

NOTAS DE POBLACIÓN

AÑO XXIX, N° 75, SANTIAGO DE CHILE



NACIONES UNIDAS



Comisión Económica para América Latina y el Caribe
Centro Latinoamericano y Caribeño de Demografía (CELADE) – División de Población

LC/G. 2186-P
Diciembre de 2002

Copyright © Naciones Unidas 2003
Todos los derechos están reservados
Impreso en Chile

La autorización para reproducir total o parcialmente esta obra debe solicitarse al Secretario de la Junta de Publicaciones. Sede de las Naciones Unidas, N.Y.10017, EE.UU. Los Estados miembros y sus instituciones gubernamentales pueden reproducir esta obra sin autorización previa. Sólo se les solicita que mencionen la fuente e informen a las Naciones Unidas de tal reproducción.

PUBLICACIÓN DE LAS NACIONES UNIDAS

NÚMERO DE VENTA: S.03.II.G.77

ISBN 92-1-322063-4
ISSN impreso 0303-1829
ISSN electrónico 1681-0333

Ilustración de portada: Curdo Bermúdez, "El balcón" (detalle), 1941.
Gentileza del Banco Interamericano de Desarrollo
Diseño de portada: María Eugenia Urzúa

COMISIÓN ECONÓMICA PARA AMÉRICA LATINA Y EL CARIBE
José Antonio Ocampo Secretario Ejecutivo

CENTRO LATINOAMERICANO Y CARIBEÑO DE DEMOGRAFIA
(CELADE) – DIVISIÓN DE POBLACIÓN
Miguel Villa, Oficial a cargo

La Revista **NOTAS DE POBLACIÓN** es una publicación del Centro Latinoamericano y Caribeño de Demografía (CELADE) - División de Población, cuyo propósito principal es la difusión de investigaciones y estudios de población sobre América Latina y el Caribe, aun cuando recibe con particular interés artículos de especialistas de fuera de la región y, en algunos casos, contribuciones que se refieren a otras regiones del mundo. Se publica dos veces al año (junio y diciembre), con una orientación interdisciplinaria, por lo que acoge tanto artículos sobre demografía propiamente tal, como otros que aborden las relaciones entre las tendencias demográficas y los fenómenos económicos, sociales y biológicos.

Comité editorial:

Jorge Bravo
Juan Chackiel
José Miguel Guzmán
Rolando Sánchez
Susana Schkolnik

Coordinador Técnico:

Juan Enrique Pemjean

Secretaria:

María Teresa Donoso

Redacción y administración:

Casilla 179-D, Santiago, Chile
E-mail: mdonoso@eclac.cl

Precio del ejemplar: US\$ 12

Suscripción anual: US\$ 20

Las opiniones expresadas en esta revista son responsabilidad de los autores, sin que el CELADE sea necesariamente participe de ellas.

SUMARIO

Presentación	7
Foreword	10
Avant-propos	13
América Latina: los censos del 2000 y el desarrollo social. <i>José Miguel Guzmán y Susana Schkolnik</i>	17
Los censos en América Latina: nuevos enfoques. <i>Juan Chackiel</i>	45
Cómo adaptar las nuevas tecnologías a las operaciones censales. <i>Arij Dekker</i>	73
Algunas reflexiones acerca de un formulario censal diseñado para la captura de datos mediante tecnología de imágenes (escáner). <i>Nelly T. Niedworok</i>	115
Los censos de población en Internet. <i>Dr. Werner Haug</i>	133
El censo de Canadá de 2006. Dirección estratégica. <i>Anil Arora</i> ..	147
¿Podrán los censos de los países en desarrollo responder a las futuras demandas de información?: el caso de Argentina. <i>Alejandro Giusti</i>	165
Preguntas para investigar la fecundidad y la mortalidad en la niñez y la mortalidad de adultos. <i>Ralph Hakkert</i>	195
La población indígena en Costa Rica según el censo del año 2000. <i>Elizabeth Solano Salazar</i>	217

**ALGUNAS REFLEXIONES ACERCA DE UN FORMULARIO
CENSAL DISEÑADO PARA LA CAPTURA DE DATOS
MEDIANTE TECNOLOGÍA DE IMÁGENES (ESCÁNER)**

Nelly T. Niedworok
Instituto Nacional de Estadística (INE)
Uruguay

RESUMEN

El Censo de Población, Hogares y Viviendas del año 1996 de Uruguay incorporó simultáneamente un conjunto de innovaciones tecnológicas, manteniendo un alto control sobre la calidad de los datos y los costos. Ello cambió radicalmente la organización del trabajo de elaboración de los datos posterior al relevamiento, destacando la importancia de la capacidad institucional para la implementación del proceso con un alto nivel de coordinación.

La automatización casi completa de la captura de datos (mediante tecnología de imágenes), la codificación de textos y la depuración de los datos resultantes, permitió —entre otros aspectos— un control de calidad más acucioso de todos los procesos, asegurando la uniformidad de criterios y la obtención de información detallada sobre los niveles de error obtenidos dentro de los rangos definidos.

El análisis pormenorizado del “mapa” de los diferentes niveles de error de la captura de datos revela la existencia de múltiples factores de importancia que deben considerarse en el diseño de un cuestionario censal.

Los avances tecnológicos producidos con posterioridad y los resultados globalmente positivos de una experiencia pionera —como la relatada— permiten predecir avances significativos en la calidad y eficiencia del procesamiento de los datos censales del futuro.

ABSTRACT

In the elaboration of the Housing and Population Census of Uruguay in the year 1996 a set of technological innovations were used, maintaining a strict control of the data quality and costs. Those innovations changed radically the organization of the data processing operation, enhancing the importance of the institutional power to secure a high coordination level in its implementation.

The comprehensive automation of the data capture (using imaging techniques), text coding and data editing allowed a more stringent quality control of every process, assuring coherence and detailed information about error levels.

The analysis of the “map” of error levels over each question shows the importance of several factors, which must be taken care of in the design of the questionnaire.

Ulterior technological advances and the positive evaluation of this first experience allow predicting significant advances in the quality and efficiency of future census data processing.

RÉSUMÉ

Le recensement de la population, des ménages et du logement de 1996 de l'Uruguay a introduit une série d'innovations technologiques, tout en conservant un strict contrôle de la qualité des données et des coûts. Cette approche a profondément modifié l'organisation du travail d'élaboration des données après le dénombrement et a mis en exergue l'importance de la capacité institutionnelle pour mettre en oeuvre le processus de façon parfaitement coordonnée.

L'automatisation presque totale de la capture de données (moyennant la technologie d'images, la codification des texte et l'épuration des données extraites) a notamment permis un contrôle de qualité plus minutieux de tous les processus, ce qui garantit une uniformité des critères et l'obtention d'une information détaillée sur les niveaux d'erreur obtenus dans les marges définies.

L'analyse détaillée de la « carte » des différents niveaux d'erreur de la capture des données met en relief l'existence de nombreux facteurs importants dont il doit être tenu compte dans la conception d'un questionnaire censitaire.

Les progrès technologiques survenus ultérieurement et les résultats globalement positifs d'une expérience pionnière comme celle présentée ici permettent d'augurer un bond significatif en termes de qualité et d'efficience du traitement futur des données censitaires.

INTRODUCCIÓN

En Uruguay, en los censos de población, hogares y viviendas de las últimas décadas y hasta el de 1985 se aplicaron, en forma sistemática, métodos y procedimientos tradicionales de elaboración, lo que suponía arduas tareas de crítica y codificación manual de los datos y el ingreso de éstos a medios magnéticos mediante digitación también manual. Con el fin de acortar los plazos de entrega de los datos al público, se procesaban en forma anticipada muestras de cuestionarios censales, lo que permitía una presentación de cuadros básicos del censo transcurrido aproximadamente un año del relevamiento. Los datos definitivos del censo de 1985 fueron entregados durante el año 1989.

Para la realización del censo de 1996, con el fin de poner más rápidamente a disposición de los usuarios las bases de datos censales completos y depurados, así como para presentar los tabulados impresos generales, se incorporó simultáneamente un conjunto de innovaciones tecnológicas.

En ese momento, una de las alternativas tecnológicas posibles era la captura de datos mediante lectores ópticos, ya ampliamente probada en otros censos de la región. Sin embargo, ésta sólo reportaba una ventaja marginal en términos de la mano de obra requerida para la captura y suponía costos importantes en la impresión de los cuestionarios, en tanto que la tarea de codificación de las preguntas abiertas quedaba sin una solución alternativa a la tradicional.

La captura mediante tecnología de imágenes —finalmente aplicada— consiste en tomar imágenes electrónicas de los cuestionarios, lo que incluye su reconocimiento para convertirlas en datos comprensibles, por ejemplo, en formato ASCII (*American Standard Code for Information Interchange*). Este reconocimiento no se restringe solamente a marcas en lugares predeterminados, sino que también permite la interpretación de textos alfanuméricos escritos manualmente, de acuerdo con ciertos parámetros. La introducción en medios magnéticos de respuestas textuales a las preguntas abiertas del cuestionario facilitó igualmente el desarrollo de sistemas de codificación automática no aplicados previamente. Asimismo, la rapidez del proceso y la condición de disponer casi exclusivamente¹ de

¹ La tecnología de captura de datos aplicada permite disponer de archivos de imágenes completas de todas las áreas cubiertas por el cuestionario y, eventualmente, traerlos con facilidad a la pantalla de una computadora para su observación. Sin embargo, dada la magnitud de la inversión en equipos que ello significaba, monto no considerado en el presupuesto, impidió aplicar estas facilidades en forma sistemática en algunas fases de la crítica y depuración de los datos.

los archivos ASCII procedentes de la captura de datos condujeron a una depuración de datos totalmente automatizada mediante programas CONCOR.

Como resultado de lo anterior cambió radicalmente la organización del trabajo de elaboración de los datos posterior al relevamiento, ya que se redujo a un mínimo el personal de nivel medio. Asimismo, gracias a la automatización casi completa del ingreso a medios magnéticos, la codificación y la depuración de los datos, fue posible poner en práctica un control de calidad más acucioso de todos los procesos al asegurar para algunos de ellos la uniformidad de criterios y, para otros, obtener información sobre los niveles de error registrados, esto mediante su control durante el proceso de captura con el fin de asegurar su permanencia dentro de los rangos permitidos.

Al tomar la decisión de aplicar estas tecnologías —básicamente la de captura de imágenes e interpretación inteligente de caracteres alfanuméricos— se asumieron varios riesgos, entre los que cabe mencionar los referidos al hecho de adoptar una tecnología no probada hasta el momento en su totalidad a la escala de un censo, así como a la capacidad de la institución para interactuar en forma relativamente compleja con el proveedor externo del servicio de captura de datos, a fin de cumplir con un cronograma eficiente de tareas complementarias en el corto tiempo establecido para la operación y asegurar, a la vez, el cumplimiento de los niveles de calidad exigidos.

Sin embargo, y a pesar de las pruebas realizadas, se estima que uno de los mayores riesgos que debieron asumir el INE y la empresa consultora proveedora de servicios fue el de la calidad del llenado del cuestionario a la escala real del relevamiento de un censo “de hecho”, efectuado en un solo día por empadronadores censales corrientes.² Es por ello que —aparte del énfasis puesto en el llenado del cuestionario durante la capacitación de los empadronadores—, uno de los puntos de mira y de negociación más importantes entre el INE y la empresa proveedora del servicio fue el diseño del cuestionario, hasta en sus más mínimos detalles, con vistas a capturar la información buscada. A fin de dar las mayores garantías de que el diseño gráfico del cuestionario cumpliera con las normas requeridas por la consultora, el operador contratado para elaborarlo fue propuesto por dicha empresa.

² Aparte de algunos voluntarios, los encuestadores son, en su gran mayoría, funcionarios públicos, preferiblemente docentes, sin remuneración monetaria y sólo con algunas horas de *entrenamiento*.

I. DISEÑO DEL CUESTIONARIO

El cuestionario aplicado en los últimos censos de población, hogares y viviendas del Uruguay hasta el del año 1985, con ligeras variantes, había probado su eficiencia en las distintas etapas de la elaboración tradicional antes mencionada. Estructurado como un cuestionario único, integraba las distintas unidades de relevamiento: la vivienda, el hogar y las personas, estas últimas con una configuración en forma de columnas. En él se contemplaban las condicionantes habituales del diseño de un cuestionario censal e incorporado al conjunto de los distintos operativos censales había dado resultados positivos.

En este sentido, según la experiencia anterior —de no mediar el cambio tecnológico en la captura de datos— las mayores dificultades se referían a la revisión de las distintas preguntas con miras a su actualización ante las nuevas necesidades presentes, definidas en amplias consultas con los usuarios y discutidas con grupos especializados de ellos, así como al mejoramiento de la captación de ciertas variables, tales como la actividad económica de las personas.

II. CONDICIONANTES ESPECÍFICAS DE LA TECNOLOGÍA DE IMÁGENES

El reconocimiento de una imagen exige que su ubicación, tamaño y forma hayan sido bien predeterminados y se encuentren libres de la interferencia de elementos ajenos. Así, deben incluirse en el cuestionario los siguientes elementos:

- Puntos de referencia adecuados para ubicar geométricamente la posición de cada ítem de información.
- Códigos que permitan identificar la parte del cuestionario que se está reconociendo (vivienda, hogar, persona).
- Una identificación del cuestionario.

Estos elementos, si bien ocupan algo de espacio, no inciden en el registro de los datos en el formulario.

Sin embargo, el reconocimiento de cada tipo de dato requerido en el cuestionario impone exigencias específicas:

- En el caso de las respuestas “cerradas”, que se registran en forma de marcas, éstas tienen que realizarse en forma nítida dentro del espacio destinado a tal fin, sin producir “desbordes” que interfieran con los

espacios destinados a otras respuestas. Para el registro de las marcas se definieron rectángulos de 1 por 6 mm, que debían llenarse con una línea recta horizontal.

- Las respuestas “abiertas”, cuyo registro se realiza en forma de palabras o números, requirieron que su ubicación y tamaño fueran determinados en forma exacta. Para ello, luego de diversas pruebas, se definió un rectángulo de 7 por 5 mm, con un adecuado entorno neutro, dentro del cual había que ubicar cada carácter con la mayor precisión posible. En el caso de palabras, éstas debían ser escritas en letras de molde mayúsculas.
- Para las respuestas “abiertas” también fue necesario definir a priori el número máximo probable de letras necesarias para el llenado correcto de cada tipo de respuesta, debido a que no era posible reconocer palabras cortadas por el paso a otra línea. Como resultado de esta exigencia se diseñaron espacios con una fila de 15 rectángulos (9.7 cm en total) para establecer el “departamento” y dos filas de iguales dimensiones para registrar la “localidad” en las preguntas sobre movilidad espacial. En el caso de las referidas a “orientación” o “carrera” cursada se dispuso una línea de 25 rectángulos (16.5 cm), y dos líneas similares para las respuestas sobre “estudio técnico” o “comercial” cursado, así como sobre “tipo de ocupación” desempeñada y “rama de la actividad económica”.
- En el caso concreto de las respuestas expresadas con números manuscritos —dado el mayor margen de error en la interpretación esperado de tal tipo de datos— éstas se registraron paralelamente en forma de marcas. Ello tuvo por objetivo aumentar la “redundancia”, vale decir, el apoyo en otro tipo de referencias que permitieran una mejor aproximación a la interpretación inteligente de la imagen o, en un caso extremo, la reparación mediante la digitación por parte del operador.

Asimismo, la tecnología de imágenes impone el uso de colores de fondo específicos para el área en torno de aquella donde deben registrarse las respuestas, de manera que resulten “invisibles” para el escáner y resalten el registro del caso.

Por otra parte, es preciso evitar que existan interferencias de áreas oscuras (ya sea de marcas o letras impresas) que, por transparencia desde el reverso de la hoja sobre un área de registro, conduzcan a falsas lecturas. Esto debe controlarse mediante la utilización de un grosor de papel adecuado (la sugerencia fue de 100 mg/m²), o con un diseño que evite específicamente ese efecto.

También se requiere controlar especialmente la calidad de la impresión con el fin de evitar manchas no deseadas, así como mantener la uniformidad del alineamiento interno y la distribución de los lugares de registro.

Por último, en la medida en que el escáner va leyendo hojas separadas en secuencia, los cuestionarios tienen que pasar —antes de la lectura— por un proceso previo de guillotinado que asegure cortes netos y tamaño uniforme de las hojas, las que deben estar libres de cualquier elemento extraño.

Según se desprende de lo anterior —salvo en el caso de los dos últimos puntos, que tienen consecuencias atinentes en particular a los procesos de control de calidad del material e introducen una tarea nueva en el proceso—, el resto de las condicionantes de la tecnología de imágenes antes referidas incidieron muy directamente en el diseño del cuestionario censal, tanto en lo que se refiere a su forma como, en particular, al tamaño de las superficies en que debían responderse las preguntas.

Aparte de lo dicho, otras condicionantes específicas del contexto determinaron la solución final adoptada.

III. CONDICIONANTES ESPECÍFICAS DEL CONTEXTO

Todas las organizaciones, en especial las estatales, tienen sus tiempos específicos para la tramitación burocrática que implica la contratación de servicios prestados por particulares, lo que conlleva, en general, plazos prefijados relativamente extensos e inamovibles. Ello exige, por consiguiente, definiciones concretas y tempranas en cuanto al producto o servicio que se va a contratar.

En nuestro caso, en el llamado a licitación se definió el cuestionario, en forma de librito, para mostrar que el área correspondiente a cada una de las unidades de relevamiento (vivienda, hogar y persona), debía ocupar el espacio de una página, con un tamaño A4 como máximo para el área de lectura mediante el escáner. Por otra parte, limitaciones presupuestarias y también de espacio de archivo impusieron la decisión de mantener el grosor del papel del formulario en 72 mg/m^2 , lo que fue aceptado por la empresa consultora contratada.

IV. RESOLUCIÓN FINAL

En la resolución final fue preciso balancear los siguientes aspectos:

- El contenido temático del cuestionario acordado previamente, en particular, respecto de la unidad dedicada a la persona:
 - Se introdujeron variantes en la indagatoria acerca de la actividad económica desempeñada, lo que significó incorporar un mayor número de preguntas independientes.

- Se agregó la investigación de una nueva variable referida a la cobertura de asistencia médica.
 - Se amplió la indagatoria acerca de la educación al introducir detalles paralelos sobre la enseñanza técnica.
- El tamaño considerablemente mayor que requiere el registro de las respuestas:
 - La superficie del cuestionario destinada al registro de respuestas a preguntas abiertas (tipo de ocupación, rama de actividad, departamento, localidad, otros) prácticamente se cuadruplicó, con el requisito adicional de tener anchos bastante extensos ya prefijados.
 - El registro de la edad de las personas ocupó más de cuatro veces el espacio normalmente destinado a escribir una cifra de dos dígitos.
 - La limitación del espacio prefijado a una sola página de tamaño A4 por unidad de relevamiento. Esto, si bien en el caso de vivienda y hogar resultaba ajustado, pero suficiente, en el de la indagatoria acerca de las personas era prácticamente imposible. Por consiguiente, en una instancia posterior se habilitó un espacio total de 27 por 28.9 cm para el área de lectura del escáner.

Como resultado de este conjunto de condicionantes, las distintas preguntas, tanto en de la página de vivienda como en la de hogar, se estructuraron en una secuencia de columnas, con separación de bloques particularizados para cada uno de los ítems generales; asimismo, mediante los planos de color en los que van incluidos los espacios para las marcas, se procuró delinear columnas que indicaran una secuencia clara del flujo de las distintas preguntas.

En el caso de la página dedicada a persona, el mayor espacio obtenido hizo posible dividir la página en dos columnas de distinta anchura, fijadas en función de la longitud determinante de los espacios destinados a las respuestas alfabéticas, lo que permitió delinear bloques relativamente distinguibles con los conjuntos de preguntas dirigidas a universos de personas diferentes según edad y sexo, así como también con las indagaciones relativas a fecundidad.

Sin embargo, en este caso, el espacio disponible fue insuficiente para una presentación clara y lineal de la secuencia de preguntas dentro de cada bloque. Varias de ellas debieron ser dispuestas en flujos horizontales. En estos casos, si bien la visualización se reforzó con elementos gráficos,

éstos fueron exclusivamente flechas algo más oscuras, por la limitación de no disponer de otros medios, como podría ser el uso de distintos colores de tintas.

V. ALGUNOS RESULTADOS

1. El doble registro de respuestas numéricas

En la medida en que el doble registro de repuestas numéricas permitió que eventuales discrepancias entre los datos interpretados fueran reparadas por un digitador, no es posible cuantificar el mejoramiento de la calidad de la lectura de números manuscritos atribuible a esta práctica. La empresa consultora que prestó el servicio no llevó registros pormenorizados de este tipo de reparaciones. Sólo puede saberse que, globalmente, un 20% de los ítems leídos se remitió a los operadores por interpretación dudosa.

Otro aspecto del doble registro que cabe señalar —por ejemplo, en el caso de la edad de las personas— es el grado de ayuda que puede prestar en la etapa de depuración de los datos, ante casos de error por parte de los empadronadores en uno u otro tipo de registro.

2. La calidad de la captura de las marcas

La calidad de la lectura para asegurar que no se excedan los límites de error aceptados fue controlada durante todo el proceso mediante el examen de muestras de cada una de las partidas de trabajo intercambiadas entre el INE y la empresa prestadora del servicio de captura. Para ello se extrajo aleatoriamente, de cada una ellas, una muestra sistemática de cuestionarios; en éstos, además de comprobar que todas sus páginas estuvieran en el archivo ASCII, se verificó directamente el contenido completo de una página en secuencia de cada uno de ellos.³ Sobre la base de esta información se aceptó o rechazó la partida.

De acuerdo con estos controles se pudo establecer que las tasas de error definitivas registradas en el archivo censal completo fueron las siguientes:

– marcas	0.42% (± 0.02)
– números manuscritos	1.06% (± 0.09)
– palabras manuscritas	1.73% (± 0.13)

³ Se tuvo así una muestra de 3 375 páginas de vivienda, 3 580 páginas de hogar y 9 709 páginas de persona.

Sin embargo, “cuando fijamos tasas de error a priori y globales como en nuestro caso, estamos implícitamente pensando en que los errores van a estar uniformemente distribuidos sobre todo el material. Un análisis más detenido de la información contradice esta hipótesis” (Labuonora, 1998).

En efecto, las tasas de error correspondientes a los distintos campos de variables dentro de cada tipo de registro presentan altos coeficientes de variación. Esto revela la existencia de factores incidentes no controlados, probablemente múltiples, entre los que es posible identificar los siguientes:

- efectos del ajuste del sistema de captura
- elementos del diseño del cuestionario
- diferencias entre las formas en que los empadronadores efectúan los registros, que en algunos casos pudieron estar también vinculadas al condicionamiento para su manejo impuesto por el propio diseño del cuestionario.

a) *Tipos de formato de campos de marcas*

Un primer aspecto que cabe analizar es el referido a los distintos tipos de formato de los campos de marcas en relación con los errores que pueden introducir los empadronadores como resultado de marcas defectuosas. Los tres tipos de formato que se pueden identificar presentan las siguientes características:

- un único espacio para marcar (tipo 0)
- espacios para marcas sobrepuestos, en sentido vertical (tipo 1)
- espacios para marcas contiguos, en sentido horizontal (tipo 2)

Puede suponerse a priori —sin considerar la incidencia de otros factores— que la lectura de marcas en los campos de tipo 0 debería contener el menor número de errores, debido a su sencillez. Por el contrario, en los de tipo 1 —en la medida en que los empadronadores pueden haber hecho marcas distintas de las requeridas (tildes, cruces, otras)— es posible que se rebase el espacio de la marca deseada y se cree confusión al ocupar parte del área correspondiente a otras categorías de la misma variable. En cuanto a los de tipo 2, pueden esperarse solamente errores mínimos, cercanos a los del tipo 0, o por lo menos de nivel intermedio, ya que, por la disposición horizontal de los espacios, resultaría más difícil que en el formato anterior que ocurrieran invasiones de marcas de un espacio a otro, sea con tildes o cruces. El análisis de los resultados muestra, sin embargo, la siguiente distribución:

Cuadro 1
**NIVELES DE ERROR DE LECTURA DE MARCAS
 SEGÚN PÁGINA Y TIPO DE FORMATO**

VIVIENDA Página 1 – Formato	Error promedio (porcentaje)	Desviación estándar	Coefficiente de variación
Tipo 0	0.30	No corresponde	No corresponde
Tipo 1	0.43	0.00495	1.15
Tipo 2	0.44	No corresponde	No corresponde
Total página 1	0.42	0.00470	1.11
HOGAR Página 2 – Formato	Error promedio (porcentaje)	Desviación estándar	Coefficiente de variación
Tipo 0	No corresponde	No corresponde	No corresponde
Tipo 1	0.40	0.00341	0.85
Tipo 2	0.08	No corresponde	No corresponde
Total página 2	0.39	0.00340	0.87
PERSONA Página 3 o más/menos Formato variación	Error promedio (porcentaje)	Desviación estándar	Coefficiente de variación
Tipo 0	0.22	0.00146	0.67
Tipo 1	0.39	0.00591	1.52
Tipo 2	0.57	0.00400	0.70
Total página 3	0.43	0.00516	1.19
RESULTADO TOTAL Formato	Error promedio (porcentaje)	Desviación estándar	Coefficiente de variación
Tipo 0	0.23	0.00113	0.49
Tipo 1	0.40	0.00457	1.14
Tipo 2	0.54	0.00369	0.68
Total formulario	0.42	0.00439	1.04

Fuente: elaboración propia.

Tal como puede apreciarse en el cuadro anterior, la lectura de marcas de tipo 0 es la que presenta el promedio de errores más bajo, significativamente menor que el general.

Paralelamente, las marcas de tipo 1 (arreglo vertical) —que, por otra parte, son la mayoría en el formulario— muestran porcentajes de errores de lectura en torno del promedio general. Llama la atención su alto coeficiente de variación, particularmente en las páginas 1 y 3.

Sin embargo, las marcas de tipo 2 (arreglo horizontal) —al contrario de lo esperado— son las que exhiben el mayor porcentaje de lecturas erradas en el total del formulario, con un alto coeficiente de variación. Según se desprende del mismo cuadro, las marcas de este tipo muestran altos promedios de error de lectura en la página 1 y, particularmente, en las

páginas 3 o más, donde alcanza a 0.57%, valor que supera sustancialmente el promedio general. Sin duda, otros factores, aparte del tipo de formato de los espacios para marcas, han incidido en estos promedios.

El esquema de las páginas 3 o más del formulario, que comprende las variables correspondientes a personas, muestra una distribución sugestiva de los errores de lectura de marcas. Las variables con marcas de tipo 2, que son “fintec” (número de años aprobados en los estudios técnicos), “hijos vivos” (número de hijos nacidos vivos tenidos por las mujeres) y “vivosactua” (hijos actualmente vivos), presentan tasas de error de lectura de 0.69%, 1.48% y 0.80%, respectivamente, niveles que superan de modo significativo los de la mayoría de las otras variables comprendidas en la página y están, sin duda, por encima del promedio general.

Este conjunto de variables, salvo “fintec”, se ubica en el extremo inferior derecho de la página, área donde se puede observar una concentración de variables que —independientemente del tipo de marca con que se registren— muestran porcentajes de error significativamente más altos que el promedio general. Un caso extremo es la variable con marca de tipo 1, “categoría” (categoría de la ocupación desempeñada), cuya tasa de errores de lectura alcanza al importante valor de 2.53%.

Tal como puede apreciarse en el cuadro anterior, la lectura de marcas de tipo 0 es la que presenta el promedio de errores más bajo, significativamente menor que el general.

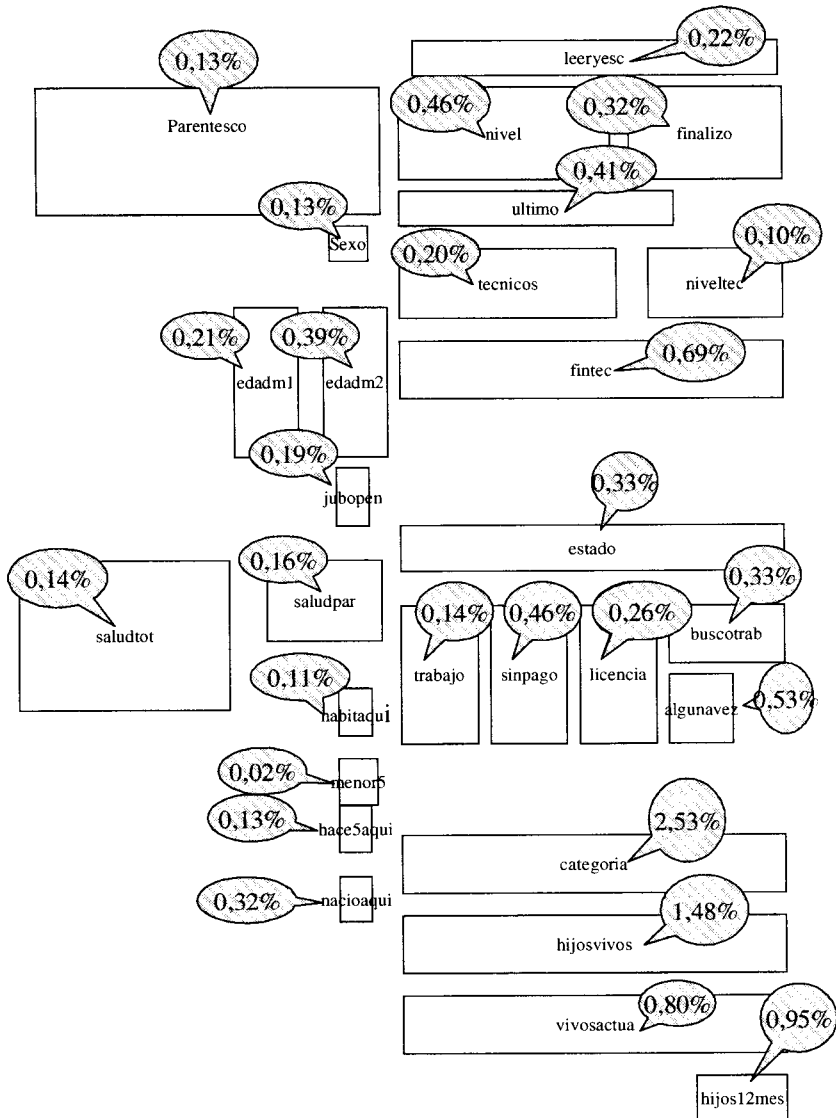
Paralelamente, las marcas de tipo 1 (arreglo vertical) —que, por otra parte, son la mayoría en el formulario— muestran porcentajes de errores de lectura en torno del promedio general. Llama la atención su alto coeficiente de variación, particularmente en las páginas 1 y 3.

Sin embargo, las marcas de tipo 2 (arreglo horizontal) —al contrario de lo esperado— son las que exhiben el mayor porcentaje de lecturas erradas en el total del formulario, con un alto coeficiente de variación. Según se desprende del mismo cuadro, las marcas de este tipo muestran altos promedios de error de lectura en la página 1 y, particularmente, en las páginas 3 o más, donde alcanza a 0.57%, valor que supera sustancialmente el promedio general. Sin duda, otros factores, aparte del tipo de formato de los espacios para marcas, han incidido en estos promedios.

El esquema de las páginas 3 o más del formulario, que comprende las variables correspondientes a personas, muestra una distribución sugestiva de los errores de lectura de marcas. Las variables con marcas de tipo 2, que son “fintec” (número de años aprobados en los estudios técnicos), “hijos vivos” (número de hijos nacidos vivos tenidos por las mujeres) y “vivosactua” (hijos actualmente vivos), presentan tasas de error de lectura

ESQUEMA DE PÁGINA 3 - PERSONA

Porcentajes de error en captura de marcas según variable



de 0.69%, 1.48% y 0.80%, respectivamente, niveles que superan de modo significativo los de la mayoría de las otras variables comprendidas en la página y están, sin duda, por encima del promedio general.

Este conjunto de variables, salvo “fintec”, se ubica en el extremo inferior derecho de la página, área donde se puede observar una

concentración de variables que —independientemente del tipo de marca con que se registren— muestran porcentajes de error significativamente más altos que el promedio general. Un caso extremo es la variable con marca de tipo 1, “categoría” (categoría de la ocupación desempeñada), cuya tasa de errores de lectura alcanza al importante valor de 2.53%.

La observación del formulario permite constatar que en las áreas donde se ubican las variables antes reseñadas y, en particular, la correspondiente a “categoría”, existen importantes efectos de transparencia de elementos impresos en tinta negra en el reverso de la página, o de superposición de espacios de registro en ambas caras de la página.

b) El caso particular de la página 2 – Hogar

En el caso de las variables comprendidas en la página 2, correspondiente a la unidad hogar del formulario, la observación de los errores de lectura de los registros de marca permite comprobar que la casi totalidad de ellos se debe a lecturas en blanco. Según se desprende del cuadro 2, en gran parte del conjunto de variables contenido en esta página el error de captura es atribuible en su totalidad a tal causa.

Si bien la página tiene un 0.39% de error de lectura (inferior al promedio), 91.2% de esa proporción se explica por la captura de un valor “blanco”, en lugar de la marca que se había registrado.

Los errores de captura que se muestran en el cuadro 2 podrían explicarse por una amplia gama de factores que sería difícil dilucidar en la actualidad. ¿Se trata de problemas de ajuste del sistema de captura o de factores asociados al registro por parte de los empadronadores? Por ejemplo, éstos pueden doblar o arrugar una hoja del formulario al escribir, lo que oportunamente que se les pidió no hacer.

VI. ALGUNAS CONCLUSIONES

Sin duda, toda experiencia pionera tiene sus costos. Ninguna prueba previa puede —en estos casos— eliminar las dudas respecto de los resultados finales del experimento a escala real. La incertidumbre induce a aumentar los resguardos para prevenir ciertos tipos de riesgos, que luego, en la realidad, parecen innecesarios. Tal puede haber sido el caso con el registro paralelo de marcas y números manuscritos, que significó sacrificar parte del escaso espacio disponible para otros aspectos importantes del cuestionario. Asimismo, ante la imposibilidad de controlar todos los riesgos

Cuadro 2

Variable página 2	Error (porcentaje total)	Porcentaje de error explicado por lectura en blanco
Totmujmar1	1.23	100.0
Lavacomun	1.15	87.8
Tothommar2	1.12	100.0
Totmujmar2	0.89	100.0
Tothommar1	0.75	100.0
Video	0.67	91.7
Vehículo	0.56	100.0
Pc	0.45	87.5
Fuentecale	0.42	26.7
Lavaprog	0.42	100.0
Totalm2	0.42	100.0
Tvcolor	0.39	85.7
Teléfono	0.36	84.6
Evacserhig	0.28	100.0
Tvbn	0.25	88.9
Usoservhig	0.25	88.9
Calefacción	0.22	75.0
Totalm1	0.22	100.0
Calentador	0.20	100.0
Freezer	0.20	85.7
Fuentecoci	0.17	83.3
Refrisimpl	0.14	60.0
Higiénico	0.11	75.0
Refriconfr	0.11	50.0
Habdormir	0.08	100.0
Lugarcocin	0.08	100.0
Microonda	0.08	100.0
Tenencia	0.08	66.7
Calefón	0.06	100.0
TOTAL	0.39	91.2

potenciales es posible que se descuiden riesgos reales, como sucedió, en definitiva, con los problemas de transparencias entre ambas caras de las hojas del formulario.

Sin embargo, el balance final puede calificarse de positivo. Las metas fueron alcanzadas en términos de plazos y niveles de error registrados.

Es la primera vez —por lo menos hasta donde se conoce— que se puede disponer de detalles tan pormenorizados sobre la distribución de los errores de captura, lo que hace posible identificar los logros y las deficiencias. Estas últimas preocupan, en la medida en que no pudo obtenerse una calidad uniforme en todas las variables del cuestionario, pero parece factible obtener mejores resultados en el futuro. La experiencia relatada corresponde al primer semestre del año 1996, por lo que los avances

tecnológicos producidos desde entonces y las enseñanzas extraídas de ella han de permitir mejorías significativas en las actividades censales del futuro.

Lo cierto es que se ha podido proporcionar la información definitiva dentro de plazos anteriormente inconcebibles, con un adecuado nivel de calidad y a costos razonables. Todo ello permite pensar que se abre un camino importante, que justifica los esfuerzos por avanzar aún más allá de los logros alcanzados.