

15.153
M 6774.1

Revista Latinoamericana de Actividades en Estadística

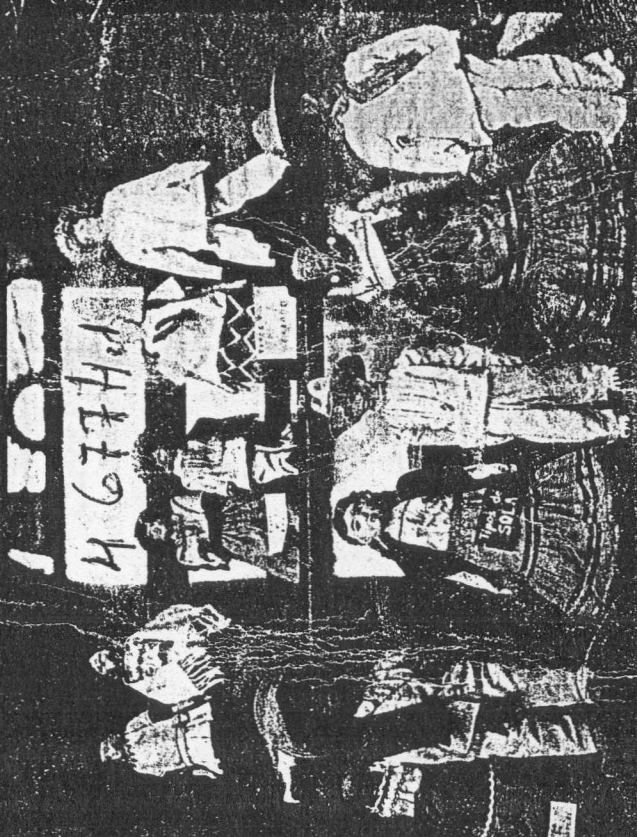
Carlos Welti
Editor

EMOGRAFIA I fue escrito para ser utilizado como material didáctico y como tal, cada uno de los capítulos está estructurado con este objetivo; en ellos se hacen explícitos los conceptos y temas a desarrollar. Se ofrece una exposición central y se presenta una serie de preguntas y respuestas para evaluar el conocimiento adquirido. Una característica sobresaliente es la inclusión de un *casquette* con ejercicios e información estadística para que el alumno pueda realizar sus propias prácticas.

Si bien el libro está dedicado a la enseñanza formal de la Demografía, el lector no especializado encontrará en él un material útil para conocer la terminología y los métodos que utilizan los demógrafos para estudiar la dinámica de la población. Los temas se abordan con rigor científico con un lenguaje sencillo y ameno, por lo que no hay necesidad de contar con un conocimiento avanzado en matemáticas para comprender la mayor parte de las explicaciones técnicas. Una proporción importante de las prácticas referidas en la bibliografía está escrita originalmente en español o traducida a este idioma, por lo que puede consultarse con facilidad.

La información estadística que sirve de base a las exposiciones y prácticas proviene de publicaciones y bases de datos recientes, ofreciendo un panorama actualizado de la población latinoamericana.

Este libro constituye un ejemplo de lo que puede hacerse para facilitar la enseñanza en otras disciplinas.



DEMOGRAFÍA I



The College of
California
Marshall
Foundation

Revista Latinoamericana de Estadística

altamente diferenciada, entre otras cosas, según sexo, edades y localización geográfica. En la misma línea, son las personas económicamente activas las que producen los bienes que se intercambian en los mercados y, por tanto, es de la mayor relevancia cuantificar la magnitud de esta población y proyectar su evolución futura. Este sería un ejemplo de uso de la Demografía en el campo de la Economía. Para un sociólogo, el conocimiento de las tendencias demográficas, de sus factores determinantes y de sus repercusiones es fundamental para la formulación y ejecución adecuada de políticas y acciones en campos tan variados como la planificación regional, la movilidad espacial de la población, la planificación familiar, etc.

4. Existe consenso en señalar que la postura malthusiana fue demasiado negativa y no previó los avances tecnológicos que expandieron la productividad y tampoco la difusión del control racional de la fecundidad. Hacer un análisis crítico del autor requiere, sin embargo, considerar la realidad de fines del siglo XVIII e inicios del siglo XIX, en que el agro era el principal ámbito productivo y la noción de rendimientos decrecientes era ampliamente aceptada.

En otro ámbito se plantea la teoría de la capilaridad social, desarrollada por Dumont para explicar el comportamiento reproductivo según la posición social estratificada. Esta tesis sostiene que el incremento de la movilidad social tiende a disminuir las aspiraciones reproductivas de los individuos, a causa de la concentración de recursos y las disposiciones emocionales que exige ascender socialmente. De este modo se reduce la natalidad en función directa a los beneficios económicos y sociales derivados del proceso de movilidad ascendente entre los estratos. Este planteamiento sostiene que la decisión de tener hijos es eminentemente racional, lo que coincide con interpretaciones vigentes y, en general, puede considerarse válido pero limitado, porque desconoce la importancia de los factores culturales y presupone un comportamiento racional en condiciones que a menudo no favorecen una decisión de este tipo.

CAPÍTULO II

Fuentes de datos

Objetivos específicos del capítulo:

- Mencionar y describir las principales fuentes de datos demográficos.
- Reconocer el uso de las principales fuentes de datos demográficos.
- Comparar la información que entregan las principales fuentes de datos demográficos.
- Definir, utilizar, y juzgar censos de población.
- Utilizar las principales fuentes de datos demográficos.
- Definir, identificar y explicar los principales errores típicos, de cobertura y contenido, presentes en los censos de población.
- Identificar formas de detección de los errores censales típicos.
- Conocer las encuestas demográficas.
- Comparar la información que entregan las principales fuentes de datos demográficos.
- Definir y explicar los principales errores típicos, de cobertura y contenido, de las encuestas demográficas.
- Definir, identificar, utilizar y juzgar las estadísticas vitales.
- Identificar y explicar los principales errores, de cobertura y contenido, típicos de las estadísticas vitales.

II.1. Tipos de datos demográficos

Es conveniente distinguir los datos utilizados en el análisis demográfico según ciertas características que reiteradamente se mencionarán en el presente texto:⁴

II.1.a. Según el objetivo que definió su recolección

Un dato se denominará primario cuando su recolección tuvo como finalidad satisfacer las necesidades u objetivos específicos de una investigación concreta. Se denominará secundario al que se utiliza en

⁴ Las distinciones realizadas tienen un propósito básicamente pedagógico y no responden necesariamente a convenciones establecidas entre los demógrafos. Así, por ejemplo, la definición de datos primarios no coincide con lo señalado en el *Diccionario Demográfico Multilingüe*, o con lo planteado por Shryock y otros (1976).

una investigación sin haber sido recogido en el contexto de ésta, es decir, información disponible o previamente existente.

Debe destacarse que esta distinción es relativa, porque un mismo dato puede ser primario para ciertas investigaciones y secundario para otras. Así, por ejemplo, toda la información disponible en el Ministerio de Educación constituye datos primarios para quienes planifiquen su recolección, pero secundarios para quienes la usen sin haber participado en la planificación de su recolección. Asimismo, los datos recolectados en una encuesta cualquiera son primarios para quienes la elaboraron, pero son secundarios para quienes utilicen la información posteriormente.⁵

⁵ La disponibilidad de bases de datos si bien no altera la condición primaria o secundaria de la información, sí permite un procesamiento "ad-hoc" de los datos superando los límites clásicos que impone la publicación oficial de los resultados.

La importancia de esta distinción está en la vinculación entre los datos y la investigación, y por tanto, con las limitaciones de uso y manejo de la información. En principio, los datos primarios parecen los más útiles (por estar directamente ligados a los objetivos de la investigación), pero no deben ser considerados como los únicos válidos para los científicos sociales. Existe la tendencia, sobre todo entre los sociólogos, a recurrir a datos primarios para realizar investigaciones, lo que normalmente implica el levantamiento de una encuesta o la necesidad de algún tipo de entrevista. Como contrapartida, a menudo no se utilizan los datos secundarios disponibles y, a veces, ni siquiera se sabe donde pueden encontrarse. Estos datos se recogen periódicamente en los países y pueden encontrarse en distintos Ministerios (Educación, Salud, Vivienda, Trabajo, Planificación, Economía, etc.), en oficinas gubernamentales como los Institutos Nacionales de Estadísticas, en Municipalidades, en el Registro Civil, etc. Existen, además, censos de diferente índole (población, vivienda, industriales, agrícolas, etc.) y encuestas (en Chile; la de caracterización socioeconómica —CASEN—, la de empleo del Instituto Nacional de Estadísticas, las políticas del Centro de Estudios Públicos y del Centro de Estudios de la Realidad Contemporánea) que se realizan con cierta periodicidad y que aportan datos de gran valor para el investigador.

En la actualidad es altamente posible el acceso a las cintas de datos, lo que permite a cada investigador efectuar los manejos de información que considere pertinentes. Por otro lado, el trabajo con datos disponibles evita, o al menos reduce, los costos que supone el levantamiento de la información. En definitiva, existe una serie de argumentos para señalar la importancia que tiene el acceso y el manejo de datos secundarios para el trabajo concreto de los científicos sociales y, en especial, para los interesados en los temas de la población.

II.1.b. Según la forma de obtenerlos

Son datos directos aquellos que se consiguen de manera simple y sin mayores supuestos, o los que representan series de números absolutos. Son indirectos los obtenidos luego de operaciones de cálculo cuya aplicación requiere de supuestos.

II.1.c. Según el grado de corrección que tienen

Se denominan datos brutos a los que se presentan sin desagregaciones, suavizamientos, ajustes o correcciones. Datos corregidos son aquellos que han sido sometidos a técnicas de suavizamiento, de ajuste, de corrección o de tipificación. Son refinados los datos que se presentan en una forma más detallada y precisa, con la intención de proporcionar una idea más exacta de lo que se pretende analizar.

Por suavizamiento se entiende la eliminación de fluctuaciones bruscas de la información, sin que esto signifique mayores cambios en los datos originales. Por ajuste debe entenderse la sustitución, parcial o total, de una serie de datos observados, por otra que —basándose en éstos— puede ser expresada como una curva matemática. La idea de corrección alude a la operación genérica de introducir modificaciones a los datos originales. En el caso de la investigación demográfica, que no es extensible al resto de las ciencias sociales, existen procedimientos de corrección relativamente convencionales, porque existen errores típicos. Al hablar de estandarización (o tipificación) se alude a una técnica de descomposición que permite controlar el efecto perturbador que ciertas características (por ejemplo, en el campo demográfico, la edad, el sexo, la nupcialidad, etc.) tienen sobre algunos datos e indicadores.

II.2. Fuentes de datos

II.2.a. Introducción

En el capítulo anterior se señaló que la Demografía era un ejemplo para el resto de las ciencias sociales en lo que se refiere a explotación de los datos disponibles. En efecto, como cualquier demógrafo sabe que la descripción y el análisis de la mortalidad, la fecundidad, la migración, la estructura y la dinámica de la población de una determinada región o país deben necesariamente incluir ciertas medidas (que serán revisadas en los capítulos pertinentes), uno de los campos donde los estudiosos de la población han concentrado sus esfuerzos es en el de captar información relevante, utilizar datos ya disponibles y evaluar y, en caso que sea necesario, corregir esta información para su uso en la investigación y descripción demográfica o en la planificación social. Para tales efectos, se han desarrollado técnicas de la

más diversa índole, y se han destinado grandes recursos a mejorar los sistemas de recolección de información sociodemográfica en todos los países.

Las principales fuentes de datos que se utilizan en el análisis demográfico son los censos, las encuestas demográficas y las estadísticas vitales.

II.2.b. El censo

Se define como "un conjunto de operaciones que consiste en reunir, elaborar y publicar datos demográficos, y también económicos y sociales, correspondientes a todos los habitantes de un país o territorio definido y referido a un momento determinado o a ciertos periodos de tiempo dados" (Naciones Unidas, 1978). La acción de recoger información censal puede expresarse como censar o empadronar. Los principales usos que ofrecen los censos para la investigación sobre población son:

- i) Se trata del procedimiento más exhaustivo para registrar los cambios en la magnitud y composición de la población en dos momentos del tiempo, datos que son esenciales para las proyecciones de población.
- ii) Es la fuente básica —sin considerar los registros de migración, que son prácticamente inexistentes en América Latina— para el estudio de la migración y juega un rol creciente en la medición de la fecundidad y mortalidad, a causa de la inclusión de preguntas que permiten la estimación indirecta de estos componentes del cambio de la población.
- iii) Al recoger datos demográficos y, a la vez, información socioeconómica individual y familiar, es posible el estudio, con limitaciones que se comentarán más adelante, de las interrelaciones entre las características demográficas y socioeconómicas de individuos y de hogares.

Las principales características de un censo de población son las siguientes:

- i) *Secreto estadístico*: los datos que se recogen no pueden servir para identificar personas específicas bajo ningún pretexto. Si en la boleta censal se pregunta el nombre de los empadronados es como mecanismo de "fami-

liarización" entre entrevistado y empadronador y también para facilitar el listado y la contabilización de los empadronados.

- ii) *Auspicio oficial*: es patrocinado por el Estado, independientemente que la agencia ejecutora sea gubernamental o privada.
- iii) *Territorio bien definido*: debe ejecutarse en una zona claramente establecida. Un censo nacional de población puede excluir ciertas regiones del país por razones de seguridad o accesibilidad, pero debe establecer precisamente los límites de las áreas que serán censadas y los de las que no lo serán.
- iv) *Universalidad*: es imprescindible que toda la población del territorio censal definido sea empadronada; nadie debe ser excluido.
- v) *Unidad censal*: en un censo de población la unidad censal es el individuo, y en el de vivienda es la vivienda. Esto no implica, por cierto, la imposibilidad de efectuar preguntas sobre el hogar como un todo. De hecho, la unidad de referencia de las boletas censales es, normalmente, el hogar.⁶
- vi) *Simultaneidad*: es necesario que la información esté referida a un momento preciso del tiempo.

Existen tres recomendaciones internacionales para el levantamiento de censos. La primera es que se levanten simultáneamente en todos los países y en los años terminados en cero. La segunda es que tengan una periodicidad constante (cada diez años). La tercera es que se ejecuten conjuntamente el censo de población y el de vivienda. Esto último, además de implicar un mejor uso de los recursos disponibles, permite ampliar el conocimiento de las condiciones de vida de la población y de la relación entre las variables habitacionales y las sociodemográficas.

La planificación total de un censo consta de tres etapas: la de pre-empadronamiento, la de empadronamiento y la de post-empadronamiento. El pre-empadronamiento incluye una fase previa en la cual el censo es sancionado legalmente (por decreto), se estructura la operación en términos administrativos y se define un cronograma inicial de actividades. Luego, siempre dentro de la etapa de pre-empadro-

⁶ Hogares particulares. Los hogares colectivos (cárceles, hospitales, comisarías, etc.) constituyen situaciones especiales.

namiento, existe un trabajo geográfico y cartográfico que actualiza los mapas nacionales y determina las zonas censales utilizando criterios relacionados con el clima, topografía, transporte, etc. Conjuntamente con lo anterior, se definen los temas y preguntas censales. Debe tenerse en cuenta que, por consideraciones de costos y de tiempo, las preguntas de un censo han de ser pocas y lo más precisas posibles.

En la definición de los temas se toman en cuenta diversos factores, como: a) las necesidades de información que tiene el país; b) la comparabilidad internacional y con censos nacionales anteriores; c) la disposición a responder (no deben incluirse temas que puedan prestarse para suspicacias o rechazos); y, d) los costos que genera la inclusión de cada tema (causados, por ejemplo, por el espacio de impresión, el tiempo de respuesta, codificación, digitación, análisis de las respuestas).

Varios temas son usuales en los censos de población:

- i) Geográficos: la ubicación de la persona en el momento del censo, su residencia habitual y el tipo de localidad (urbana o rural) en que fue empadronada.
- ii) Demográficos: datos sobre el hogar y relación de parentesco entre sus componentes. Siempre se pregunta edad, sexo, estado conyugal y lugar de nacimiento de cada individuo. Es muy frecuente que existan preguntas para captar migración, aunque éstas varían según los países. De manera creciente se incluyen preguntas para estimar, de manera indirecta, mortalidad y fecundidad.
- iii) Educación: usualmente se pregunta por alfabetismo y por nivel de instrucción alcanzado por las personas. Es habitual que se incluya una pregunta sobre asistencia actual a la escuela.
- iv) Actividad económica: normalmente se incluyen preguntas para averiguar si el individuo es económicamente activo o inactivo, cuál es la ocupación específica que tiene, en qué rama de actividad económica se desempeña (sectores primario, secundario y terciario) y cuál es su categoría ocupacional (patrón, empleado, cuenta propia, familiar no remunerado, etc.).
- v) Otras: dependiendo del país, se pregunta por rasgos como etnia, religión, idiomas que domina, impedimentos físicos, etc.

En general, se recomienda que el cuestionario sea lo más reducido y precodificado posible y que cada pregunta esté redactada de la manera más simple y precisa. Los censos contemplan, durante la etapa de pre-empadronamiento, ensayos del cuestionario, pruebas de campo y censos pilotos en zonas seleccionadas. Esto permite evaluar los requerimientos concretos que existirán al momento de la operación definitiva, calcular el tiempo medio de respuesta y medir la calidad de las preguntas.

Si bien el empadronamiento es el levantamiento mismo de los datos, como etapa se inicia en el proceso de capacitación de los empadronadores y supervisores. La experiencia del censo piloto se considera en esta capacitación.

Existen dos tipos de empadronamiento: el *de facto* y el *de jure*, también denominados de hecho y de derecho. El primero empadrona a las personas en el lugar donde se encuentran en el momento del censo, para lo cual normalmente se utiliza la pregunta relativa al lugar donde pasó la noche anterior. El segundo empadrona a las personas en su lugar de residencia habitual. Es decir, si un empadronador encuentra en la vivienda a un sujeto que no reside de manera habitual en ella, no lo entrevistará o contabilizará, incluso en el caso que haya dormido allí la noche anterior. En todos los casos, el empadronamiento lo realiza un individuo y su gestión es supervisada por un superior.

Luego del levantamiento del censo viene la etapa de post-empadronamiento, que incluye la recepción y verificación de los cuestionarios, la codificación (manual o por lectora óptica), la aplicación de programas de crítica y asignación, el procesamiento estadístico (obtención de tabulados) y la publicación de los tabulados. A veces se considera la realización de una encuesta de evaluación, con el fin de verificar errores de cobertura y calidad de la información. En algunas ocasiones se ha utilizado el muestreo para adelantar resultados provisionales.

Los resultados de un censo se van haciendo públicos de manera escalonada. A los pocos días pueden entregarse las cifras de la población total del país y algunas de sus principales características (sexo, localización geográfica). Posteriormente pueden entregarse cifras provisionales. Sin embargo, es difícil que los datos definitivos, con todas las tabulaciones programadas realizadas, estén listos antes de un año de levantado el censo. Es importante se-

ñalar que la operación censal no termina con el levantamiento de los datos, sino cuando se publican sus resultados.

Pese a los problemas que generalmente presenta la información (que serán tratados en el punto II.3), normalmente el censo es la fuente de datos de mayor importancia, porque solicita información de toda la población, aborda aspectos relevantes para el trabajo de cualquier científico social y es oficial. Por otro lado, la explotación de sus datos ha crecido últimamente con la disponibilidad de tecnología y "software" computacional, que permite manejar la información de un censo de manera desagregada geográfica y administrativamente con un computador personal. A lo anterior hay que agregar un factor de demanda, por cuanto la información a escala desagregada es cada vez más necesaria a causa de la creciente descentralización política administrativa.

II.2.c. Las encuestas demográficas

Son utilizadas como fuente de datos para análisis demográficos referidos, básicamente, a los componentes del cambio de la población. En general, permiten profundizar el conocimiento sobre la fecundidad, la mortalidad y la migración. Este último componente presenta dificultades adicionales, que serán tratadas en el capítulo pertinente. Las encuestas, habitualmente, recogen información a nivel de cada individuo, pero también consideran criterios para definir hogares. Las encuestas demográficas pueden ser de dos tipos:

- i) *Prospectivas*: registran todos los movimientos y hechos demográficos, y eventualmente de otro tipo también, de la población bajo observación durante un periodo de tiempo. Así, se aplica un cuestionario de manera reiterada a una misma persona o grupo de personas dentro de un hogar durante el periodo de estudio. A causa de lo anterior también se llaman encuestas de "visitas repetidas". A través del formulario se recogen datos y se registran los principales hechos demográficos (nacimientos, defunciones, enfermedades, matrimonios, migraciones, etc.) que ocurren en el lapso definido. La información que aportan estas encuestas es de gran valor

y permite un análisis riguroso de los potenciales determinantes de los hechos mencionados. Sin embargo, su costo es muy elevado y a menudo experimentan altas tasas de deserción muestral.

- ii) *Retrospectivas*: un cuestionario se aplica en una sola ocasión; las respuestas permiten reconstruir la historia demográfica de los individuos entrevistados. En este caso se utilizan, para la mortalidad y la fecundidad, preguntas que permiten la estimación directa e indirecta de ambas variables.⁷ En el caso de la migración, normalmente las preguntas indagan la historia de cambios de residencia de los individuos y las motivaciones para salir del lugar de origen y elegir el de destino.

La principal ventaja que tienen las encuestas es que, al usar el muestreo, implican menores costos y están en condiciones de contener una cantidad de preguntas significativamente mayor que las del censo. Desde este punto de vista, son más adecuadas para los estudios demográficos de mayor profundidad. Su principal problema radica en la existencia de errores de muestreo.

Existe una amplia experiencia en encuestas demográficas, que ha sido muy valiosa para países con información censal y de registro de hechos vitales deficientes. Debe destacarse la existencia de dos proyectos de encuestas demográficas realizadas a nivel mundial, la primera en la década de 1970 (WFS-World Fertility Survey) y la segunda durante la de 1980 (DHS-Demographic and Health Surveys), mediante las cuales se ha recogido información demográfica típica (fecundidad y mortalidad, básicamente) para numerosos países del mundo. Esta experiencia, que tiene un símil sociológico en las investigaciones mundiales sobre actitudes políticas o conductas sexuales, ha permitido estimar y comparar los principales indicadores sobre fecundidad y mortalidad infantil en distintas regiones del mundo.

⁷ Existe medición directa de la fecundidad y la mortalidad cuando no se recurre a preguntas retrospectivas o cuando, aun recogiéndose la información a partir de éstas, la estimación no requiere cálculos o supuestos muy complejos. Una estimación indirecta normalmente se basa en preguntas retrospectivas y requiere además supuestos adicionales y cálculos complejos.

II.2.d. Estadísticas vitales

Son registros oficiales sobre los principales hechos vitales que ocurren en una población, tales como nacimientos, defunciones, matrimonios, separaciones, adopciones, etc. En el pasado, estos hechos eran certificados por organismos eclesiásticos, pero en la actualidad son registrados por organismos gubernamentales, normalmente llamados registros civiles.

Un registro civil consiste en la inscripción legal y compulsiva, con carácter continuo y permanente, de los hechos vitales. El carácter compulsivo puede darse tanto a través de sanciones legales, como mediante la necesidad de constatar el registro para acceder a servicios básicos. Esta última situación puede ejemplificarse en el requerimiento que tienen los padres chilenos de presentar el certificado de inscripción de nacimiento de sus hijos para postular a asignaciones familiares o para inscribirlos en el sistema educacional. El registro civil tiene, por tanto, finalidades legales y también estadísticas o de recopilación de información.

El formulario de estadísticas vitales se denomina informe estadístico y en el caso chileno registra, para los nacimientos, datos sobre la fecha de ocurrencia e inscripción, orden de nacimiento, tiempo de gestación, lugar del parto y de la inscripción, tipo de atención del parto, tipo de nacimiento (simple o múltiple), sexo, peso y estatura del recién nacido, etc. El formulario de nacimientos también incluye información sobre los padres (en especial la madre), como lugar de residencia, nacionalidad, estado civil, edad, nivel de instrucción y actividad económica. En el certificado de defunción se registra información sobre el hecho mismo (causa de la muerte, lugar, fecha y existencia de certificación médica), y sobre el difunto (sexo, edad, nivel educacional, actividad económica, estado civil, etc.). Si el fallecido tenía menos de un año de edad, se incluye información adicional sobre la fecha y el orden de nacimiento, el tiempo de gestación y datos básicos sobre los padres.

Como se desprende de lo señalado sobre la información disponible en las estadísticas vitales, estos registros son de una gran utilidad para el estudio de la mortalidad y la fecundidad. La información que entregan permite investigar los niveles de estas variables en un país y en sus distintas regiones. Además, según el tipo de información incluida en la ficha de registro, posibilita el análisis de factores aso-

ciados a la mortalidad y a la fecundidad, tales como el nivel educacional, la edad, la zona de residencia, el estrato socioeconómico, la cantidad de embarazos anteriores, etc. En el caso de la migración, en cambio, su utilidad es muy reducida a causa de la inexistencia —o mala calidad— de información al respecto, excepto en aquellos países que disponen de registros permanentes de población (países nórdicos), o en los cuales los movimientos dentro del país son controlados rigurosamente (Cuba).

II.3. Problemas de las fuentes de datos demográficos

II.3.a. Introducción

En términos generales, los errores existentes en las principales fuentes de datos demográficos pueden dividirse en dos grandes tipos: a) los provenientes de deficiencias en la contabilidad de las personas o hechos, que se denominan errores de cobertura; y, b) los derivados de deficiencias en la información recogida, que se llaman errores de contenido. El estudio de los errores existentes en las fuentes de información demográfica es de gran importancia para evitar conclusiones equivocadas de los datos que entregan. Además, se ha detectado la existencia de errores típicos, que pueden ser eliminados o compensados a través de diversas técnicas. Un examen más detenido de los errores que registra la información obtenida a través de estas fuentes, demuestra la necesidad de un estudio particular de cada una de ellas (Chackiel y Maccio, 1979; Shryock, Siegel and associates, 1976).

II.3.b. Problemas de la información censal

En el caso del censo de población,⁸ los errores de cobertura se relacionan, básicamente, con la sub o sobreenumeración de personas. Estos errores tienen su origen en diversas etapas de la operación censal. Pueden originarse en deficiencias del trabajo carto-

⁸ Para un análisis de los errores detectados en los censos latinoamericanos realizados durante la década de 1980, puede revisarse el texto: *Los censos de población del 80, taller de análisis y evaluación*, Indec, 1985.

gráfico, errores en la instrucción de los empadronadores o inadecuada coordinación y supervisión de la recolección de datos. Pese a que, por definición, el censo debe registrar a todos los habitantes que se encuentran o residen habitualmente en las zonas definidas para el empadronamiento, a menudo sucede que una cantidad de personas no es contabilizada, como también que otras son empadronadas más de una vez.⁹

Normalmente, la cantidad de personas omitidas es mayor que el número de censadas. Debido a lo anterior, en general los censos tienen omisión neta, es decir, el número de personas que contabilizan es menor que la población realmente existente en las zonas censadas. Por cierto, esta omisión varía sustancialmente según el país y según la zona dentro de cada país, así como según otras variables (la edad, el estrato social, etc.).

Por ejemplo, la omisión en los censos de población realizados durante la década de 1980 fue más elevada en Guatemala (13.8%) y Ecuador (7.5%) que en Argentina (1.5%) y Brasil (1.8%) (Arretx y Chackiel, 1985). A menudo ocurre que ciertas regiones del país presentan menor subregistro que otras. Por su parte, es común que el grupo de los menores de un año tenga una omisión mayor que la prevalectante en los restantes grupos de edad.

Existe además la omisión completa de áreas geográficas, que se origina, principalmente, por circunstancias de accesibilidad, clima o error cartográfico. La omisión completa de viviendas es otro tipo de falla y se produce tanto por problemas del empadronador (no sabe diferenciar una vivienda de otra, no se percata de la existencia de una vivienda, etc.), como por otras circunstancias (viviendas que casualmente se encontraban sin moradores al momento del censo, etc.). Finalmente está la omisión de personas, lo que puede suceder cuando la información censal es entregada por el jefe de hogar y el empadronador no puede certificar a ciencia cierta si las personas que identifica el informante son todas las que durmieron allí la noche anterior o que residen en la vivienda.

El error de contenido afecta prácticamente a todas las preguntas de la boleta y tiene su origen en diversas etapas de la operación. Puede surgir por una desacertada elaboración del cuestionario (mala re-

⁹ Según Spiegelman, esta situación sería más común en los censos *de jure* que en los censos *de facto* (Spiegelman, 1972).

dacción, inadecuada selección de las preguntas, distribución inconveniente de éstas).¹⁰ También puede presentarse durante la recolección misma de los datos, ya sea por una incompleta preparación de los entrevistadores o por desconocimiento del informante. Por último, es posible que se produzcan errores adicionales en los procesos de codificación o de digitación.

Uno de los errores más comunes en la información censal es el relacionado con la declaración de la edad. Debido a equivocaciones y desconocimiento de los informantes, habitualmente la distribución de la población según edad que registra el censo tiene grupos subrepresentados y otros sobrerrepresentados. Es muy común, por ejemplo, que las edades terminadas en cero o cinco registren un número de personas significativamente mayor que las existentes en las edades inmediatamente antecedentes y consecuentes. De igual modo, existen ciertas edades específicas que ejercen atracción de respuesta, como las edades 7 ó 21. Por diferentes razones, existen otras edades que la gente tiende a rechazar (los 18 años, por el servicio militar; los 13 años, por superstición, etc.).

También se han observado serios problemas en las preguntas sobre números de hijos tenidos, ya que los informantes tienden, con el avance de la edad, a olvidar el número exacto de hijos, sobre todo si algunos han fallecido. Asimismo, a menudo las no respuestas son contabilizadas como mujeres sin hijos o, inversamente, las mujeres que no declaran hijos se registran, por error del empadronador, como no respuesta. Como estos dos últimos errores son usuales, se han desarrollado procedimientos de ajuste basados en ecuaciones de regresión sobre datos empíricos de censos de diferentes lugares del mundo (Naciones Unidas, Manual X). Son comunes los errores en las respuestas a preguntas relacionadas con la actividad económica, lo que en algunos censos prácticamente ha impedido la utilización de los datos recogidos sobre este tema (por ejemplo, el caso del censo de 1980 en México). A causa de lo anterior, algunos censos de la región han incluido una pregunta adicional para las personas que se declaran inactivas para ratificar o rectificar dicha declara-

¹⁰ Por ejemplo, siempre se recomienda, aunque no siempre se cumple, ordenar las preguntas partiendo con las que deben ser contestadas por toda la población y terminando con las que aluden a grupos específicos.

ción inicial (Honduras, 1988; El Salvador 1992 y Nicaragua, 1995).

II.3.c. Errores en las encuestas

En el caso de las encuestas tenemos los errores provenientes del muestreo y aquellos que tienen otro origen. Dentro de los primeros están las deficiencias en el marco muestral, lo que puede significar que algunas unidades de la población que se desea estudiar tengan probabilidad de selección nula, y por lo tanto, se presente omisión.¹¹ Otras causas de errores muestrales pueden encontrarse en un diseño mal definido o en equivocadas reglas de asociación entre unidades de observación y unidades muestrales (hogares y viviendas, por ejemplo). Se ha demostrado que para una población con una tasa bruta de natalidad de 40 por mil, una muestra de 1 000 personas entregaría un intervalo de confianza con un nivel de significación de un 5% —entre 28 y 52 por mil—, mientras que para ese mismo nivel de significación, una muestra de 100 mil personas arrojaría un intervalo de confianza sustancialmente más preciso, concretamente entre 39 y 41 por mil (Tapinos, 1988).

Dentro de los errores de contenido se consideran los mismos comentados en el caso de los censos. Al compararlos con estos últimos, en los de las encuestas se conjugan elementos que tienden a disminuir el riesgo de error de contenido —quienes recogen los datos normalmente están mejor preparados y tienen más experiencia, el tiempo disponible para realizarla es generalmente mayor que el existente para levantar el censo—, con otros que tienden a elevarlo —cuentan con una mayor cantidad de preguntas y éstas son de una complejidad más elevada. En todo caso, los errores en la declaración de la edad y en la respuesta sobre número de hijos tenidos que se detectan en las encuestas, usualmente son de menor magnitud que los registrados en los censos de población.¹²

¹¹ Lo anterior no implica, necesariamente, inexactitudes en la estimación de parámetros, ya que si los individuos omitidos no difieren en sus características de los incluidos en el marco muestral, los resultados de la muestra representarán adecuadamente la realidad del colectivo.

¹² Para un detenido análisis de los errores no muestrales en las encuestas de hogares véase: Naciones Unidas, 1983, *Errores no muestrales en las encuestas de hogares: fuentes, evaluación y control*.

II.3.d. Errores en las estadísticas vitales

Pueden originarse en una falta de recursos del sistema (personal, infraestructura e información), en errores, omisiones de la legislación, en la escasa utilidad de reportar el hecho, en una actitud reacia de la población a certificar este tipo de eventos, en problemas de coordinación dentro de la administración pública y en deficiencias de los informes vitales (o incapacidad de los informantes —o de los encargados— para llenarlos a cabalidad).

La fuente más conocida de error es la omisión o subregistro, originada por la no inscripción de algún hecho vital (nacimiento, defunción, matrimonio, etc.). Antes de seguir adelante conviene distinguir entre "alcance" y "cabalidad". Por alcance se entiende la medida en que el sistema es aplicado a toda la población, ya que en algunos países ciertos grupos de población pueden estar oficialmente eximidos de entregar información al registro de hechos vitales (diplomáticos, militares, aborígenes, grupos segregados, etc.). Por cabalidad se alude a la medida en que la población da cuenta de los hechos vitales. La omisión está ligada con este último concepto y varía significativamente entre los distintos países según el desarrollo alcanzado por el sistema de estadísticas vitales, lo que se vincula con el grado de avance socioeconómico de cada nación.

Esta última afirmación debe ser complementada con dos elementos de juicios adicionales. El primero se refiere a la existencia de países con una actividad histórica de los registros de hechos vitales, los que están más consolidados que sus homólogos en países de semejante nivel de desarrollo socioeconómico. Es el caso de Guatemala en Latinoamérica, ya que desde la conquista fue la sede de los archivos de la Capitanía General, con lo cual se forjó en ese país una infraestructura y una cultura muy favorable para la inscripción de hechos vitales. El segundo tiene relación con la existencia de heterogeneidad en la distribución de la omisión dentro de cada país (Macció, 1992).

Otro error típico de los registros de hechos vitales se relaciona con la oportunidad, y adquiere la forma concreta de inscripción tardía. Es común que la gente registre ciertos hechos vitales, sobre todo los nacimientos, luego de algún periodo de tiempo de ocurrido el hecho, ya sea por ignorancia, por comodidad o por razones de fuerza mayor. Lo anterior puede alterar los indicadores demográficos, sobre todo si la tendencia al registro tardío tiene fluctuaciones mar-

casas año a año. Se ha observado un incremento de la tasa de natalidad derivada de estadísticas vitales que se había calculado para los 5 ó 6 años anteriores, a causa de la inscripción masiva de los niños no registrados anteriormente, en virtud de la promulgación de leyes que obligan a presentar el certificado de nacimiento de los niños para su matrícula en la escuela.

La duplicación de inscripciones es otro problema que afecta a las estadísticas vitales, aunque los avances informáticos y los progresos en la legislación pertinente tienden a reducir drásticamente su ocurrencia. Dos casos típicos de duplicación son la familia que se traslada y vuelve a inscribir al niño en el lugar de destino, y la madre que inscribe al niño en dos o más comunas para recibir beneficios o subsidios que se entregan contra presentación del certificado de nacimiento.

Los errores de contenido son una deficiencia extendida en la información de estadísticas vitales. Estos pueden ir desde la no respuesta hasta la certificación no especializada. En el caso de las defunciones, ocurre a menudo que la certificación no es realizada por un médico y, por tanto, la información sobre la causa del hecho carece de confiabilidad. Asimismo, los problemas con la declaración de la edad suelen ser importantes toda vez que la información es brindada por personas distintas al afectado. También es común el falseamiento en las respuestas sobre lugar de residencia habitual, con el fin de evitar trámites o para obtener atención en algún lugar considerado más favorable. Por último, también debe destacarse la existencia de errores atribuibles a deficiencias administrativas del sistema de registro de hechos vitales, las que pueden reflejarse en que los mismos datos frecuentemente son codificados y tabulados por distintas dependencias gubernamentales, con diferentes criterios y énfasis. Lo anterior puede generar serias discrepancias en los indicadores entregados por las distintas oficinas del gobierno, con las consiguientes confusiones y controversias técnicas entre las diferentes reparticiones estatales.

II.4. Evaluación de la información de fuentes de datos demográficos

II.4.a. Introducción

Por evaluación de la información se entiende "la apreciación en conjunto de la calidad de ella". Los proce-

dimientos de evaluación de la información sociodemográfica pueden agruparse de la siguiente manera:

- i) Comparación de los datos observados con alguna configuración esperada.
- ii) Comparación de los datos observados con los correspondientes de países similares.
- iii) Comparación con datos provenientes de otras fuentes de información demográfica y no demográfica.
- iv) Utilización de la ecuación compensadora y de técnicas indirectas más complejas (poblaciones teóricas).
- v) Evaluación directa a través de nuevas recolecciones de datos (encuesta de reempadronamiento censal).
- vi) Exámenes sobre la coherencia interna de la información.

II.4.b. Procedimientos de evaluación de la información de fuentes de datos demográficos

El primer paso de la evaluación es una acurada y minuciosa observación de los datos, ya sea en los cuadros, gráficas o en los indicadores que se desprendan de los mismos. Por ejemplo, la distribución según edad de la población tiene, normalmente, cierto perfil y solamente mirando las cifras pueden identificarse sesgos en las declaraciones de la edad, tal como se aprecia en el caso de Honduras (1961). En el cuadro II.4.b.i es posible ver que las edades 40 y 45 presentan atracción, lo que va en detrimento de las edades 41 y 44. En el caso de los nacimientos, la existencia de omisión diferencial entre los sexos puede evaluarse según la cantidad de hombres y mujeres registrados como nacidos vivos en un año calendario.¹³

Por otro lado, una observación cuidadosa de los datos es un factor básico para detectar errores como falta de respuestas, o confusión entre No-Responde (NR) y una categoría nula; es el caso, por ejemplo, de una mujer que no responde a la pregunta de cuántos hijos nacidos vivos ha tenido y el encuestador anota erróneamente la respuesta en cero hijos. En definitiva, una esmerada observación de los datos es siempre necesaria y conveniente.

¹³ Como se verá más adelante, existe una distribución esperada bastante exacta de la cantidad de niños que nacen por cada 100 niñas.

CUADRO II.4.b.i
Errores en la declaración censal de la edad.
Contraste entre dos países

Honduras, Censo de 1961: Distribución de la población entre 40 y 45 años cumplidos por años simples y según sexo				Cuba, Censo de 1981: Distribución de la población entre 40 y 45 años cumplidos por años simples y según sexo			
Edad	Total	Hombres	Mujeres	Edad	Total	Hombres	Mujeres
40	29 077	14 085	14 992	40	111 661	56 489	55 172
41	9 642	5 021	4 621	41	119 692	60 368	59 324
42	13 894	7 184	6 710	42	105 474	52 974	52 500
43	11 493	5 813	5 680	43	103 485	51 685	51 800
44	10 337	5 133	5 204	44	97 192	48 729	48 463
45	17 332	8 621	8 711	45	96 349	48 125	48 224

FUENTE: Honduras, Censo Nacional de Población, 1961; Cuba, Censo Nacional de Población, 1981.

Un segundo paso, que siempre debe darse en la evaluación de la información, es el análisis de su coherencia interna. Cuando se tienen los datos originales pueden utilizarse programas computacionales de revisión (mallados de crítica), que relacionan, según reglas lógicas, los datos disponibles y detectan errores (como hombres que declaran hijos tenidos o niños menores de 10 años que aparecen como casados, etc.). Es más, cuando se dispone de los datos originales pueden desarrollarse programas computacionales de asignación que, según reglas determinadas, imputan los valores más probablemente concordantes con la realidad en casos de errores y "no respuestas" en censos y encuestas. En todo caso, lo más conveniente para el usuario es el acceso a los datos originales. Cuando esto no es posible puede buscarse la correspondencia entre los distintos tabulados. Por ejemplo, todos los tabulados según estado conyugal para el conjunto de la población deben arrojar el mismo número de casados, solteros, convivientes y viudos, ya que lo contrario indicaría la presencia de errores o mala presentación de los datos.

Al evaluar los datos disponibles, un tercer paso es su comparación con la información obtenida en situaciones parecidas o con los valores que arrojan otras fuentes, ya sean de índole demográfica o no demográfica. Deben cotejarse los resultados obtenidos con las proyecciones demográficas previas, sin que ello signifique considerar a que éstas últimas son las exactas. Es recomendable comparar los valores censales con los provenientes de estadísticas vitales o de encuestas demográficas disponibles; si

no es posible (o no confiable), puede recurrirse a las encuestas permanentes de hogar, a las encuestas periódicas de empleo, a los registros de consumidores, etc. Finalmente, conviene contrastar los valores obtenidos con los de otras regiones que presenten características semejantes. Es posible suponer subregistro de muertes infantiles porque las tasas de mortalidad infantil derivadas de las estadísticas vitales son demasiado pequeñas en comparación con las existentes en otros países de semejantes características sociodemográficas, pero con registros de hechos vitales con menor omisión de muertes infantiles.¹⁴

Si se dispone de información adicional sobre el hecho o la población que se está midiendo, y se acepta su exactitud, el error de cobertura puede cuantificarse de la siguiente manera:

$$(E - R) / E * 100$$

donde E es el número de personas o hechos vitales estimado correcto y R el número de personas o hechos vitales censados o registrados empíricamente. Un ejemplo de la situación anterior puede ser una

¹⁴ Nótese, sin embargo, que un valor lógico en ciertos indicadores demográficos puede ser engañoso cuando existen errores que se compensan en el numerador y denominador. Así, una tasa de mortalidad infantil "esperable" a la luz de los valores existentes en países de semejantes condiciones sociodemográficas, puede ocultar una elevada omisión en el registro de muertes infantiles y, a la vez, una omisión parecida en el registro de nacimientos.

encuesta nacional, efectuada con los más rigurosos estándares de calidad, que arroja una estimación de los nacimientos ocurridos durante el año calendario previo a su realización. Esta cifra sería E, y permitiría calcular el porcentaje de cobertura de las estadísticas vitales sobre nacimientos, ya que R sería el número de nacimientos registrados en el sistema de estadísticas vitales. Otro ejemplo estaría en el cálculo de la cobertura censal, considerando a E como la cantidad de personas que según las proyecciones demográficas existió en el momento del censo y a R como los montos censales.

El valor obtenido es el porcentaje de cobertura del censo o de las estadísticas vitales. Si el valor —de E-R— es positivo implica subenumeración y si el valor es negativo implica sobreenumeración o sobreregistro.

Un cuarto paso, algo más complicado, está en la utilización de la ecuación compensadora, la fórmula básica del cambio de población. En términos generales, este paso implica comparar las cifras demográficas en un momento definido como final, con aquellas que se desprenden del efecto, sobre la población existente en el momento estipulado como inicial, del balance de los componentes de la dinámica de la población entre el instante inicial y el final.

Adicionalmente, los demógrafos han generado una gran cantidad de procedimientos, basados tanto en la ecuación compensadora como en poblaciones modelos del tipo estable, que, de manera indirecta, permiten detectar errores de cobertura o sesgo en la información (Brass, 1974; Chackiel y Macció, 1979; Keyfitz, 1979). Por último, siempre cabe la posibilidad de una evaluación directa de los datos por medio de reentrevistas o recogiendo nuevamente la información.

Luego de la evaluación de la información, y ante la eventual presencia de errores, hay varias formas de corrección, algunas de las cuales ya han sido señaladas. Se mencionaron los programas de crítica y asignación, que permiten encontrar y superar, bajo ciertos supuestos, errores de contenido. Las soluciones que entregan los programas de asignación se basan en probabilidades y, por tanto, pueden no ser acertadas. Además, no son utilizables cuando la cantidad de errores o de "no responde" es un porcentaje alto del total de respuestas. También se señaló la posibilidad de recoger nuevamente la información, que es una manera de detectar errores de cobertura

y solucionar algunos errores de contenido (llenar preguntas sin respuesta, resolver respuestas incoherentes, etc.).

Por ejemplo, una solución al problema de la omisión está en encontrar un factor de expansión que lleve los resultados a su verdadera magnitud. Para esto se requiere ubicar el porcentaje de omisión, siguiendo la fórmula planteada anteriormente, y aplicarlo a los valores recogidos.¹⁵ Por ejemplo, si un censo ha arrojado una cifra de 10 millones de habitantes y se considera que la omisión censal fue de un 50 por ciento, el problema de subenumeración se resuelve, a nivel nacional, multiplicando la cifra obtenida (10 millones) por 2. Sin embargo, esta solución es demasiado simple, porque a nivel de subpoblaciones (grupos de edad, condición urbano-rural, sexo, estratos sociales), tal como se ha señalado insistentemente, la omisión no se distribuye homogéneamente y, por tanto, si se aplicara un factor de corrección único (de 2) para todos los subgrupos señalados, los datos podrían quedar más sesgados que en su presentación original. Por cierto, si se conoce la omisión en cada subgrupo, lo que es difícil de lograr, puede utilizarse este método para resolver el problema de la subenumeración.

Técnicas más elaboradas de corrección de información proveniente de las fuentes de datos demográficos han sido desarrolladas para la solución de los errores de la declaración de la edad. Éstas se basan en el carácter típico de tales errores. En efecto se han detectado tres tipos fundamentales de error: la preferencia de dígitos (0 y 5 habitualmente); el traslado de edades (por ejemplo, a la gente muy anciana se le atribuyen 100 años de edad); y la preferencia o rechazo de edades específicas (por ejemplo, 60 y 13 años, respectivamente). Los especialistas en análisis demográfico deben recurrir a índices para detectar errores de la información y a procedimientos para corregirlos. Sin embargo su uso no es muy frecuente en otros ámbitos de las ciencias sociales.

¹⁵ Debe destacarse que el cálculo del porcentaje de omisión cambia según el valor que se considere como correcto, el que no siempre está disponible y sobre cuya exactitud normalmente no existe certeza total. Debido a lo anterior, es prácticamente imposible encontrar el porcentaje real de subenumeración y en la práctica se trabaja con una estimación.