

NACIONES UNIDAS

CONSEJO
ECONOMICO
Y SOCIAL

PROPIEDAD DE
LA BIBLIOTECA



C.1



LIMITADO

ST/ECLA/Conf.32/L.9/Add.1
23 de mayo de 1968

ESPAÑOL
ORIGINAL: INGLES

SEMINARIO SOBRE ORGANIZACION Y
LEVANTAMIENTO DE CENSOS DE POBLACION
Y HABITACION PARA AMERICA LATINA

Organizado por la Comisión Económica para América Latina, la Oficina de Estadística y la Oficina de Cooperación Técnica de las Naciones Unidas, con la colaboración del Instituto Interamericano de Estadística, el Centro Latinoamericano de Demografía y la Dirección de Censos de los Estados Unidos

Santiago de Chile, 20 a 31 de mayo de 1968

ELABORACION ELECTRONICA DE DATOS CENSALES

II. Algunos principios para el uso de computadoras en la elaboración de datos censales

Preparado por la Oficina de Estadística de las Naciones Unidas



INDICE

	<u>Párrafos</u>	<u>Página</u>
A. Introducción	1	1
B. Uso óptimo de la capacidad	2- 8	1
C. Descubrimiento y corrección de errores	9-21	5
D. Obtención mediante computadora de los cuadros definitivos para su reproducción mediante foto-offset	22-24	8
E. Resumen	25-26	9
Anexo		11

A. Introducción

1. Casi todos los censos de la serie de 1970, así como los posteriores, se elaborarán electrónicamente. En algunos casos, las operaciones de entrada/salida se automatizarán en alto grado por el uso de dispositivos para la lectura de marcas o la lectura óptica de caracteres, en las operaciones de entrada, y de dispositivos electrónicos de composición fotográfica, en las operaciones de salida. En este trabajo se analizarán algunos principios de la técnica de elaboración, con el fin primordial de ofrecer las explicaciones y orientaciones necesarias a los países que se inician en estas actividades.

B. Uso óptimo de la capacidad

2. Todas las computadoras tienen una unidad central de elaboración que contiene la memoria central, es decir, el dispositivo que almacena el programa que se ejecutará, junto con los datos que se elaborarán (datos de entrada y datos intermedios) y los definitivos que resultarán de la elaboración (datos de salida). La unidad central de elaboración se complementa con un conjunto de dispositivos de entrada/salida, que constituye el equipo periférico o auxiliar que incluye unidades para leer o perforar tarjetas o cintas de papel, mecanismos propulsores de cintas magnéticas, dispositivos para almacenamiento de acceso directo (discos magnéticos, unidades de tambor, etc.), dispositivos de lectura óptica de caracteres, de impresión, etc.

3. La manipulación de los datos dentro de la memoria central es mucho más rápida que la lectura, perforación y escritura de los datos en el equipo auxiliar. Como técnica adecuada de programación se aconseja, en lo posible, mantener siempre ocupada la memoria central, sin dejarla ociosa mientras se espera que se lea o escriba algún registro. La lectura y perforación de tarjetas o cintas de papel, así como la lectura de marcas o caracteres, son mucho más lentas que la lectura y escritura de registros magnéticos (en cintas, discos, etc.). Lo habitual, por lo tanto, es registrar los datos de entrada que deben usarse más de una vez, en cintas magnéticas u otros medios magnéticos de almacenamiento. Incluso se dan casos en que se usa una computadora

/especial, más

especial, más pequeña, para operaciones comparativamente lentas como el paso de tarjeta a cinta, de cinta a tarjeta y de cinta a impresión, dejando a la computadora más grande la tarea de comunicarse con el mundo "exterior" utilizando exclusivamente cintas magnéticas.

4. Al tabular los datos censales, normalmente se usa parte de la memoria central para preparar simultáneamente un conjunto de cuadros diferentes (véase el anexo). En algunos casos pueden reservarse partes de la memoria central para todos los cuadros requeridos; de hacerse así, para la tabulación sólo se necesitará una operación de entrada para cada registro, y se evitará la necesidad de seleccionar los registros de entrada. Si en la memoria central no hay espacio suficiente para hacer la tabulación en una sola pasada, pueden aplicarse los métodos siguientes, por separado o en combinación:

- 1) Repetir las pasadas de los datos de entrada, sin clasificación y con programas modificados, hasta preparar todos los cuadros básicos. Se dice que un cuadro es básico cuando no es una suma de otros cuadros.
- 2) Clasificar los registros de entrada antes de cada tabulación, entre tabulaciones o haciendo ambas cosas. En el caso extremo se puede emplear la clasificación para hacer que sólo se necesite una parte de la memoria central en la tabulación. Esa parte se usa entonces para hacer la acumulación correspondiente a una celda del cuadro por vez. Este caso extremo sólo se menciona aquí para ilustrar la tesis siguiente: a mayor memoria, menos clasificación; a menor memoria, más clasificación.
- 3) Agregar un dispositivo de almacenamiento de acceso directo a la memoria central. El acceso a los registros almacenados en este dispositivo es directo, como en la memoria central, y no en serie, como sucede con las cintas magnéticas. Los discos y tambores magnéticos y hasta algunas huinchas o tarjetas magnéticas como la unidad de celdas de datos de la IBM o el sistema de memoria de libre acceso controlado por tarjetas de la NCR, son ejemplos de dispositivos de almacenamiento de acceso directo.

5. Antes de examinar con más detenimiento los tres métodos señalados, se señalarán aquí las proporciones que debe tener un censo típico. En

el documento de las Naciones Unidas Principios y recomendaciones relativos a los censos de población de 1970 se señalan 22 cuadros considerados de primera prioridad. El número total de ubicaciones o celdas en dichos cuadros es de 5 000, aproximadamente. En esta cuenta no se incluyen totales. Por ejemplo, se estima que para el cuadro 7 se necesitan 40 y no 63 celdas, que son las necesarias cuando se incluyen las distribuciones marginales (véase una explicación de las distribuciones marginales en el anexo). Muchos de los cuadros se referirán al nivel más bajo de la escala jerárquica de subdivisiones geográficas. Por lo tanto, es fácil ver que el número necesario de celdas de cuadros puede llegar al orden de los millones y aún a las decenas de millones, si se toma en cuenta la distribución geográfica. Para evitar en estos casos la clasificación y la repetición de pasadas, es indispensable usar algún dispositivo de almacenamiento de acceso directo. Por ejemplo, podría usarse este tipo de dispositivo cuando el material se reúne y prepara en un orden distinto al de la tabulación. Por lo general, estos dispositivos ofrecen una solución satisfactoria cuando el material de entrada se considera en su conjunto como una sola unidad de elaboración, y cuando no se puede terminar tabulación alguna en tanto no se prepare el último registro de entrada. Sin embargo, en muchos casos la forma en que se recolectan los resultados censales da por resultado una preclasificación útil. Es lo que sucede normalmente en los censos de población de facto. Los datos sobre una zona de empadronamiento pueden elaborarse separadamente de los de otras zonas en cuanto estén preparados, lo que se ha denominado procedimiento por lotes. Si se aceptan los cuadros recomendados que se mencionaron antes, sólo se necesitarán unas 5 000 celdas para efectuar una tabulación completa con una sola pasada. Si se hace una preclasificación adicional (por edades y sexos, digamos), la capacidad necesaria baja a algunos pocos cientos. (Las capacidades señaladas aquí sólo se refieren a los cuadros, y han de añadirse a las necesidades del propio programa y de un posible programa de supervisión.)

6. Los tres métodos, como se dijo antes, pueden combinarse. Cabe destacar asimismo que los dispositivos de almacenamiento de acceso directo pueden desempeñar un papel de mucha importancia, aunque su tamaño no les permita almacenar a la vez todos los cuadros necesarios.

7. El almacenamiento en dispositivos de acceso directo es más lento que el almacenamiento en la memoria central. Por lo tanto, en algunos casos sería más económico tabular los casos más frecuentes en la memoria central y los censos frecuentes en dispositivos de acceso directo. La computadora puede programarse de modo que indique dónde debe hacerse la tabulación para reducir al mínimo el tiempo de elaboración. Los datos de entrada a los que no se reserva espacio en la memoria central o en un dispositivo de almacenamiento de acceso directo (ya sea porque no se han agregado estos dispositivos o porque se ha sobrepasado su capacidad), pueden permanecer en cinta magnética para su futura elaboración. En algunos casos quizá convenga guardar datos excedentes en tarjetas perforadas. Esta técnica ofrece la posibilidad de reducir el volumen de datos de entrada entre una pasada y otra.

8. La tabulación de un registro de entrada consiste en determinar, en cada cuadro, si debe añadirse algún dato del registro a cualquiera de las celdas del cuadro. Si es preciso añadir datos al cuadro, el problema estriba en hallar los números de serie de celdas del cuadro que deben modificarse. Una vez encontrados esos números, se hacen las que adiciones. Como es natural, el número de serie que se calcule para un cuadro debe caer dentro de los límites del cuadro. Una manera de asegurarse de que así sea es reservar a cada cuadro un grupo residual en el cual podrían añadirse los registros con códigos erróneos. Sin embargo, esto crearía discrepancias caóticas entre los cuadros y no resolvería el problema cuando no estuviera claro si un registro pertenece o no a un cuadro determinado. El método que se aplica normalmente es el de facilitar la elaboración de los datos mediante un programa especial de comprobación. Esto obliga a una elaboración de exactitud tan grande que excede en mucho a las exigencias estadísticas.

C. Descubrimiento y corrección de errores

9. La precisión se logra a través del control de volumen y de la compaginación o verificación. Normalmente, el empadronador resume - en una lista de control - unos pocos datos básicos de los cuestionarios, como números de personas, hogares, locales de habitación, empresas industriales, predios agrícolas, etc. Al resumir las listas de control por subdivisiones geográficas se logran dos propósitos.

Primero, se obtiene rápidamente un reducido número de datos estadísticos provisionales. Segundo, se crea un marco para verificar la cabalidad y singularidad de los datos; esto es el control de volumen, que tiene importancia fundamental en todo el manejo de material y datos.

10. La compaginación o verificación tiene dos fases: el descubrimiento de errores y la corrección de errores. Parte de ella se hace y debe hacerse a mano. Sin embargo, la experiencia ha enseñado que la verificación manual no basta. Los registros deben aparecer libres de errores en la memoria central. Pueden deslizarse errores en la perforación, en la lectura de marcas o en la lectura óptica de caracteres. La computadora puede programarse para la primera fase solamente, o para ambas fases combinadas. En el primer caso, la computadora imprime una lista de los registros erróneos y señala la naturaleza de cada error. Este procedimiento puede combinarse con el rechazo de los registros de entrada erróneos; de hacerse así, las correcciones se introducirían a mano. En el segundo caso, la aparición de un error inicia un proceso automático de corrección, que puede ser determinista o estocástico. Es determinista el proceso en que una condición dada define un valor único. Por ejemplo, asignar el sexo femenino siempre que se han anotado varios hijos nacidos vivos es un proceso determinista. En los procedimientos estocásticos, una condición dada define una distribución de frecuencia de posibles valores para un código erróneo u omitido. El valor mismo se determina al azar. Por ejemplo, determinar una edad omitida eligiendo al azar un número con una distribución dada constituye un proceso estocástico.

11. Los sistemas de corrección, como es natural, deben estar concebidos de modo que eviten la introducción de nuevos errores o inconsistencias. Una manera simple, pero muy discutible, de formular las normas estocásticas, es suponer que los datos omitidos o erróneos están distribuidos en la misma forma que los datos correctos. A veces los errores aparecen en forma muy sistemática. Puede suceder que un codificador haya interpretado siempre mal determinadas respuestas (tal vez a causa de diferencias lingüísticas) o que haya memorizado incorrectamente ciertos códigos. Es posible que personas que profesan una religión sean más reacias a declararla que personas que profesan otras religiones. Es difícil formular reglas para corregir automáticamente, pero, una vez formuladas, los beneficios son considerables. Se puede eliminar la verificación de la codificación y de la perforación, o hacerla sólo por muestreo.

12. Las correcciones automáticas deben controlarse constantemente desde el punto de vista estadístico. El programa de compaginación debe informar del número de errores por tipos, por codificadores y por perforadores en cada lote, y de vez en cuando deben comprobarse los requisitos para aplicar las normas estocásticas.

13. No todos los errores pueden corregirse automáticamente: los que resultan de traspapelar cuestionarios, por ejemplo.

14. La compaginación no siempre se limita a cada registro en sí, sus códigos y las relaciones entre sus códigos. También puede abarcar las relaciones entre registros del mismo grupo, como los correspondientes a todos los miembros de un hogar. La labor de crítica se combina a veces con la preparación de registros resumidos, y también con la transferencia de datos desde los registros principales o el traspaso de datos paralelos de censos diferentes.

15. La compaginación es un requisito previo para el funcionamiento del programa de tabulación. Esto no significa que la verificación anterior a la tabulación sea suficiente para que los cuadros resulten aceptables, ni significa tampoco que toda la verificación deba hacerse antes de comenzar la tabulación, esto último por razones de índole práctica. Los dos ejemplos siguientes ilustrarán estos dos puntos.

16. Ejemplo 1. En un censo de población se preguntó acerca del tipo de diploma de formación vocacional. Se asignó el código de "Otros diplomas" a la respuesta "Licencia para conducir". En cada registro en sí, el código correspondiente a "Otros diplomas" era aceptable. En la tabulación, la frecuencia con que aparecía este código suscitó dudas e hizo que se descubriera el error.

17. El ejemplo 1 muestra que el estudio de las proporciones que ofrece un cuadro suele revelar errores que no pueden descubrirse, o no se han descubiertos, en los registros. La computadora puede programarse para analizar los cuadros y para indicar los casos que merecen investigarse.

18. Ejemplo 2. Puede suceder que, en un censo de población, el plan de tabulación incluya un cuadro sobre el país en su conjunto, con una clasificación cruzada completa de la industria por ocupaciones. Tal vez la capacidad sea suficiente para comprobar por separado los códigos industriales y ocupacionales en el plano de los lotes, pero insuficiente para hacer una comprobación práctica y económica de la combinación de ambas clases de códigos en igual plano. Si se excluye la clasificación cruzada de la verificación y tabulación originales, la elaboración por lotes puede hacerse con bastante exactitud. Una vez terminado este proceso, se pueden seleccionar las clasificaciones cruzadas y se pueden comprobar cotejándolas con un archivo de combinaciones aceptables. Después de hacer las correcciones necesarias, puede prepararse el cuadro nacional.

19. El ejemplo 2 plantea la cuestión de si los errores descubiertos en la segunda labor de compaginación afectarán a los cuadros que se hicieron directamente, después de la primera compaginación. Si se ha terminado la tabulación, no son ya aplicables las rigurosas exigencias derivadas del uso de la computadora y el problema queda reducido a determinar el significado estadístico de los errores; siempre que sea razonable suponer que cualquier decisión basada en los cuadros podría ser influida por una nueva serie de correcciones, será preciso efectuar esas correcciones, pero, si no es así, no habrá que hacerlas.

20. Es posible efectuar las correcciones de manera sistemática dando una pasada al programa de tabulación, en su totalidad, e insertando los registros de entrada erróneos como componentes negativos y sus posibles sustituciones

/como componentes

como componentes positivos. Los cuadros obtenidos, que contendrán únicamente las diferencias entre los dos componentes, se usan entonces para poner al día el conjunto de los cuadros. De este modo no es preciso volver a pasar todos los registros de entrada para hacer las correcciones.

21. Los cuadros básicos, esto es, los originados directamente a partir de los registros de entrada para cada caso particular, se añaden a su vez a los nuevos cuadros en una o más jerarquías. La compaginación de los cuadros no se limita a la serie básica, sino que ha de llevarse a cabo después de cada etapa aditiva. La compaginación no consiste únicamente en descubrir los errores existentes, sino que, en realidad, constituye un análisis y una interpretación preliminares del material estadístico.

D. Obtención mediante computadora de los cuadros definitivos para su reproducción mediante foto-offset

22. Como bien se sabe, las computadoras pueden utilizarse para imprimir cuadros con títulos, talones, números de página, etc., en forma tal que puedan ser impresos por el sistema offset. Algunas calculadoras poseen juegos de caracteres para la impresión simultánea en diferentes alfabetos, como romano-griego, romano-árabe, romano-hebreo, etc., con lo cual pueden imprimirse en una pasada cuadros en dos alfabetos.

23. Quizás menos conocido sea el método de foto composición electrónica. En uno de estos sistemas, que se utiliza actualmente, se programa la computadora para que haga pasar un rayo luminoso a través de la imagen del carácter tipográfico que se desee reproducir, el cual se proyecta sobre película fotográfica. Otro sistema, que está actualmente en período experimental, consiste en programar a la computadora para que produzca electrónicamente una página de texto en la pantalla de un tubo de televisión, la cual es fotografiada, utilizándose esta fotografía, como en el sistema anterior, para la reproducción en offset. Se dispone de muchas familias de tipos de imprenta y la impresión apenas se diferencia de la habitual. La elaboración automática de los cuadros tiene las siguientes ventajas en relación con los procedimientos corrientes: alta velocidad, bajo costo (al menos en los países desarrollados) y mayor exactitud. Según se ha indicado anteriormente, la compaginación constituye un requisito previo para el programa de tabulación y es también un prerrequisito para la obtención automática de los cuadros.

24. Por regla general no se reproducen totalmente los resultados compaginados del programa de tabulación, lo que se debe a varias razones. Para facilitar la elaboración de los datos se hacen a veces cuadros correspondientes a subdivisiones geográficas mínimas, que sólo adquieren significado en un nivel más alto de la jerarquía. En otros casos, se hallan los errores, pero resulta imposible corregirlos, por lo que han de desecharse los cuadros correspondientes. En algunos cuadros pueden figurar todas las categorías posibles, con inclusión de aquellas que carecen de datos numéricos y que, por tanto, no han de aparecer en los cuadros impresos. La elección de los cuadros que han de imprimirse y la decisión relativa a sus modificaciones, se determinan, hasta cierto punto, por los resultados de la tabulación. No es posible en consecuencia, su total planeación previa.

E. Resumen

25. La tabulación constituye la operación central de la elaboración de los datos del censo. Cuando se emplea la computadora para tabular, equivale a una calculadora rapidísima que posea un elevado número de registros. Las computadoras pueden verificar en gran medida la codificación y realizar tareas de perforación de tarjetas; son igualmente capaces de leer, no sólo tarjetas perforadas o cintas de papel, sino también marcas y caracteres (escritos a máquina o a mano); además, pueden realizar la primera fase de la compaginación - el descubrimiento de errores - y, hasta cierto punto, la segunda, esto es, la corrección. A cada etapa de la elaboración de datos sigue una etapa de compaginación, en la cual se analizan los resultados obtenidos en la etapa anterior y se despeja el camino para la siguiente. Existe además un control de volumen, mediante el cual se logra que ningún material reciba por error más o menos pasadas de las necesarias. Las computadoras pueden también ser programadas a fin de preparar los originales de los cuadros para su reproducción por offset.

26. El grado de automatización que debe alcanzar cada país ha de determinarse teniendo en cuenta las circunstancias de cada caso. En términos generales, las computadoras han de tener la capacidad suficiente para mantener el ritmo de la codificación y de las demás etapas preparatorias de los materiales de entrada. La computadora debe mantener bajo control permanente a las operaciones de oficina y al sistema de codificación,

/mediante una

mediante una especie de "sistema de pronta alarma". Los datos reunidos pueden mostrar características insospechadas que exijan la modificación del sistema codificador. Todo ello obliga a que, tanto la computadora como los programas, estén en buenas condiciones de funcionamiento al levantar el censo. Esto significa, de acuerdo con la experiencia, que los preparativos para la elaboración deben iniciarse entre 18 y 12 meses antes del día de levantamiento del censo. Si se levanta un censo de prueba, es preferible hacerlo durante la fase de elaboración, en cuya virtud, el día de referencia para iniciar los preparativos deberá coincidir aproximadamente con el día de levantamiento del censo de prueba. Es de la máxima importancia disponer de bastante tiempo para el establecimiento del sistema de elaboración. Desde que se comience la planeación censal, y a lo largo de todo este período, ha de concederse la necesaria atención a las tareas de elaboración.

/Anexo

Anexo

Un ejemplo de tabulación mediante computadora

En el presente ejemplo se dispone en la memoria central de la computadora de 400 palabras dispuestas para la tabulación (ubicaciones, zonas y celdas), numeradas del 4200 al 4599, cada una de las cuales puede actuar como un contador. El sector disponible de la memoria central se utilizaría en el caso presente para originar los cuadros siguientes:

Cuadro 1: Población por año de edad y sexo

Cuadro 2: Población por estado civil, edad y sexo

Según se observa más adelante, hay 102 códigos de edad, 2 códigos de sexo y 5 códigos de estado civil. En consecuencia, se precisan 204 palabras para el cuadro 1 y se reservan para el mismo las palabras números 4200 a 4403. Para el cuadro 2, las edades forman 15 grupos, por lo que se necesitan 150 palabras para las cuales se reservan los números 4404 a 4553. Obsérvese que se han reservado palabras únicamente para las partes básicas de los cuadros y no para la distribución llamada marginal o total. En la versión del cuadro 2 que figura a continuación se aclara este concepto. Las zonas sombreadas representan distribuciones marginales.

Grupo de edad	Hombres					Mujeres					Total									
	Estado civil					Estado civil					Estado civil									
	1	2	3	4	9	TOTAL	1	2	3	4	9	TOTAL	1	2	3	4	9	TOTAL		
Menos de 15 años																				
de 15 a 19 años																				
de 20 a 24 años																				
....																				
75 y más años																				
No consta																				
TOTAL																				

/No se

No se utilizan las palabras que van del número 4554 al 4599. Es necesario añadir un "uno" a la palabra correspondiente en la zona reservada al cuadro 1 y a la palabra correspondiente en la zona reservada al cuadro 2, para cada registro de entrada. El número de la palabra correspondiente, que es la dirección, es una sencilla función de los códigos de sexo, edad y estado civil, aunque no lo es de la versión externa de los códigos, sino de la interna. La versión externa se utiliza para codificar los cuestionarios y perforar las tarjetas, en tanto que la interna es calculada por la computadora. Según se muestra a continuación, ambas versiones pueden ser idénticas o la versión interna puede variar de un cuadro a otro, en cuyos casos las funciones son las siguientes:

$$\begin{aligned} \text{Dirección para el cuadro 1} &= (\text{sexo}-1). 102 + (\text{edad}-1) + 4200 \\ \text{Dirección para el cuadro 2} &= (\text{sexo}-1). 75 + (\text{edad}-1).5 \\ &\quad (\text{estado civil}-1) + 4404 \end{aligned}$$

Las versiones diferentes de los códigos son las siguientes:

<u>Código externo para la edad</u>	<u>Código interno para la edad. Cuadro 1</u>	<u>Código interno para la edad. Cuadro 2</u>
00 para menos de 1 año	001 para menos de 1 año	01 para menos de 15 años
01-99 para 01 a 99 años	002-100 para 01 a 99 años	02 para 15 a 19 años
XK para 100 y más años	101 para 100 y más años	03 para 20 a 24 años
YY cuando no consta la edad	102 cuando no consta la edad
		14 para 75 y más años
		15 cuando no consta la edad

Código externo para el sexo = Código interno para el sexo

- 1 hombre
- 2 mujeres

/Código externo

Código externo para el estado civil

- 1 para solteros
- 2 para casados
- 3 para viudos
- 4 para divorciados
- 5 cuando no consta

Código interno para el estado civil

Cuadro 2

- 1 para solteros
- 2 para casados
- 3 para viudos
- 4 para divorciados
- 5 cuando no consta

Ejemplo de cómputo de direcciones:

Códigos externos

Direcciones y códigos internos

	<u>Códigos externos</u>			<u>Cuadro 1</u>		<u>Cuadro 2</u>				
	Edad	Sexo	Estado civil	Edad	Sexo	Di- rec- ción	Edad	Sexo	Estado civil	Di- rec- ción
Registro 1	21	1	2	22	1	4221	3	1	2	4415
Registro 2	YY	2	9	102	2	4403	15	2	5	4554
Registro 3	75	2	4	76	2	4377	14	2	4	4547

En la página siguiente se describe la parte disponible de la memoria central. Cuando, al tabular los tres registros indicados, hay que añadir un "uno" a las palabras, se expresa esta operación mediante el símbolo "más 1".

