



NACIONES UNIDAS

CONSEJO  
ECONOMICO  
Y SOCIAL



LIMITADO

C.1

ST/ECLA/Conf.43/L.4  
11 de julio de 1972

ORIGINAL: INGLES

COMISION ECONOMICA PARA AMERICA LATINA

SEMINARIO SOBRE LA PREPARACION Y USO DE TABULACIONES  
DE CENSOS DE POBLACION Y HABITACION

Organizado por las Naciones Unidas, por intermedio de  
la Comisión Económica para América Latina, la Oficina  
de Estadística de las Naciones Unidas, el Fondo de las  
Naciones Unidas para Actividades en Materia de Población  
con la colaboración del Centro Latinoamericano de Demografía

Santiago de Chile, 14 al 19 de agosto de 1972

REVISION Y CORRECCION POR COMPUTADORA DE UN  
CENSO DE POBLACION Y HABITACION

Por Howard G. Brunsmann  
de la Oficina del Censo de los Estados Unidos



El programa de detección y corrección de errores tiene por objeto mejorar la exactitud de los datos censales. Como se indica en el documento correspondiente de las Naciones Unidas 1/, los errores pueden derivar del hecho de que la información contenida en el cuestionario censal no represente correctamente los hechos o de errores cometidos en las diversas etapas del procesamiento de los datos, desde la boleta censal hasta las tabulaciones estadísticas resultantes.

El documento de las Naciones Unidas se refiere a tres tipos de errores que pueden detectarse. Ellos son: a) omisión de datos; b) datos inadmisibles; y c) datos no consistentes.

Los tres tipos podrían ocurrir tanto en los datos registrados en la boleta censal como en el procesamiento de los datos. En la boleta podría omitirse el dato porque el declarante no supo qué responder. Los datos registrados podrían ser inadmisibles: por ejemplo, pudiera no existir el lugar de nacimiento declarado. O los datos registrados podrían ser inconsistentes con otras características. Por ejemplo, el último grado de enseñanza cursado podría ser demasiado alto conforme a la edad de la persona. Las instalaciones de baño declaradas pueden no guardar relación con el suministro de agua. Algunos errores pueden corregirse con un examen detallado de la boleta. Si se ha omitido el sexo del declarante, podría suplirse esa información con un examen de su nombre de pila. Un lugar de nacimiento inadmisibles puede resultar de una pequeña falta de ortografía. Muchos de esos errores sólo pueden corregirse volviendo a interrogar al declarante. Evidentemente este procedimiento no es práctico para aplicarlo en forma masiva a un censo de todo el país.

Muchos de los errores resultan de las diversas operaciones de procesamiento. Los más comunes derivan de una codificación manual incorrecta y del proceso de conversión de los datos a una representación que pueda utilizar la máquina. Los errores de este tipo pueden corregirse volviendo a examinar la boleta censal, pero es muy oneroso buscar la ficha correspondiente y determinar qué corrección debe hacerse. Y si acaso el error radica en el registro de los datos en la boleta, no se gana nada con esa búsqueda. Así también, si los errores son relativamente escasos, no se distorsiona gravemente la exactitud de los datos cuando esos errores se corrigen por imputación.

---

1/ Naciones Unidas, The use of computers for error detection and correction in population and housing census, abril de 1971.

La falta de coherencia entre los datos puede advertirse en un examen caso por caso de cada registro o con la preparación de tabulaciones cruzadas muy detalladas de las variables. Por ejemplo, la falta de coherencia entre la asistencia escolar y la edad podría descubrirse con una tabulación detallada de grado de enseñanza por edad, también detallada. Entonces podrían introducirse los ajustes en el cuadro antes de su publicación. Pero esos ajustes podrían provocar incoherencias con los cuadros que muestran el grado de enseñanza cursado, pero no por edad, para las zonas más pequeñas. Por lo tanto, es preferible descubrir los casos de inconsistencias y corregirlas antes de preparar tabulación alguna.

Como se indica en el documento de las Naciones Unidas, el papel de la computadora puede limitarse a la detección de registros con error, sin la corrección automática de esos errores. Cuando se usaba el equipo de registro unitario, la tabuladora podía procesar un grupo de tarjetas perforadas y separar en un casillero aparte todas las que contenían un tipo específico de omisión o incongruencia. Estas tarjetas se corregían a mano y se volvían a colocar en el archivo. De la misma manera, el programa para la computadora puede imprimir un informe que muestre el contenido de todas las tarjetas defectuosas y la naturaleza del defecto. El programa puede producir duplicados de las tarjetas defectuosas de modo que sólo los renglones defectuosos necesiten corregirse.

El hecho de que la computadora sea capaz de advertir ciertos tipos de errores, no debe usarse como disculpa para eliminar todas las verificaciones. La información registrada puede ser incorrecta aunque sea plenamente coherente. El programa para la detección de errores podría emplearse para identificar las unidades de trabajo que arrojan un número excesivo de errores. La presencia de los errores advertidos por la computadora podrá sugerir que hay otros. Las unidades de trabajo que tienen demasiados errores podrían entonces someterse a una verificación más completa.

En muchos casos la verdadera naturaleza del error no salta a la vista. Una codificación por sexo inadmisibles puede resultar cuando la perforista pasa por alto un ítem. En ese caso, el código de relación con el jefe del hogar "4" puede aparecer en la columna destinada a sexo donde sólo son aceptables los códigos "1" y "2". Un error de este tipo hará aparecer datos incorrectos para muchas de las características. Es mejor reprocesar tales registros antes que asignar los datos por imputación. Así también un código equivocado de identificación de la unidad de vivienda en la tarjeta de una persona puede llevar a un error aparente en el recuento de personas en la unidad de habitación, y pueden aparecer varios grupos de personas con la misma identificación pero sin unidad de habitación. Estos casos también deben corregirse por procesamiento y no por imputación.

/La corrección

La corrección automática de errores puede consistir en dar el mismo tratamiento a los 3 tipos de errores posibles: datos inadmisibles, datos admisibles pero que producen incoherencias con otras variables, y omisión de datos. El procedimiento es muy aplicable a las características que tienen muchas categorías, como la ocupación, o industria, o país de nacimiento. La inclusión de una categoría de "no declarado" obliga al usuario a usar cierto criterio en la interpretación de los datos. Por ejemplo, en algunos grupos de edad, el número de personas que no declara su asistencia escolar, pudiera ser mayor que el número que no asistió a la escuela. La interpretación de tales datos sería bien diferente si se supone que el grupo que no declaró esa información no asiste a la escuela, o se distribuye en la misma proporción que los que declaran la asistencia. Por lo tanto, la inclusión de la categoría "no declarado" advierte al usuario que debe emplear los datos con precaución. Tal advertencia tal vez no existiría si el programa de la computadora (o más bien dicho el programador) hubiera decidido cómo clasificar esos datos.

Para algunas características es preferible que el programa asigne un código a todos los casos en que se advierten errores. El programa puede asignar valores a las variables que faltan tomando en cuenta otras características del registro.

Es conveniente que el programa asigne un código para la clasificación por sexo en todos los casos a fin de no aumentar los casilleros en cada cuadro en un tercio, como ocurriría si se incluyera una categoría de "no declarado" junto con total, varones y mujeres. El programa puede tomar en cuenta si el declarante es la esposa, o el jefe del hogar con esposa presente; también puede considerar la fecundidad declarada, o si la actividad declarada es la de ama de casa. Aun cuando no haya estos indicios, es preferible imputar el sexo tomando en cuenta la relación con el jefe del hogar y el estado civil.

La imputación del dato de edad permite evitar la difícil decisión de incluir o no a las personas cuya edad no ha sido declarada en los cuadros que se refieren solamente a la población de 15 años y más, de 5 años y más, de 5 a 29 años, etc.

El programa también puede imputar una edad que falta sobre la base de las características de la persona o la presencia y características de personas emparentadas. Los niños no declaran actividad económica. La asistencia escolar y el nivel de instrucción no se declaran para los menores

de cinco años. Suele haber una relación relativamente estrecha entre la edad de la esposa del jefe del hogar y la de éste, y entre la edad de la esposa del jefe del hogar y la del hijo mayor. Hay también una estrecha relación entre la edad y el nivel de instrucción para los que asisten a la escuela. De la misma manera, es mejor imputar un valor al estado civil y al parentesco cuando no se declaran.

La asistencia a la escuela debiera abarcar a las personas de 6 a 29 años. El programa podría tomar en cuenta la escolaridad y la edad de las personas que no declaran estar asistiendo. El grupo que no declara podría concentrarse entre las personas de 25 a 29 años, que raras veces asisten a un establecimiento educacional. Es así como los valores imputados por la computadora pueden ser más racionales que los que suponga el usuario.

Es mucho más probable que los analistas acepten las imputaciones de la computadora si cuentan con un informe de las características que han sido modificadas y cómo han sido modificadas. Este informe puede ser un recorte resumido. También puede incluir informes de todos los casos en que se han introducido cambios. El analista puede entonces revisar el informe y presentar enmiendas de aquellos en que no considera atinada la acción de la computadora.

En el último año he estado trabajando con la Oficina Ejecutiva del Censo de Nicaragua preparando un programa de corrección por computadora de su Censo de Población y Habitación de 1971. El programa se está trabajando en una IBM 360/25 con 32 K.

La entrada del programa consiste en tarjetas perforadas con una tarjeta para cada unidad de habitación y una tarjeta por persona. La tarjeta de la unidad de habitación va primero y luego vienen las tarjetas de las personas que viven en esa ciudad. Las personas se agrupan por hogares dentro de la unidad de habitación.

La salida está constituida por una cinta binaria en que cada variable se expresa en un número binario de uno o dos bytes. La salida también contiene una indicación de cuáles variables fueron cambiadas en el proceso de corrección. Es así posible preparar tabulaciones que den el número de casos en que la variable ha sido modificada y la distribución de las variables en los casos en que no ha habido alteración, junto con la distribución de los casos que dio el proceso de corrección.

/El programa

El programa imprime un informe para cada unidad de trabajo que muestra:

- 1) El número total de registros
- 2) Número de registros con uno o más errores, y como porcentajes del total de registros
- 3) Total de errores y como porcentaje del total de registros
- 4) Total de registros de habitación
- 5) Total de registros de habitación con uno o más errores
- 6) Total de errores de habitación y distribución de los errores de habitación por tipo
- 7) Iguales cifras y distribuciones para los registros de población

Además del reporte por unidades de trabajo, el programa puede producir un informe sobre el contenido de las tarjetas de entrada antes de la corrección, e indicar qué variables han cambiado y cómo han cambiado. Este informe, basado en cada uno de los registros puede prepararse para todas las tarjetas, para las tarjetas de las unidades de habitación y para sus ocupantes, cuando ha habido imputaciones para el grupo. O puede omitir este informe de las tarjetas por separado.

El informe de todas las tarjetas puede verificarse manualmente. En la mayoría de los casos ha resultado satisfactoria la revisión por computadora. Cuando no es satisfactoria puede prepararse un juego complementario de tarjetas para las unidades de habitación y sus ocupantes. Estas tarjetas pueden reemplazar a las defectuosas.

Por desgracia no hemos podido elaborar un programa de corrección actuado por tarjetas-parametros como las que tiene el programa CENTS. Hemos tomado en cuenta la conveniencia de facilitar el proceso de adaptación del programa a otros países. Hemos usado fundamentalmente subrutinas e instrucciones. Las variables están catalogadas con títulos mnemónicos. La edad se titula AGE (¿o debiera ser EDAD?). Sólo se requieren pequeños cambios para tomar en cuenta que la variable pudiera estar en un lugar distinto en el registro de entrada o salida. El Computer Methods Laboratory de la Oficina del Censo de Estados Unidos podría adaptar el programa al censo de otro país en una pequeña fracción del tiempo que demoró preparar el programa de Nicaragua.

/En el

En el proceso de corrección, primero se lee la tarjeta de la unidad de habitación y de los miembros del primer hogar de esa unidad. Entonces se procesa este primer hogar. Luego se leen las tarjetas del próximo hogar y se procesan. La tarjeta de la unidad de habitación se edita cuando se llega a una nueva unidad. Entonces se transcriben a cintas las tarjetas de la unidad de habitación y de todos sus ocupantes. Por este método es posible verificar la coherencia entre las características de personas emparentadas y usar esas relaciones entre personas en el procedimiento Hot Deck. Por ejemplo, cuando está presente la esposa del jefe del hogar, se supone que la edad de ese jefe, de no estar declarada, guarda la misma diferencia con la de su esposa que la que hubo entre el jefe de hogar anterior y su esposa. A veces este método lleva a resultados absurdos y debe ser descartado. Por ejemplo, supóngase que luego de un hogar en que el jefe tenía 20 años y la esposa 50, sigue otro en que el jefe es de edad desconocida, pero cuya esposa tiene 25 años. Como el primer jefe de hogar tiene 20 años y 30 años menos que su esposa, pero el segundo, no puede ser 30 años menor que su esposa de 25.

Me gustaría señalar algunos de los métodos de correcciones automáticas que efectúa el programa:

#### Sexo

El sexo se obtiene por imputación, cuando no se declara, empleando los criterios siguientes: Se imputa como femenino si declara haber tenido uno o más hijos o ser esposa del jefe de hogar. Se imputa como masculino si su esposa está presente en el hogar. En cualquier otro caso, se imputa el sexo al jefe de hogar empleando el procedimiento Hot Deck, y tomando en cuenta la presencia de hijos propios en el hogar y el estado civil, de no haber hijos presentes. Impútese el sexo al hijo del jefe del hogar usando el Hot Deck. Cuando no se declara el sexo de un hijo, es probable que no se declare el de otros del mismo hogar. Por lo tanto el Hot Deck para este ítem contiene el sexo de los últimos cinco niños. Se da una rotación a estos valores a fin de evitar repeticiones demasiado altas del mismo sexo. En los demás casos impútese el sexo del Hot Deck tomando en cuenta la relación con el jefe del hogar, la edad y el estado civil.

#### Edad

La revisión y corrección de la edad es una de las tareas más complejas del programa. Se anula una declaración de menos de 10 años para el jefe del hogar o esposa del jefe. De la misma manera un hijo del jefe del hogar sólo puede ser de 15 a 49 años más joven que el jefe, y el padre de éste debe ser por lo menos 15 años mayor.

/Impútese la

Impútese la edad a la esposa a partir de la edad del jefe del hogar, según la diferencia de años entre la pareja del hogar anterior. Si esto no da un resultado valedero, impútese la edad a la esposa a partir de la edad del hijo mayor, según la diferencia entre el hijo mayor y la esposa del hogar anterior. Si esto no da resultado, impútese la misma edad que la de la esposa anterior.

Impútese la edad del jefe del hogar a partir de la diferencia de edad entre el jefe de hogar anterior y su esposa. Si no hay esposa, impútese la edad a partir de la edad del hijo mayor y de la diferencia entre el jefe de hogar anterior y su hijo mayor. Si no hay hijos, impútese la misma edad del jefe de hogar anterior del mismo sexo y estado civil.

Para otras personas que asisten a la escuela, impútese la edad a partir del grado que cursan y de la diferencia entre edad y curso de la persona anterior que asiste a la escuela.

Para los hijos del jefe del hogar que no asisten a la escuela, impútese la edad a partir de la edad de la esposa (o jefe de hogar por sexo si no hay esposa) y de la diferencia de edad entre el niño y la esposa o jefe de hogar anterior. Como en el caso del sexo del hijo, estas diferencias se mantienen y se procede a su rotación cada cinco casos para evitar repeticiones.

Para un padre o nieto, impútese la edad a partir de la edad del jefe del hogar y de la diferencia entre el padre o nieto anterior y el jefe de hogar, por relación. En todos los demás casos y cuando fallen los procedimientos señalados y no den una edad razonable, impútese el dato según las características de la persona anterior con la misma relación y edad probable.

#### Alfabetismo

Si no se declara, impútese como alfabeto si ha cursado tres o más años de enseñanza; como analfabeto en caso contrario.

Los temas incluidos en la corrección de los datos de habitación incluyen los siguientes:

#### Estado de ocupación

Cámbiese el estado de ocupación según que haya (o no) registros de población para esa unidad.

/Personas que

Personas que ocupan la unidad de habitación

Quando haya discrepancias, cambie el número de personas de la unidad conforme al número de registros de población.

Cuartos y dormitorios

Aplíquese el procedimiento Hot Deck para imputar el número de cuartos a partir de la unidad anterior del mismo tipo, cuando no se declaran ni el número de cuartos ni el número de dormitorios.

Calcúlese el número de dormitorios por el procedimiento Hot Deck a partir del número de cuartos y diferencia entre éste y el de dormitorios de una unidad anterior.

Servicios sanitarios

Aplíquese el procedimiento Hot Deck para imputar la provisión de agua corriente por tipo de estructura, cuando no se declara información alguna sobre servicios sanitarios.

Aplíquese el procedimiento Hot Deck para imputar la provisión de agua corriente, según la unidad anterior que tuviera retrete o baño de uso exclusivo, retrete o baño compartido, o ninguno, cuando este ítem no haya sido contestado o la respuesta sea incoherente.

Aplíquese el procedimiento Hot Deck para imputar los servicios de retrete y baño, a partir de la unidad anterior, según la provisión de agua corriente.



