultados censales.

Ya fueron mencionados algunos temas que se relacionan con el empadronamiento, tales como el Diseño de la Boleta y la Capacitación a los encuestadores. Sin embargo, cabe resaltar aquí un problema que afecta directamente al

procesamiento de datos: la cantidad de ignorados (falta de información) en las respuestas.

El bajo número de ignorados es fundamental cuando se trata de corrección automática (ver ítem D.5), y la mejor manera de controlarlo es a través de los supervisores de recolección. Ellos deben examinar el trabajo de cada empadronador, teniendo especial cuidado con las faltas de respuesta, si es posible obligando el retorno al campo para completar los cuestionarios.

Se observó que en un censo había empadronadores que, al encontrar una vivienda ocupada pero vacía en el momento de la visita, preguntaban al vecino el número de ocupantes y llenaban un cuestionario sólo con los nombres, relación de parentesco y a veces el sexo, dejando todas las demás preguntas en blanco. Algo muy parecido pasó en otro censo que tenía mucha presión de sus dirigentes para que fuera hecho en un solo día. Como consecuencia se pudo observar un aumento desmedido de viviendas desocupadas y variables sin respuestas para las personas.

D- Procesamiento

D.1- Organización del Material

Para la recolección, hay que enviar a los agentes todo el material necesario, como cuestionarios, manuales del supervisor y empadronador, formularios de control, lápices, etc., en cantidades proporcionales a su área de acción. El material impreso debe ser, siempre que sea posible, de un mismo tamaño, o de tamaños múltiples, para facilitar el aprovechamiento máximo del papel en la impresión, su acondicionamiento para transporte, y el archivo. Con el objeto de calcular las cantidades de material a ser enviadas a cada localidad, se debe usar un programa de computador que parte de la población estimada de cada localidad y usa factores de multiplicación por tipo de material. Estos factores deben llevar ajustes porcentuales para compensar los gastos adicionales en la capacitación y eventuales pérdidas por inutilización. Este programa estima también el peso total del material a ser enviado (para calcular los fletes o valores de los contratos de transporte) y, como subproducto, la cantidad total de material a ser impreso.

Luego de terminada la recolección, las boletas censales deben ser agrupadas en unidades mayores para facilitar el desplazamiento entre las diversas estaciones de procesamiento siguientes, y facilitar también el control de producción de cada una de las etapas. En general se usan

carpetas con un tamaño prefijado, de manera que contengan un mismo número de cuestionarios. Los cuestionarios deben ser identificados con el número de la carpeta y con un número secuencial de acuerdo a su posición dentro de la misma.

Este acondicionamiento puede ser ejecutado descentralizadamente, por los agentes de recolección, o cuando se reciban los cuestionarios en el centro de producción.

La ventaja de su ejecución en el campo, además de los beneficios de un proceso descentralizado (ver ítem IV.3), es la de mejorar las condiciones de transporte del material, proporcionando mayor protección a las boletas contra extravíos individuales, daños, y que no se mezclen.

Por otro lado, la centralización del agrupamiento en carpetas permite que este proceso sea controlado automáticamente por computador, a través de un sistema que, a partir de la información del material recibido, emite listados con la identificación del contenido y cantidad de cuestionarios de cada carpeta. 4/

El sistema automatizado genera como subproductos:

1- Impresión de la identificación de las carpetas en etiquetas engomadas -

Esta identificación automática garantiza la uniformidad de las informaciones comunes a los cuestionarios de cada carpeta (básicamente la ubicación geográfica hasta un nivel intermedio de sector o sección), permitiendo la disminución de la digitación de estos datos (una sola vez por carpeta).

2- Generación de un archivo de control para las etapas posteriores -

Este archivo, con la identificación de cada carpeta y el número de boletas en cada una, sirve como auxilio a la administración de las etapas de codificación y digitación.

3- Dimensionamiento de los lotes de producción en las etapas de crítica y tabulación -

Debido al gran volumen de información censal, es necesario dividir el material en lotes de procesamiento, siendo cada lote un conjunto de carpetas. Estos lotes deben tener tamaños similares, con el fin de facilitar los controles de tiempos de procesamiento, y para su

4/ Instituto Brasileiro de Geografia e Estatística, Metodologia do Censo Demográfico de 1980, Rio de Janeiro, 1983.

dimensionamiento se deben tener en cuenta los siguientes parámetros:

- Capacidad máxima de la cinta magnética utilizada.
- Tiempos promedios de ejecución de los programas de crítica y tabulación.
- Agrupación de Departamentos pequeños en un único lote.
- Separación de Departamentos grandes en varios lotes.
- Orden de llegada de los cuestionarios al centro de procesamiento.
- Orden de prioridad de divulgación de resultados.

Las estaciones de procesamiento que manipulan las boletas, es decir, Agrupamiento en carpetas, Codificación y Digitación, deben estar localizadas lo más cerca posible unas de las otras, si posible en un mismo sitio, con el objeto de minimizar el traslado del material censal.

Por limitaciones de espacio físico, debido al gran volumen de personal y material involucrados, a veces se separan estas estaciones en edificios distintos, como fue el caso de más de un país. El transporte del material entre las estaciones de procesamiento provocó problemas de embalaje y desembalaje, necesidad de ambientación del material (en el caso de lectura óptica), y posible deterioro y/o pérdida de cuestionarios, aumentando la necesidad de un control de material y del tiempo de procesamiento, además de los costos de transporte que varían dependiendo de las distancias entre las estaciones.

Finalmente, la organización del material en carpetas es importante también para la etapa de Control Cuantitativo (ver ítem D.4), cuando se pretende garantizar el procesamiento de todos los cuestionarios.

D.2- Codificación Manual

En otras épocas, la etapa de Codificación, era también denominada de Crítica y Codificación Manual, porque los responsables ponían un gran énfasis en el componente de crítica, con una verificación visual de la boleta y una "corrección" de los datos inconsistentes.

Como normalmente esta tarea es realizada por varias personas, es imposible garantizar una uniformidad de criterios en lo que se refiere a la crítica. La experiencia ha mostrado (con el uso del computador para efectuar la crítica y corrección automática), que esta fase debe ser reducida al mínimo, es decir, a la codificación de las preguntas abiertas, evitando la corrección o asignación de

respuestas a no ser de casos obvios y preestablecidos. La crítica manual debe ser restringida a una revisión general de la estructura y ordenamiento de las informaciones por boleta o grupos de boletas, como por ejemplo la verificación de las variables de Ubicación Geográfica y de identificación de los cuestionarios. Sin embargo, hay una crítica previa, de terreno, por el supervisor, la cual debe hacerse siempre, con el fin de evitar los errores sistemáticos de cada empadronador.

La crítica y corrección manual, además de involucrar un innecesario aumento del tiempo de codificación, demanda al codificador mucho de su atención, la cual debería estar totalmente concentrada en la tarea misma de codificación de las preguntas abiertas. Por otra parte, puede introducir también distorsiones en el comportamiento de las variables, de difícil detección y de casi imposible corrección, ya que sería necesario reejecutar la codificación para eliminar los códigos errados y redigitar nuevamente los cuestionarios.

En la parte de Codificación propiamente tal, es importante resaltar la necesidad de un control de calidad por muestreo, principalmente para los códigos de ocupación y rama de actividad. Estas variables, si son codificadas sin un rígido control y procedimientos uniformes establecidos a través de la experiencia del personal sustantivo, pueden ser distorsionadas de tal manera que se torne muy difícil su explotación.

La codificación es una etapa que requiere conocimientos multidisciplinarios. La definición de las categorías de cada variable es una labor netamente sustantiva, mientras que la atribución de los códigos a las categorías debe ser realizada por personal de computación, con el fin de evitar la creación de valores de difícil manipulación (por ejemplo, códigos alfabéticos o el uso de caracteres especiales).

Durante la ejecución, el control del flujo de documentos y la supervisión de la aplicación correcta de los códigos, así como la asistencia a los codificadores en los casos más difíciles, debe haber personal con experiencia en este tipo de trabajo que se haga responsable. En un determinado censo esta labor estaba a cargo sólo del personal de computación, que no poseía los conocimientos suficientes para resolver todos los problemas.

D.3- Grabación

El almacenamiento de la información en un medio magnético para su posterior utilización en computador, tiene actualmente dos opciones que son las más usadas en los censos: la digitación y la lectura óptica. Luego de un

relato de las experiencias existentes en cada tipo, se hace una comparación entre ambos mecanismos, con sugerencias para su utilización.

D.3.1- Digitación

Debido a innumerables razones, que dependen del tipo de máquina que se usa, organización del trabajo, calidad y responsabilidad de los operadores, se puede transcribir otro código que el que indica la boleta, omitir digitaciones, duplicar, etc. Sobre esta fuente de error se debe establecer un control de calidad lo más amplio posible, con verificación de las digitaciones, control del flujo de carpetas con boletas y controles específicos de cada equipo. Por lo tanto, el proceso de grabación convencional debe comprender dos etapas: la digitación y la verificación. La segunda es necesaria para disminuir a un porcentaje insignificante los errores provocados por el proceso de digitación.

a) Digitación

Con la aparición de los terminales "inteligentes", la digitación tuvo una tendencia a sofisticarse, con la inclusión de elementos de crítica en los programas de entrada de datos, y la consiguiente necesidad de asignación de variables por los digitadores. La idea era que el digitador, al encontrarse con una variable inconsistente, la corrigiera usando algunas reglas preestablecidas más la información adicional de la misma persona, o de otras personas, de la misma boleta, dado que las tenía a la mano. Los cuestionarios deberían ser considerados correctos al término de la entrada de datos.

Este tipo de procedimiento es adecuado para una encuesta de dimensiones reducidas, para la cual el método de corrección automática no es aconsejable. Sin embargo, en el caso de los censos demográficos, la corrección de los datos en el momento del ingreso tiene una serie de desventajas, a saber:

1- No hay una uniformidad de criterios en la asignación de valores a las variables con error.

2- La programación de las reglas de crítica aumenta los programas de entrada de datos, disminuyendo la velocidad de ingreso.

3- La velocidad de los digitadores disminuye debido a las constantes paradas para la toma de decisión sobre la corrección.

4- Los digitadores necesitan un mayor entrenamiento y calificación para decidir las condiciones de imputación de

variables.

5- Una vez decidida la asignación de una o más variables, ésta queda "congelada", sin opción de cambios en las reglas de corrección (ver ítem D.5).

Por estas razones, la crítica en la digitación de censos debe limitarse a la eliminación de los errores de digitación, es decir, garantizar la igualdad entre los cuestionarios y el archivo recién grabado. En general cada etapa debe estar capacitada para corregir (minimizar) sus propios errores.

El procedimiento más aceptable actualmente es la crítica de los códigos válidos para cada variable. En caso de error, el programa da la oportunidad al digitador de reingresar el valor defectuoso. Si el nuevo valor coincide con el anterior, eso significa que éste es el valor informado en la boleta, y como tal se lo acepta para posterior corrección en el proceso de asignación automática (ver ítem D.5). Las únicas variables para las cuales se debe exigir total corrección son las variables de Ubicación Geográfica, en cuyo caso no se recomienda la corrección automática, por estar relacionadas directamente con el Control Cuantitativo (ver ítem D.4).

En un determinado censo se empezó la digitación con un programa muy complejo, con muchas de las reglas existentes en el Plan de Asignación Automática, pero luego los responsables se dieron cuenta de los problemas resultantes y cambiaron el programa de entrada para verificar sólo los códigos válidos, aceptando los valores doblemente ingresados.

b) Verificación

La etapa de verificación es una repetición, total o por muestreo, de la digitación, en donde se hace una comparación entre el código digitado por primera vez y el digitado por el verificador. En caso de discrepancia el verificador redigita el código nuevamente. Esta etapa es necesaria para controlar el nivel de error de los digitadores y mantenerlos siempre atentos. También es usada como una herramienta de detección para eliminar a los digitadores ineficientes.

En algunos censos la Verificación fue eliminada con el argumento de que los digitadores estaban bien entrenados, y no había tiempo ni recursos para efectuarla. En este caso es preferible su ejecución por muestreo, y disminuir al máximo el factor de selección, pero mantenerla vigente.

D.3.2- Lectura Optica

El proceso de Lectura Optica fue usado con relativo éxito en algunos países pero en otros hubo problemas, derivados de la sofisticación del método. Resulta que se había comprado una sola lectora, y como no existía una firma de mantención en el país, a cada falla del equipo era necesario que llamar un experto del extranjero, lo que aumentaba el tiempo de parada. Por problemas presupuestarios, el contrato de mantenimiento no fue renovado, lo que complicó aún más las paradas del equipo.

En general, la impresión de los cuestionarios exige una precisión no existente en las imprentas locales, lo que suele obligar a la importación de los cuestionarios. En el caso de un censo, por restricciones presupuestarias fue imposible importar cuestionarios adicionales para ser usados en la lectora, en los casos de rechazo en la lectura de algunos cuestionarios muy maltratados.

D.3.3- Ingreso Optico vs Digitación

"¿Qué método se debe utilizar para el ingreso de la información: una lectora óptica ó un sistema de digitación?". Para contestar esta pregunta con un razonable grado de seguridad, es necesario que se examinen primero las características de cada proceso, y a partir de eso intentar obtener alguna conclusión:

-La digitación involucra una cantidad mayor de personal, que necesita un entrenamiento adecuado, espacio para trabajar, seguridad social, etc. Este costo depende del nivel de salarios de la mano de obra en cada país.

-El tiempo de digitación también es mayor, variando de acuerdo al número de operadores que se tenga.

-La digitación es una operación manual, y por lo tanto, sujeta a errores, dependiendo del grado de entrenamiento de cada digitador.

-El equipo básico de la lectora óptica es mucho más caro que el empleado en digitación.

-La lectora óptica exige ajustes y mantenimiento frecuentes.

-La lectora óptica es muy sensible. Los cuestionarios necesitan un papel de buena calidad, y mucha precisión en su impresión, condiciones que no son satisfechas por las imprentas comunes. Generalmente, los cuestionarios son impresos en el exterior, lo que aumenta la dependencia externa.

-Para la lectora óptica, las respuestas a las preguntas deben ser marcadas con mayor cuidado, y durante todo el proceso censal los cuestionarios necesitan un cuidado especial para evitar que se deterioren.

-El equipo de digitación, sobretodo si son microcomputadores, puede ser aprovechado posteriormente para otros trabajos.

En resumen, la digitación es un proceso más sencillo y barato, mientras que la lectora óptica es más rápida y precisa, pero sólo si se observan todos los controles técnicos necesarios en una forma adecuada. Para los censos de mayor envergadura, el costo de la lectora óptica es relativamente menor y se justifica más fácilmente. Las desventajas del proceso pueden ser resueltas a través de procedimientos cuidadosos, y por lo tanto, su uso es aconsejable.

Para los censos menores, la decisión es más difícil, dependiendo de si se encuentra una manera de abaratar el costo de la lectora, ya sea a través de su utilización en otras encuestas posteriores al censo, o sea por arriendo o "leasing" de la lectora.

D.4- Control Cuantitativo

El Control Cuantitativo, a veces llamado de Control Primario, tiene los siguientes objetivos:

1. Garantizar que se ingresen todas las boletas y que no haya repetición.
2. Garantizar que dentro de una boleta se ingresen todas las personas, y que no haya repetición.
3. Verificar que se obedezca la estructura básica de los registros para cada boleta (control de identificación y estructura).
4. Verificar las variables de Ubicación Geográfica.

El proceso de Control Primario se basa en los datos informados por los jefes de zona o supervisores de empadronamiento, utilizándose los totales de vivienda y personas (a veces por sexo) de cada sector. Estos valores, con sus respectivos códigos de ubicación geográfica, son digitados, verificados y almacenados en un Archivo de Control. Como subproducto de esta digitación se puede obtener muy rápidamente una publicación de resultados provisionarios, con los valores del total de viviendas y personas para todos los niveles geográficos.

Una vez que se tenga el Archivo de Control, el proceso consiste en hacer un pareo entre las informaciones contenidas en este archivo y los totales obtenidos por conteo de los registros de los archivos del censo propiamente tal. Las diferencias son analizadas y corregidas, volviéndose a ejecutar el proceso hasta que desaparezcan, o se reduzcan a un nivel insignificante.

El Control Primario puede ser aplicado en partes, dependiendo de las facilidades operacionales de cada equipo, y de hecho es lo que ocurre con los ítems 2, 3 y 4, en los cuáles el control es ejecutado durante el ingreso de datos (si se hace por digitación). El ítem 1 se ejecuta cuando se tiene un lote de cuestionarios disponibles para procesamiento.

Este proceso es imprescindible en un Censo, sobre todo en razón de su primer objetivo (garantía de procesamiento total de las boletas). Todas las etapas anteriores tienen sus controles manuales, plausibles de error; por eso la necesidad del Control Cuantitativo. Sin embargo, hubo países que no lo ejecutaron, lo que acarreó consecuencias negativas previsibles. En otro caso se empezó sin el Control Primario, pero al compararse los resultados de la primera unidad administrativa con los resultados provisorios, se pudieron apreciar serias dificultades y fue necesario ponerlo en práctica.

D.5- Consistencia de Datos

La etapa de Consistencia tiene como objetivo la eliminación de los errores lógicos existentes en las variables, para evitar que aparezcan incoherencias o absurdos en las tabulaciones, produciéndose un archivo limpio y un informe estadístico de las correcciones hechas. Los errores pueden surgir en cualesquiera de las fases anteriores al almacenamiento: empadronamiento, codificación manual y digitación. La calidad final de los datos en cuanto a la representación de la información recolectada originalmente, depende más del control que se haga de las causas de los errores en las etapas previas a la corrección automática, que del uso sofisticado de ésta.

El procedimiento de Corrección Automática de los errores fue usado con éxito en la gran mayoría de los censos de América Latina. Si se compara con el proceso de Corrección Manual usado anteriormente, él tiene una serie de ventajas tales como evitar la tarea tediosa de volver a los cuestionarios originales para determinar la corrección de un error, economía de tiempo y recursos humanos, uniformidad de criterios, capacidad de cambio en los criterios, reejecución, mantención de la estructura de frecuencias de

los códigos de cada variable, y capacidad de producción de estadísticas de asignación.

La corrección automática de los datos no significa que se mejorará la calidad de los datos en cuanto a la representación de la realidad a que se refieren. Al efectuar la asignación de un valor a una variable, no se intenta acertar la respuesta que dio o habría dado el entrevistado, sino que hacerla consistente con las otras informaciones, manteniendo la distribución de los valores de la variable a lo largo del archivo.

Sin embargo, hay que advertir, en primer lugar, que su aplicación debe ser dirigida por quienes tengan conocimientos y experiencias sobre el comportamiento de las variables, evitando la asignación indiscriminada. En segundo lugar, no se la debe usar para tratamiento de errores sistemáticos o de alto porcentaje. Para efectos de orientación práctica, la asignación de hasta el 2 por ciento de los casos que deben responder a la variable es considerada un buen objetivo, y de 2 a 5 por ciento es considerada relativamente aceptable (estos porcentajes son distintos de acuerdo a la importancia de cada variable dentro de la boleta censal). Cuando los errores sobrepasan el 5 por ciento es recomendable dejar el archivo con un dato malo o ignorado, conforme sea el caso, y no intentar corregirlo en esta variable.

La tarea crucial en esta etapa es la confección de un Plan de Consistencia, conteniendo todas las reglas de verificación de rangos de variables y de relación entre variables, así como los criterios de asignación en caso de error. Para su establecimiento, hay que examinar la distribución de frecuencias de las variables de un archivo real, por ejemplo el Censo Piloto, la Muestra de Adelanto, o una parte inicial del censo definitivo. Esto se hace a través de la ejecución de un programa específico, que calcula las frecuencias, absolutas y relativas, de los códigos de cada variable (a estos valores se les llaman de Marginales). La mayor o menor concentración de códigos inválidos va a determinar el grado de sofisticación del programa.

Como requisito adicional se debe preparar un archivo de prueba con todos los errores posibles, además de algunos casos sin error, para verificar la perfecta ejecución del programa. Este archivo deberá mantenerse vigente durante toda la etapa de corrección. A cada cambio en el programa para aplicar nuevas reglas o corregir las antiguas, debe corresponder una agregación de los casos correspondientes en el archivo de pruebas, y una nueva ejecución para probar los cambios en el programa.

El programa de distribución de frecuencias debe ser ejecutado también sobre el archivo limpio, con el objeto de comprobar la ausencia de códigos inválidos, y verificar la mantención de las estructuras de código de cada variable.

Por último, el programa de asignación debe producir un informe estadístico sobre sus acciones, conteniendo informaciones tales como:

- .número de boletas y personas leídas y grabadas.
- .frecuencia de ocurrencia de cada error.
- .frecuencia de ocurrencia de errores por variable.

Con este material, el personal sustantivo debe mantener un control rígido sobre el funcionamiento del programa de corrección automática, analizando los posibles problemas y efectuando cambios, si es necesario, a las reglas de crítica y asignación.

Durante la etapa inicial de aplicación es aconsejable ejecutar el programa de corrección sobre el archivo limpio, para verificar si realmente las asignaciones son consistentes. En esta segunda ejecución no deberá haber ninguna asignación.

Los criterios de asignación y los porcentajes de error encontrados deben ser puestos a disposición de los usuarios, y si es posible, deben ser publicados en conjunto con los tabulados, de modo que permitan un análisis más riguroso de los resultados.

Los problemas encontrados en la aplicación de la Asignación Automática, con mayor o menor frecuencia, fueron los siguientes:

1- Inexistencia de un archivo de pruebas -

Esta ha sido una situación bastante frecuente, que tuvo como consecuencia el hecho de que las pruebas se han aplicado a un subconjunto de un archivo real que no poseía todos los errores. Se procedió en un sentido contrario al recomendable: los errores se detectaron (o no) durante la ejecución del proceso, y no en la fase de desarrollo, situación que contribuyó, además, a obvias pérdidas de tiempo.

2- Eliminación del examen de los informes de asignación y distribución de frecuencias luego de la corrección -

Por un defecto no detectado en el programa, esto ocasionó que algunos errores permanecieran en el archivo limpio. La simple verificación de las frecuencias de salida habría eliminado este problema.

3- Asignación para errores de alto porcentaje -
La única asignación permitida en estos casos es para el código de ignorado, pero presionados por presentar resultados comparables a otras fuentes de datos, los responsables programaron un "método científico" para inventar el valor de una pregunta de las más importantes de la boleta censal, la cual era "¿qué hizo en la semana pasada?".

4- Inexistencia de documentación sobre los criterios de asignación -

En general, siempre existió una versión preliminar del Plan de Consistencia, con la cual se diseñó inicialmente el programa de asignación. A su vez, este programa sufrió múltiples cambios para adaptarse mejor a la realidad del censo, pero el Plan de Consistencia permaneció desactualizado. Además, fueron raros los países que publicaron algo sobre los criterios de asignación y respectivos porcentajes de error.

5- Eliminación de los archivos de pre-corrección -

Estos deben ser siempre guardados, con el fin de permitir una reejecución del proceso en caso de que se cambien los criterios, o se detecte algún sesgo en las asignaciones.

D.6- Análisis de los Resultados

Una vez ejecutado el proceso de limpieza, la etapa siguiente debería ser la de Tabulación, a través de la cual se imprimen las tablas y cuadros para publicación de los resultados. Antes de la Tabulación, debe tener lugar una etapa generalmente inexistente en la mayoría de los censos, la cual es la de Análisis de los Resultados. Ésta consiste en la producción de una serie de indicadores que permitan a los investigadores verificar la existencia de algún problema en los datos.

Es evidente que el Plan de Tabulaciones también contiene elementos de juicio, pero no siempre están en el formato más indicado para este tipo de estudio, por eso es importante que se definan cuáles son los índices, indicadores, resultados comparativos, etc., que componen el Plan de Análisis, y que estos resultados sean producidos ANTES de la Tabulación.

D.7- Tabulación y Publicación Básica

La Tabulación es una tarea bastante directa, no debería tener problemas. Trátase de ejecutar un programa, o un

conjunto de programas, para calcular e imprimir los cuadros con los resultados censales. Sin embargo, lo que pasa con la mayoría de los países es que, como ésta es una de las últimas tareas censales y aparentemente una de las más sencillas, siempre es postergada hasta el momento en que ya está lo suficientemente retrasada como para forzar una definición.

El mayor problema encontrado en todos los censos de los países latinoamericanos en la década de los ochenta fue el retraso en la Definición del Plan de Tabulaciones, el que debe estar listo una vez que se tenga el diseño definitivo de la boleta censal. Como siempre se usa el Plan del censo anterior como base, agregándose o eliminándose cuadros, no hay razón para que no se obtenga esta definición a tiempo. Con esta relegación del Plan para el último momento, la programación de las tablas es perjudicada, ya que debido a la urgencia de obtención de los resultados se ocasionan errores de programación perfectamente evitables.

Un segundo problema, que es consecuencia de la inexistencia de la etapa anterior (Análisis de los Resultados), surge cuando se usan los cuadros del Plan de Tabulaciones para la verificación de posibles defectos en los datos o programas anteriores.

Lo ideal sería que los analistas definieran con bastante anticipación, un Plan de Tabulaciones compuesto de dos grupos: los tabulados para análisis y los cuadros para divulgación.

E- Explotación de la Información

E.1- Almacenamiento en Base de Datos

Al publicarse los resultados censales, o incluso antes de su difusión, los usuarios e investigadores, internos y externos al organismo encargado de la operación censal, suelen solicitar "tabulaciones especiales", es decir, cuadros que no fueron incluidos en el Plan Tabular. Es imposible que este Plan Tabular de Divulgación contemple todos los cruces de información que satisfagan los intereses de los usuarios, por eso es necesario hacer flexible el proceso de producción de resultados "ad hoc", a través de la creación de una base de datos censales.

Estos sistemas, con menor o mayor grado de sofisticación, tienen como objetivo eliminar las distancias entre el dato censal y el usuario final, proporcionando una diseminación eficaz de la información, con la utilización de

terminales para acceso directo a la base de datos.

Para este almacenamiento se puede sugerir la utilización del sistema RAPID que, entre otras ventajas, tiene desarrolladas las "interfaces" para los lenguajes SPSS y SAS.

E.2- Creación de Archivos Derivados

Acompañando la creación de la base de datos, la tendencia actual se orienta en el sentido de reducir al mínimo el número de tablas publicadas, y producir archivos magnéticos (cintas) con los microdatos, que son distribuidos por un precio nominal, con una documentación apropiada para facilitar su explotación.

Aquí también aparece la necesidad de procesamiento de sub-archivos para estudio de temas específicos (migración, fecundidad, etc). En este caso, para facilitar la explotación y disminuir los tiempos de procesamiento, el proceso aconsejado es la creación de un archivo sólo con los casos pertinentes (mujeres de 15 años y más, por ejemplo), y con las variables involucradas en el estudio (las de fecundidad y alguna variable básica para estratificación de los resultados). Este proceso tiene la gran ventaja de producir archivos bastante reducidos y, por lo tanto, más adecuados para la utilización de los paquetes de programas, pero exige que el usuario (investigador) elija con anterioridad el universo de casos y variables a ser investigados.

Otra manera de aminorar el costo y el tamaño de los archivos censales es a través de la producción de muestras, que dependiendo de sus magnitudes, puedan ser explotadas con un razonable grado de fidelidad en los resultados. En general, estas muestras son suficientes para la investigación, a escala del país o de divisiones administrativas mayores, de la mayoría de las variables censales.

E.3- Divulgación de la Información

La creación de bases de datos y de archivos derivados deben tener una publicidad y propaganda adecuadas, con el objeto de poder informar a todos los posibles usuarios acerca de su existencia. De nada vale tener una buena base de datos o muestras impecables si el usuario final no las conoce ni sabe como usarlas.

IV- RECOMENDACIONES ADICIONALES

En este capítulo se consideran algunos asuntos que, por su naturaleza, se encuadran en más de un ítem del capítulo anterior, como es el caso de la Utilización de Paquetes de programación. También se discuten el aprovechamiento de los microcomputadores en el procesamiento censal, y el tópico de la Descentralización de operaciones, que parece ser la tendencia dominante para la ronda de los censos de los años 90.

IV.1- Utilización de Paquetes

La falta de la cantidad necesaria de programadores experimentados en un procesamiento censal, agregada a la escasez de tiempo disponible, hacen que la utilización de paquetes para programación sea una solución inevitable.

Básicamente, los paquetes disponibles para procesamiento estadístico tienen las siguientes ventajas:

- 1- Facilidad de aprendizaje.
- 2- Facilidad y rapidez de programación, con la consiguiente disminución de la posibilidad de errores, y la flexibilidad para efectuar rápidamente cambios en los programas.
- 3- Comandos específicos dirigidos a satisfacer las necesidades de los usuarios.
- 4- Han sido extensamente experimentados, lo que implica una garantía contra el mal funcionamiento.
- 5- Su instalación está disponible en la mayoría de los equipos.
- 6- Disponibilidad de personal con conocimiento para apoyo en su utilización.

Las desventajas de los paquetes pueden ser resumidas en dos:

- 1- No son totalmente flexibles para satisfacer exactamente, y a medida, las necesidades de los usuarios, lo que significa que éstos deben adaptarse a las restricciones del paquete.
- 2- A veces no son los más eficientes en tiempo de ejecución, ocupando mucho del equipo.

La opinión generalizada es que las ventajas sobrepasan por mucho a las desventajas, ya sea porque en términos de limitación y rigidez de programación, los paquetes se adecúan a la mayoría de las necesidades. En el rubro de eficiencia en la ejecución, los equipos son cada vez más rápidos y baratos, lo que disminuye el valor de este argumento en contra.

La experiencia de los diversos países en la utilización de paquetes fue altamente positiva, y debe ser seguida. Los más usados fueron el CONCOR para la asignación automática y el CENTS para la tabulación, ambos desarrollados por el Bureau de Censos de los Estados Unidos. Algunos países usaron también el SISTMARG (desarrollado en el CELADE) para la obtención de la distribución de frecuencias de variables, o el SPSS para frecuencias simples y algunos cruces básicos. En el censo de un país se usaron paquetes desarrollados internamente en su Instituto de Estadística, para la crítica, asignación, tabulación, y para la impresión de los tabulados.

IV.2- Microcomputadores

La tecnología computacional está evolucionando a una velocidad tal que se torna difícil prever sus posibilidades de utilización en los censos. Los microcomputadores están aumentando su capacidad de procesamiento y disminuyendo sus costos, haciéndolos accesibles para las actividades más básicas.

Actualmente, los microcomputadores son utilizados en la actividad censal para análisis de datos y estudios de investigación, con posterioridad al levantamiento propiamente tal. Algunos países usaron minicomputadores en el sistema de entrada de datos, con una unidad concentradora central y varios terminales para la digitación. De acuerdo a los estándares recientes, estos minicomputadores no tienen ninguna diferencia con los "micros" en términos de capacidad, y de este modo se considera el ingreso de datos como un área de actuación para los "micros", ahora y en el futuro.

Es evidente que la mayor o menor utilización de los "micros" en un censo depende del volumen de información a ser manejada. En el caso de los países grandes, es difícil pensar en la sustitución de los equipos actuales, principalmente para las etapas de consistencia de datos y tabulación. Para los países de menor población, la utilización de los "micros" en la consistencia y tabulación dependería más del desarrollo de paquetes específicos, que de la capacidad de procesamiento de los equipos.

Aunque no directamente relacionados en la línea de producción de datos censales, los microcomputadores tienen reservados para ellos papeles importantes dentro de las actividades de apoyo, y se puede decir que serán indispensables en un futuro muy próximo. Por ejemplo, los "micros" pueden ser utilizados en el control de cronogramas y del presupuesto, llevar la contabilidad censal, procesamiento de textos para toda la documentación necesaria, programas para control de material, manuales de códigos, etc.

La mayor utilización de los "micros" en el procesamiento está también relacionada con el tema de la descentralización de las actividades censales.

IV.3- Descentralización

Todos los censos, en forma general, ejecutan el empadronamiento en cada vivienda y envían los cuestionarios para ser procesados centralizadamente. Es decir, ya existe algo de descentralizado en el proceso censal (aunque obligatorio, porque sería muy difícil hacer que las personas vayan a los empadronadores). La discusión es en qué medida se podría aumentar esa descentralización, transfiriendo operaciones para los locales de recolección, o creando subcentros de procesamiento.

¿Cuáles son las ventajas de la descentralización?. En primer lugar, la reducción del flujo y transporte de documentos. En segundo lugar, la división de un gran proceso en varios procesos menores (iguales), que los hace más manejables. En tercer lugar, los cuestionarios serían procesados más cerca del punto de recolección, por personal conocedor de los problemas de cada área, facilitando así la consistencia de la información.

Como desventajas puede mencionarse el costo para la adquisición de los varios equipos a ser distribuidos, y las dificultades para control del proceso. Existe también el problema de la mantención de una uniformidad de criterios, principalmente en la etapa de codificación, así como la dificultad en aplicar las correcciones y ajustes a las distintas versiones de los programas. Algunos países tendrían una desventaja adicional para encontrar personal calificado en las provincias, obligando al traslado de personal de la sede.

Las grandes etapas del procesamiento censal, luego del empadronamiento, son: codificación, digitación, control cuantitativo, consistencia automática, y tabulación. Las tres primeras necesitan físicamente de las boletas para su ejecución, por lo tanto, al descentralizar, sería mejor

hacerlas las tres juntas, so pena de no cumplirse la primera ventaja (reducir el flujo de documentos). La descentralización de la digitación está íntimamente relacionada con la utilización y aprovechamiento de los microcomputadores, su decisión va a depender de la disponibilidad de los equipos, el entrenamiento del personal, etc. Si el ingreso es realizado por lectura óptica, lo aconsejable es que no se descentralice, porque la lectora (o las lectoras) deben estar concentradas para la optimización del proceso, y hacerlo sólo con la codificación no valdría la pena.

Para descentralizar también las etapas de corrección automática y tabulación, sería necesario que los equipos en los subcentros fueran lo suficientemente grandes para soportar estos procesos. Como los microcomputadores actualmente no poseen una gran capacidad de almacenamiento de datos y tampoco tienen paquetes estadísticos para la crítica y tabulación, la descentralización de estos procesos no es aconsejable. En los casos en que los subcentros puedan contar con un computador de mayor capacidad, aún así la descentralización no es aconsejable porque no se tendría ninguna de las ventajas, solamente las desventajas.

Resumiendo, la tendencia a la descentralización es un hecho, que debería ser seguido, pero tomándose los cuidados adecuados. Al descentralizarse la codificación, también debe hacerse lo mismo con la digitación y el control cuantitativo. La corrección automática y tabulación deben ser centralizadas. De todas maneras, el Censo no debería servir como prueba para el proceso de descentralización, sino que ésta debería ser experimentada antes, en una encuesta de menor magnitud, o bien en el Censo Piloto.

En el censo de un país se usó la descentralización con éxito, al tenerse dos unidades de procesamiento (en sus dos ciudades mayores), para codificación, ingreso de datos y parte del control cuantitativo. Los archivos así obtenidos fueron transmitidos por telecomunicación al computador central, que procesó centralizadamente la parte final del control cuantitativo, la consistencia automática y la tabulación.

V- CONCLUSION

La experiencia de los censos de los países de América Latina en la década del ochenta, aunque heterogénea, debe utilizarse para mejorar la ejecución de las futuras operaciones. En un intento por resumir todo lo discutido, pueden mencionarse las recomendaciones más importantes con el objeto de garantizar la calidad de los resultados.

En primer lugar, en las etapas de Planeamiento y Preparación, se requiere tener un experto en computación involucrado desde su inicio. Es conveniente asignar el máximo posible de presupuesto a las tareas de computación, inclusive para compra de material y contratos de mantención de los equipos. Es preciso elaborar un cronograma con todos los eventos, y mantenerlo actualizado, dejando establecido quiénes son los responsables de cada tarea. Se debe tener especial cuidado en el diseño de la boleta. Es imperioso preparar de la mejor manera posible a los empadronadores.

En la etapa de Empadronamiento, debe buscarse la eliminación total de las no-respuestas a través de un control por los supervisores. Ha de tenerse un Control Cuantitativo. Hay que evitar las causas de los errores en las etapas previas a la corrección automática y analizar cuidadosamente los resultados de la corrección automática. Por último, definir, con anticipación, el Plan de Tabulaciones de Análisis y de Divulgación.

De todas estas recomendaciones, sin duda la mayor importancia se concentra en la realización de un buen empadronamiento, porque no hay sistema de procesamiento, por sofisticado que sea, que consiga mejorar la calidad de los datos censales en cuanto a la representación de la realidad a partir de un levantamiento mal ejecutado.

Finalmente, se sugiere la utilización de paquetes para agilizar la programación, y se consideran las tendencias futuras de aumentar la utilización de los microcomputadores, con una descentralización de actividades.

VI- BIBLIOGRAFIA

Bureau of the Census, POPSTAN- Un estudio de caso para los Censos de Población y Habitación de 1980, Washington, 1979.

Documentación producida sobre los Censos por los Institutos de Estadística de América Latina, archivos internos del CELADE.

Instituto Brasileiro de Geografia e Estatística, Metodologia do Censo Demográfico de 1980, Rio de Janeiro, 1983.

Silva, A., Plan Básico propuesto para la Corrección de los Datos del Censo de Población y Vivienda de 1982, Quito, inédito.

Silva, A., Informes de misiones de asistencia técnica de CELADE a los países de América Latina, Santiago de Chile, circulación interna.

Tacla, O., Breves Comentarios, según etapas censales, con especial énfasis en ingreso de datos, Buenos Aires, mayo de 1985, presentado en el Taller de Análisis y Evaluación de los Censos de Población del 80.