D- 02945.00

Centro Latinoamericano de Demografía

CELADE

BIBLIOTECA
SEP 2009
NACIONES UNIDAS
SANTIAGO
CHILE

EVALUACION Y CORRECCION DE DATOS DEMOGRAFICOS

Introducción
I. Conceptos básicos
II. El error en las estadísticas demográficas

Juan Chackiel y Guillermo Macció

Serie B, N° 39
Agosto, 1978
Santiago de Chile

900035849 - BIBLIOTECA CEPAL

NACIONES UNIDAS
BIBLIOTECA

Este documento forma parte de un texto más amplio, constituido por 10 capítulos, cuyo título es Evaluación y Corrección de Datos Demográficos.

La publicación independiente de estos dos capítulos y de otros que aparecerán posteriormente, se hace con la finalidad de utilizarlos como material docente.

Los autores agradecen, desde ya, los comentarios y sugerencias que les puedan hacer llegar, pues serán de gran utilidad para la versión definitiva.

I N D I C E

	Página
INTRODUCCION	1
I. CONCEPTOS BASICOS	7
1. Evaluación de datos demográficos	7
2. Definición de otros conceptos	8
3. Principios básicos que deben considerarse en el manejo de los datos	9
II. EL ERROR EN LAS ESTADISTICAS DEMOGRAFICAS	12
1. Error de cobertura	13
2. La medición de los errores de cobertura	15
3. Error de contenido	16
4. Las fuentes de errores	20

INTRODUCCION

Una revisión de la literatura sobre métodos y técnicas de análisis demográfico producida en los países desarrollados -en particular anglosajones y franceses- pone en evidencia que los temas relativos a evaluación, corrección y ajuste de datos han recibido y reciben, aún hoy, escasa atención. Este hecho no puede causar extrañeza ya que para los investigadores de esas realidades, la calidad de la información básica que sirve de materia prima en su labor no llega a constituir motivo de inquietud o duda. Los datos a su alcance ostentan un grado de cobertura, periodicidad y confiabilidad que en la mayoría de los casos puede hacerse uso de ellos tal como son presentados. Otra consecuencia de esa realidad, además de la explicable resistencia de los demógrafos a introducir correcciones o ajustes, es la poca importancia que se concede al asunto en los programas de enseñanza de demografía donde, a veces, ni siquiera figura o es tratado sólo muy superficialmente.

Muy otra es, en cambio, la situación de los países de América Latina donde, como es bien sabido, los datos demográficos son escasos y en su mayoría deficientes; cuadro que por lo demás vale también para la mayor parte de Africa y Asia. Sucede, sin embargo, que a pesar de los males comunes que aquejan a estas tres regiones, América Latina en su conjunto presenta una serie de peculiaridades que, como en otros campos, *la sitúan a mitad de camino* entre los países más avanzados y aquéllos que se ubican en los niveles más bajos medidos en los términos usuales de desarrollo. Esta circunstancia tiene una clara repercusión sobre los temas que se tratan en este trabajo; conviene entonces retenerla desde un principio, porque en cierta forma determina o da la pauta al perfil actual de los asuntos aquí tratados.

Se dijo antes que en general los datos demográficos son escasos y deficientes, pero no son tan pocos ni tan malos como para desecharlos o basar la investigación en modelos sustitutivos o analogías. Esta particularidad que

se acentúa como una consecuencia, muy favorable por cierto, del continuo proceso de acumulación de nuevos y mejores datos censales y de encuestas levantadas en fechas recientes, confiere a la investigación que en el presente se lleva a cabo, una peculiaridad adicional.

Dentro de ese contexto, el trabajo del demógrafo se ve en la necesidad de pasar, en la gran mayoría de los casos, por una etapa previa destinada a detectar, juzgar y desentrañar, interpretándolas, las debilidades que afectan sus datos básicos y eventualmente introducirle correcciones.

Con el correr del tiempo se han ido acumulando una serie de técnicas destinadas a enfrentar los problemas antes señalados. Hoy, su repertorio es bastante extenso, tanto como para justificar un cuerpo de conocimientos que tiene perfiles bastante definidos y que cobra una cierta identidad dentro de la demografía. Su evolución y desarrollo no ha seguido un camino rectilíneo todo lo contrario; se nutre sobre todo de éxitos y fracasos, de aproximaciones empíricas destinadas fundamentalmente a resolver problemas muy concretos en el tratamiento de información. No debe extrañar entonces que, dentro del bagaje de técnicas y artificios al alcance del demógrafo, subsistan recursos con muy diverso grado de eficiencia. No es ni más ni menos que el resultado del propio proceso, esencialmente histórico y acumulativo, traducido en respuestas a problemas muy específicos. Sólo en fechas recientes ha cobrado fuerza, probablemente como consecuencia de ese mismo proceso de acumulación de conocimientos y de maduración de ideas en torno al tema, el interés y aún la necesidad por dotar a la evaluación y corrección de datos de una base teórica más general que permita -entre otras cosas- sistematizarla desde un punto de vista conceptual, jerarquizar sus técnicas y ayudar al demógrafo en la tarea de detección de errores o deficiencias para interpretarlos y corregirlos.

Se piensa que este proceso de conceptualización -por llamarlo de alguna manera- puede tener repercusiones beneficiosas en el papel del demógrafo y su postura frente a la realidad que desea desentrañar. Hasta ahora, por lo general, el investigador ha jugado y juega un papel pasivo, resignado, frente a la calidad de los datos básicos; son elementos que le vienen dados. Poco o nada puede hacer para cambiar su naturaleza; a lo más, desarrollar el

ingenio, instrumentos de análisis y artificios para discernir entre lo erróneo y lo cierto. Se piensa, sin embargo, que tarde o temprano deberá producirse un cambio, de suerte que el demógrafo pase a jugar un papel mucho más activo en el proceso de generación de esos datos. Los esfuerzos que se hagan para poner en evidencia la importancia que dentro del análisis demográfico reviste la etapa de evaluación y corrección, contribuirán sin duda a acercar esta participación más activa.

Otro asunto sobre el que conviene hacer aquí aunque sólo sea una mención, es el relacionado con la actitud del demógrafo frente a los datos y al uso del instrumental que tiene a su disposición para juzgarlos y, llegado el caso, corregirlos. Dos parecen ser los riesgos principales que acechan su quehacer: por un lado, el de la ingenuidad; por otro, el del escepticismo. Ingenuidad en la aplicación rígida o mecánica de técnicas sin adoptar las necesarias precauciones respecto de las limitaciones, supuestos y, en ocasiones, sesgos que subyacen en todas ellas; ingenuidad también expresada en exagerado optimismo respecto de la bondad de los métodos, niveles de tolerancia de los supuestos, regularidades obtenidas, a veces meramente visuales. Ingenuidad, por último, en creer que las posibilidades de análisis se agotan con la aplicación de unas pocas técnicas. Este riesgo ha sido, en no pocos casos el origen de conclusiones falsas o analogías sólo válidas por la coincidencia numérica, Escepticismo, frente a la posibilidad de desentrañar una determinada realidad a partir de información fragmentaria, distorsianada y, desde luego, incompleta. Actitud que conduce a restringir las posibilidades de estudio y a veces hasta una postura agnóstica o cómoda. Justamente la aplicación del principio, de rehabilitación y rescate de los datos ha contribuido al desarrollo de nuevas técnicas cuyo uso permite hoy desentrañar, a partir de datos que hasta hace poco habían sido sub-utilizados o simplemente descartados, facetas o características hasta ese momento desconocidas.

Una advertencia necesaria, muy ligada a la aplicación de técnicas de evaluación y corrección de datos tiene que ver con la permanente actitud crítica del investigador frente a las herramientas de que dispone y a su forma de aplicarlas. No existe un método único capaz de proveer el mejor ajuste ni hay una técnica que reúna las cualidades de una receta universal. Por el

contrario, el trabajo de corrección y evaluación se rige por los intentos sucesivos y las pruebas alternativas. La decisión final sobre lo que usualmente se denomina el "error más probable" o "el ajuste más plausible", reside, en última instancia, en el juicio y la perspicacia del investigador. Por eso, el éxito de esta etapa depende fundamentalmente de la actitud crítica que se mantenga y del grado de discernimiento que se aplique tanto frente a las debilidades y virtudes del arsenal en uso como al significado en la naturaleza y magnitud de los errores que se detecten.

Las reflexiones precedentes constituyen, en cierto modo, los fundamentos sobre los cuales se apoyan los capítulos que siguen. La primera dificultad ha sido la de establecer un orden coherente y más o menos lógico para el desarrollo de los distintos temas. Otro escollo viene dado por la necesidad de seleccionar y jerarquizar las distintas técnicas de examen y corrección. Ambos, el orden y la jerarquización, constituyen apenas un modesto intento que de seguro podrá mejorarse. En este ejercicio -conviene destacar una vez más-, la atención se ha centrado en el estado actual de la investigación demográfica y en el tratamiento que requieren los datos latinoamericanos.

El trabajo se compone de diez capítulos. Los dos primeros están destinados a examinar aspectos teóricos en torno al error que afecta a la información. El tercero se concentra en las comúnmente denominadas evaluaciones directas, es decir, aquellas operaciones que van hasta la instancia misma de generación del dato. Los siguientes se concentran en las técnicas usualmente llamadas de evaluación indirecta que, a diferencia de las precedentes buscan poner en evidencia las imperfecciones mediante el estudio de la consistencia interna de la información y la contraposición de datos de distinto origen o recolectados en diversos momentos. A través de esta línea de trabajo, la demografía ha alcanzado sus éxitos más recientes.

Por la importancia que reviste el examen de la variable edad en todo trabajo demográfico y por la abundancia de esfuerzos que se han hecho para controlar la bondad de la información que sobre ellas se recoge, los capítulos VI y VII se han destinado en exclusividad a este asunto. El que le sigue recoge una serie de modelos teóricos de población, en particular los estables. Aun cuando, como es bien sabido, las poblaciones latinoamericanas se han

ido apartando de las condiciones de estabilidad, el uso de estos modelos y de las relaciones que de ellos se derivan, constituyen instrumentos útiles para fines de evaluación.

El capítulo IX está dedicado a las técnicas de estimación y corrección a través de información retrospectiva recogida en censos y encuestas. El tema trasciende los límites de la evaluación y corrección de datos propiamente dicho; se insertaría mejor en un capítulo más general de estimaciones. No obstante su aporte como fuente independiente de los registros, el estrecho parentesco entre ambos asuntos y la importancia que para el caso latinoamericano revisten estas técnicas, sobre todo en los últimos años, parecen argumentos suficientes para justificar su inclusión aquí.

El capítulo X y final está destinado a presentar algunas técnicas de uso frecuente en el ajuste de funciones demográficas.

Por último, es conveniente insistir en que este trabajo no debe ser considerado en ningún caso como un recetario o un manual donde el investigador encuentra respuestas directas a sus problemas de análisis. Pretende ordenar y darle un perfil a un cuerpo de conocimientos que se desarrolla y crece con rapidez. Tiene la intención principal de servir como guía y ayudar al desarrollo de la actitud crítica del demógrafo tanto frente a los datos que constituyen la mediación entre la realidad y la posibilidad de su conocimiento como al uso apropiado de las técnicas que están a su alcance.

I. CONCEPTOS BASICOS

1. Evaluación de datos demográficos

En una primera apreciación, el concepto de evaluación parece ser muy obvio, y quizás lo sea, pero como se ha creado cierta confusión acerca de su significado, es conveniente precisar qué significa ese término en el contexto de esta materia.

Probablemente la confusión mencionada provenga de la definición adoptada por las Naciones Unidas en su Diccionario Demográfico Plurilingüe[1/] en que dice textualmente: "Aunque la palabra evaluación suele emplearse como sinónimo de estimación, se aplica más bien a valores puramente conjeturales, que no están basados en ningún dato fidedigno, y sirve sólo para indicar el orden de magnitud del valor de que se trata". Aunque esta definición pretenda desligarse del concepto de "estimación" no lo logra, sino que simplemente hace referencia a estimaciones más burdas, de menor confiabilidad.

En realidad, la evaluación es algo mucho más amplio, es una operación mucho más compleja que pretende dar un juicio sobre cuán bueno o malo es el dato que se está examinando, es un intento por detectar los errores que afectan a la información, la naturaleza y la magnitud de los mismos. Para ello la información disponible es sometida a examen mediante técnicas muy variadas, pero el dato en sí no es afectado, no se modifica, se trata solamente de derivar juicios o indicadores sobre su calidad.

Esta operación puede hacerse de dos maneras denominadas comúnmente "evaluación directa" y "evaluación indirecta", que pueden aplicarse con independencia una de la otra, pero que por lo general son complementarias.

1/ Naciones Unidas, Diccionario Demográfico Plurilingüe, ST/SOA/Ser. A/29, Nueva York, 1959.
Esta definición subsiste en la nueva versión del Diccionario, actualmente en prensa.

La evaluación directa es un procedimiento que consiste en volver al origen del dato, lleva implícito una vuelta al terreno. Se trata de reproducir nuevamente, bajo condicicnes completamente controladas, el mecanismo de obtención de la información. El nuevo dato, así obtenido, constituye el elemento de comparación con el dato original sometido a evaluación. Por ejemplo, si se trata de evaluar un censo de población, se podría volver al terreno y realizar, en algunas áreas predeterminadas, en forma muy cuidadosa y con personal altamente calificado, un nuevo censo a efectos de establecer cuán confiable ha sido el empadronamiento censal original.

En cambio, la evaluación indirecta, se refiere a aquellos análisis a los cuales se someten los datos disponibles, mediante relaciones cuyo comportamiento se supone conocido, e información colateral, con el fin de establecer la naturaleza y el grado de error que los afectan. Podría decirse que es un trabajo esencialmente de "escritorio". Es en el contexto latinoamericano actual un quehacer fundamental del análisis demográfico y a él está destinado la casi totalidad de este texto.

2. Definición de otros conceptos

Una vez cumplida la etapa de evaluación, esto es, emitido un juicio sobre la calidad de los datos, frecuentemente le sigue una serie de operaciones para las cuales no existe una definición clara y aceptada universalmente (corregir, ajustar, suavizar, disciplinar).

Cuando se habla de corregir, ajustar y suavizar, ya se está en disposición de afectar el dato con la intención de mejorarlo o hacerlo más coherente. Estas tres expresiones están muy ligadas entre sí y, aunque es difícil precisarlas, parece conveniente aclarar los matices que las diferencian.

Para el término ajustar o ajuste se tienen dos definiciones más o menos similares. El Diccionario Demográfico Plurilingüe[2/] dice al respecto: "Muchas veces es necesario sustituir una serie de datos observados por otra

serie más regular, llamada serie ajustada. Un método de ajuste consiste en hacer pasar una curva regular por entre los puntos que representen la serie dada". El Diccionario de Términos Estadísticos[3/] tiene como una de las acepciones, la siguiente: "... para denotar el ajustamiento de una curva matemática a cualquier tipo de datos estadísticos susceptibles de ser representados en correspondencia con tiempos o espacios variables, e.g. datos de regresión o series cronológicas".

Se entenderá por suavizamiento el eliminar fluctuaciones bruscas sin afectar mayormente la información. La diferencia con el ajuste está en la intensidad de la corrección aplicada. Kendall y Buckland en el Diccionario ya citado expresan que es el "Proceso de eliminación de las fluctuaciones de una serie ordenada de modo que la serie resultante sea "suave" en el sentido que las primeras diferencias sean regulares y las de órdenes superiores pequeñas ...". "De hecho el concepto está estrechamente ligado al de ajustamiento de una curva de tendencia."

Corrección sería el término más general que engloba a los dos anteriores y aún más. En este concepto se incluye todo tipo de modificación que se introduce en los datos, a los efectos de eliminar errores e irregularidades. En este sentido se pueden mencionar los ajustes, suavizamientos, pero también, por ejemplo, la modificación del total de la población por haberse detectado un faltante de personas en el relevamiento censal. De aquí la justificación al título que se ha dado a este documento.

3. Principios básicos que deben considerarse en el manejo de los datos

Basándose en su propia experiencia, Brass[4/] sugiere que se tomen en consideración una serie de principios generales de gran utilidad en el análisis demográfico y, en particular, cuando se procede a realizar la evaluación y la corrección de los datos. A continuación se hace referencia a cada uno de estos principios:

3/ Kendall, M. y Buckland, W., Diccionario de Términos Estadísticos, IASI, Rosario, Argentina, 1959.

4/ Brass, W., Seminario sobre Métodos para Medir Variables Demográficas, 24 de setiembre de 1971, CELADE, Serie DS, N° 9, San José, Costa Rica, 1973.

a) Serendipity.[5/] Se trata de un concepto usado en otras ciencias, en particular en la sociología. Al respecto el Webster's Third New International Dictionary[6/] expresa que es una supuesta capacidad para encontrar cosas valiosas o aceptables, no buscadas. En esto juega un papel importante la imaginación y la suerte del investigador. Pueden hallarse relaciones o analogías que no se buscaban y que resultan útiles. También ocurre que en busca de cierta información se encuentra a veces otra que posibilita una nueva dirección o nuevos hallazgos en la investigación. Por ejemplo, supongamos que averiguando datos de estadísticas vitales en un estante de una biblioteca, con el fin de tratar de realizar estimaciones de fecundidad en el pasado, se encuentra por casualidad información confiable, brindada por una encuesta realizada en esa época, y a través de la cual se llega a nuevas estimaciones no contempladas en el plan de trabajo original.

b) Rehabilitación. Este concepto lo usa Brass por analogía con la terminología médica. Cuando se corrige la información, debe hacérselo en su justa medida, no más allá de lo que requiere. Debe tratarse de respetar al máximo la información observada, pues de lo contrario se corre el riesgo de sobrecorregirla e introducir errores mayores a los que tenía originalmente e incluso alterar el significado de los datos. Brass da un ejemplo tomado de la medicina: el de una persona herida en una pierna que probablemente puede recuperarse más rápido si se le auxilia con un bastón en lugar de usar una silla de ruedas. Es más, esta última terapia podría resultar, por otros efectos, contraproducente para su recuperación definitiva.

c) Consistencia o coherencia. Según este principio corresponde exigir coherencia tanto en los procedimientos utilizados como en los resultados que se obtienen. Un ejemplo de coherencia de procedimientos sería: adoptado un método de estimación de una variable demográfica que suponga que la fecundidad es constante en el tiempo, no debería usarse otro método basado en un

---

5/ En ausencia de un vocablo español de aceptación general, se ha preferido mantener el término en inglés, no obstante en algunos textos aparece traducido como "chiripa".

6/ Gove, Philip Badcock (Ed.), Webster's Third New International Dictionary G. y C. Merriam Company, Massachusetts, 1966.

supuesto descenso pronunciado de la fecundidad. En cuanto a los resultados, es de esperar que haya consistencia con otros resultados conocidos. Por ejemplo, es difícil aceptar una tasa bruta de natalidad superior al 45 por mil en un país con una población envejecida donde los menores de 15 años de edad representan menos del 30 por ciento de la población total.

d) Robustez. En general, los métodos para derivar estimaciones demográficas descansan en determinados supuestos, cuya verificación en situaciones concretas condiciona la validez de las estimaciones. Se dice que el método es robusto cuando los resultados no se alteran demasiado por el no cumplimiento de esos supuestos a la hora de aplicarlo en casos reales. En otras palabras, la validez de los resultados no se resienten seriamente como consecuencia del no cumplimiento de las hipótesis de trabajo.

e) Ausencia de norma estricta (No rule). Este principio sostiene que la aplicación de artificios o técnicas para evaluar y corregir deben convertirse en rutina. Por el contrario, cada caso debe tratarse en forma independiente de acuerdo a la información disponible, a la calidad de ésta y a la situación concreta que se está analizando. Puede decirse que no hay método que sirva en todas las circunstancias y tampoco hay un solo método para aplicar. En este sentido lo adecuado es usar varios métodos, razonablemente confiables para la situación que se maneja, y así darse un repertorio probable de estimaciones, de las cuales se seleccionará la que con buen juicio parece ser la más correcta. Este principio refuerza la postura crítica del investigador tanto frente a los distintos métodos disponibles como a los resultados obtenidos.

f) A los cinco principios de Brass, que denominamos así por haber sido él quien los explicitara y los presentará sistemáticamente, debe agregarse el conocimiento de las circunstancias históricas y culturales del país al cual pertenecen los datos. Hay múltiples ejemplos de errores en la información, derivados de hechos históricos circunstanciales a los cuales es necesario recurrir para explicar ciertos resultados. Un censo puede presentar una fuerte omisión si se levanta en momentos de violencia en un país, dado que produce una tendencia de la población a no colaborar por temor a que el censo esté

vinculado con esa situación. Es posible que la sanción de una nueva ley sobre asignaciones familiares conduzca a un aumento en las inscripciones de los nacimientos para acogerse a este beneficio, lo que puede desembocar, si no se tiene en cuenta esto, en estimaciones erróneas de la tendencia de la fecundidad.

II. EL ERROR EN LAS ESTADISTICAS DEMOGRAFICAS

La demografía, como disciplina social, se diferencia, entre otras cosas, de las comúnmente llamadas ciencias naturales y exactas en que la información que constituye su materia prima adolece de errores de distinto tipo. En los estudios de población no puede experimentarse en laboratorios con un control total de las variables, como sucede en la física o en la química. Los componentes demográficos deben estimarse a partir de datos recogidos mediante censos, registros o encuestas, operaciones éstas que están sujetas a irregularidades de diversa índole que afectan la calidad final de los datos. Aún más, esos datos en sí mismos proveen una imagen parcial y a veces sesgada de los fenómenos que hacen el objeto de la demografía.

La magnitud de los errores en la información está directamente vinculada al grado de desarrollo de las estadísticas demográficas, lo que a su vez depende de la situación socio-económica y cultural de los países. En América Latina, en particular, se encuentra un amplio espectro en cuanto a la calidad de las estadísticas, pero con un fuerte predominio de países con datos que pueden calificarse como insuficientes.

Se puede hablar fundamentalmente de dos tipos de errores, a los que Spiegelman[7/] llama de cobertura y de contenido. A éstos corresponde agregar los errores propios del muestreo, cuando se trata de datos recolectados mediante encuestas.

7/ Spiegelman, M., Introducción a la Demografía, Fondo de Cultura Económica, México, 1972

1. Error de Cobertura

Es el caso en que las personas o los hechos que se desean investigar no se registran o se registran más de una vez. Esto conduce a errores por defecto o por exceso. En la práctica, lo más común es que ocurra lo primero, es decir que los casos registrados sean menores que los reales. Se trata, pues, de un concepto esencialmente cuantitativo.

En las estadísticas vitales lo común es que los nacimientos y las defunciones registrados sean menores a los ocurridos. A esto se le suele llamar subregistro. Cuando los casos registrados superan a los ocurridos se habla sobreregistro. Estos aspectos están ligados directamente a los conceptos de alcance y de cabalidad de las estadísticas vitales.

El alcance de los registros hace referencia a "la medida en que el sistema de inscripción es aplicado a toda la población".[8/] Es posible que por definición el sistema de recolección de los hechos vitales no se extienda a toda la población o a todo el territorio de un país. Por ejemplo, supóngase que por dificultades de acceso, una provincia de determinado país no cuenta con oficina de registro, por lo tanto, el alcance está limitado y se produciría un subregistro de los nacimientos y defunciones del total del país.

La cabalidad es"la medida en que se da cuenta de todos los nacimientos y defunciones allí donde se aplica el sistema de inscripción ...."[9/]

El sobreregistro es más frecuente entre los nacimientos. En algunos países es probable que para atender determinados trámites que exigen certificado de nacimiento, resulte más cómodo para los interesados volver a inscribir a un niño, sobre todo si se está en una provincia alejada de la que corresponde a la primera inscripción.

En el caso de las encuestas y de los censos de población, los errores de cobertura son llamados comúnmente omisión y sobreenumeración.

8/ Naciones Unidas, Métodos para Evaluar la Calidad de los Datos Básicos Destinados a los Cálculos de la Población, Manual II, ST/SOA/Serie A/ Nº 23, Nueva York, 1955.

9/ Naciones Unidas, "Métodos para Evaluar ..." op. cit., pág. 20.

Se dice que hay omisión cuando un individuo no ha sido captado por el censo o encuesta. Este término muchas veces se hace extensivo a los registros como sinónimo de subregistro.

Es posible considerar distintos tipos de omisión:

a) Omisión completa de áreas geográficas. Por diversas razones en las operaciones de relevamiento censal, a veces queda fuera un área geográfica o localidad determinada y, por lo tanto, no se empadronan las personas que allí viven. Esto en general sucede por muy diversas razones, pero conviene destacar los obstáculos geográficos,o deficiencias en la cartografía censal y circunstancias locales. En este caso el error no afecta tanto la estructura de los resultados censales, aunque sí el total, debido a que se excluirían hombres, niños y mujeres de todas las edades y condiciones. Por el contrario, cuando se trata de un área donde las características de su población difieren mucho del resto del país (áreas mineras, zonas de migración diferenciadas, etc.), este tipo de omisiones tiene consecuencias más graves.

b) Omisión completa de viviendas. En este caso se empadrona el segmento censal correspondiente, pero quedan sin enumerar una vivienda o más. En áreas urbanas, ocurre con frecuencia que los empadronadores omiten el registro de viviendas que por su ubicación resultan poco visibles o es difícil identificarlas como tales. Esto se da sobre todo en barrios de alta concentración de viviendas precarias. Otro caso es el de habitantes ausentes de su casa durante el día, a la que el empadronador censal después de varios intentos fallidos decide no regresar. Un ejemplo de esto se refiere a las personas que viven solas en áreas rurales y que durante el día no se encuentran en su casa por estar dedicadas a labores agrícolas en el campo.

El efecto que este tipo de omisiones tiene sobre la estructura de la población depende en esencia de las características demográficas y socio-económicas de las familias excluidas.

c) Omisión de personas aisladas. En este caso se empadrona la vivienda, pero quedan una persona o más no captadas por el empadronamiento. Es lo que ocurre comúnmente en América Latina con los niños menores de cinco años de edad que no son declarados por el informante, lo cual conduce a un faltante notable en ese

grupo de edades. Basándose en la experiencia, este tipo de error es el que tiene un mayor peso sobre la calidad de los censos, derivado precisamente de las características diferenciales de los individuos omitidos.

El problema de la omisión puede enfocarse desde otro punto de vista, considerando aquellos sectores específicos de la población más expuestos a ser omitidos: niños de corta edad, trabajadores en la agricultura, personas que viven solas, hombres solteros, etc. Desde luego estos sectores necesariamente serán omitidos por alguna de las formas vistas anteriormente.

La sobreenumeración censal es el caso opuesto a la omisión, es decir, ocurre cuando una persona es empadronada más de una vez. Esto le puede suceder a muchos individuos, pero en el balance general se espera normalmente que el porcentaje de omisión sea superior al de sobreenumeración.

Una circunstancia que puede conducir a un doble empadronamiento es cuando el levantamiento del censo se prolonga por varios días. Si una persona se traslada en ese período de un lugar a otro del país, es probable que sea censado dos veces, aunque por supuesto una de ellas es incorrecta. Otro caso puede darse en los censos de derecho, en que una persona es correctamente censada en el lugar de residencia habitual, pero además se lo empadrona indebidamente en el lugar donde se encuentra presente. Este hecho adquiere relevancia cuando existen sectores importantes de la población en esta situación y el error se comete sistemáticamente. Por ejemplo, en época de zafras, en ciertos países, los trabajadores se trasladan en masa de una región a otra y es probable que si el censo se realiza en ese momento, se produzca la sobrenumeración debido a una mala interpretación del concepto de residencia habitual.

2. La medición de los errores de cobertura

Los errores de cobertura se miden, en general, de la siguiente manera:

$$\frac{E - R}{E} \; 100$$

siendo: E = número estimado como correcto (población o hechos vitales)

R = número censado (o registrado) sin corrección

A veces se calcula este porcentaje sobre el valor registrado, en lugar de sobre el estimado:

$$\frac{E - R}{R} \quad 100$$

Es necesario, para evitar confusiones, aclarar siempre si el porcentaje de omisión o subregistro se ha calculado sobre el valor estimado o sobre el valor censado (o registrado).

En los casos de sobreenumeración y sobreregistro, R, será mayor que E y, por lo tanto, el porcentaje será negativo.

La forma de determinar el valor estimado como hipotéticamente correcto (E) es parte del trabajo de evaluación y, con este fin, serán estudiados varios métodos en los capítulos restantes.

Como se mencionó anteriormente, el cuadro general de los países de América Latina indica la existencia de omisión en los censos y subregistro de los hechos vitales. La omisión censal es fluctuante, pero según las estimaciones realizadas varía generalmente entre el 4 y el 10 por ciento del valor estimado. Más heterogéneo y más grave es el panorama de las estadísticas vitales donde se encuentran países con registros considerados razonablemente completos (con menos del 5 por ciento de subregistros) hasta otros en donde prácticamente no se dispone de sistemas de estadísticas vitales o éstas son de alcance muy limitado, pudiendo establecerse subregistros superiores al 50 por ciento y aún más.

3. Error de contenido

En este caso el individuo y sus datos han sido incluidos, pero en forma incorrecta. Los errores de este tipo son muy diversos; entre ellos se destacan los derivados de la mala declaración por parte del informante. Así, por ejemplo, se declara una edad que no es la correcta o un estado civil que

no es el verdadero. Se trata esencialmente de un error de tipo cualitativo: se está falseando una característica de la persona.

Este tipo de deficiencias en los datos adquiere importancia cuando se produce sistemáticamente, conduciendo a un sesgo en los resultados. Si, por ejemplo, las mujeres que tienen entre 10 y 15 años declaran sistemáticamente una edad mayor, este hecho producirá no sólo un faltante en el grupo 10-14, sino que, además, tenderá a aumentar la edad media de la población femenina. El error de contenido, generalmente, altera el conocimiento de las estructuras demográficas de la población.

No es fácil detectar estos errores y menos aún cuantificarlos, aunque existe cierta metodología destinada a esos fines. Tampoco se sabe mucho acerca del comportamiento y de la distribución de esos errores. En ese sentido más adelante se hará hincapié fundamentalmente en aquellos que inciden en la estructura por edades de la población.

Spiegelman, en su libro Introducción a la Demografía,[10/] presenta una serie de formas de medición de los errores de contenido de las estadísticas demográficas (censos, encuestas, registros), bajo los siguientes supuestos:

a) Los censos, encuestas o registros son una prueba en una serie de recolecciones de pruebas hipotéticas hechas al azar bajo las mismas condiciones e independientes entre sí.

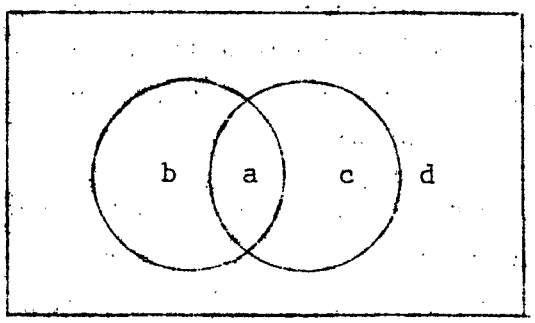
b) Es posible contar con registros o enumeraciones ideales que no contienen error alguno. Este es quizás el trabajo más difícil del demógrafo: encontrar información lo más confiable posible que sirva como estándar, con el cual se contrastan los datos reales.

Considérese ahora un atributo cualquiera que se declara en un censo, como por ejemplo el estado civil y, dentro de él, a la "clase" de los solteros. Se puede confeccionar, con esta base, el cuadro que se muestra a continuación.

10/ Spiegelman, Mortimer, Introducción a la ...., op. cit., pág. 63.

Recuento ideal	Recuento censal		
	Número de personas en la clase	Número de personas que no están en la clase	Total
Número de personas en la clase	a	b	a+b
Número de personas que no están en la clase	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

que se ilustra así:



a = Es el número de personas que según el censo y el recuento ideal pertenecen a la clase, es decir son solteras y no se ha cometido error.

b = Número de personas que según el recuento ideal son solteras, pero que en el censo no aparecen como tales.

a+b= Es el número total de solteros según el recuento ideal.

c = Son los que han sido censados como solteros pero que según el recuento ideal no lo son.

a+c= Es el total de personas censadas como solteras.

d = Son las personas que no pertenecen a la clase según los dos recuentos.

n = Total de personas, es decir el universo considerado.

Con estos elementos es posible definir un conjunto de medidas, muchas de las cuales se usan corrientemente sin tener en realidad conciencia de su fundamento teórico.

a) Tasa de diferencia neta. La diferencia neta entre el recuento real y el ideal se define como: (a+c) - (a+b) = c-b y la tasa de diferencia neta:

$$\frac{c-b}{n} \; 100$$

Esta tasa mide la importancia del error censal con respecto a (a+c), es decir a los solteros censados. Si la tasa es mayor que cero quiere decir que en el censo se han sobreestimado las personas que pertenecen a esa clase y un resultado negativo expresaría la situación inversa. En este caso se introduce un sesgo por compensación entre los valores de b y c, debido a que en ambos casos se está en presencia de errores, pero con sentido contrario.

b) Tasa de diferencia bruta. Se define de la siguiente manera:

$$\frac{b+c}{n} \; 100$$

En este caso no existe compensación pues se suman los valores de los que no están en la clase y debieran estar, más aquellos que están en la clase pero que en la situación ideal no pertenecen a ella. Es una medida del error total que puede considerarse como un indicador de la varianza simple de respuesta.

c) Indice de cambio neto relativo al recuento ideal. En este caso se considera la diferencia neta en relación con el total de personas que idealmente pertenecen a la clase:

$$\frac{c-b}{a+c} \; 100$$

Es muy similar a la medida que usamos corrientemente para medir el error de cobertura, pero aplicado al caso particular de un sector de la población que en el ejemplo presentado son los solteros. El resultado sería negativo si se tienen menos solteros en el recuento real, lo que sucedería por mala declaración del estado civil.

d) Indice de estabilidad. Se define como:

$$\frac{a}{a+b} \; 100$$

Este índice mide la estabilidad de una característica que aparece como correcta (es común al censo y al recuento ideal) a través del tiempo. Si la declaración en los censos ha mejorado, el valor de este índice debería aumentar como consecuencia de un aumento de a, en desmedro de b y c.

4. Las fuentes de errores

El error puede generarse en todas las etapas de la producción de la información, desde la planificación del sistema de recolección de datos hasta la publicación de los resultados finales; por lo tanto, una evaluación completa requiere el análisis de cada una de las actividades involucradas en el proceso de obtención del dato.

a) Etapas previas: planeación y organización. Tómese como ejemplo la preparación de un censo de población. Para este caso, en esta etapa se considera la legislación, el presupuesto, calendario del censo, organización administrativa, labor cartográfica, plan de tabulaciones, preparación del cuestionario, experimentación censal y publicidad.

No es difícil comprender que cualquier irregularidad que se cometa en alguno de los aspectos mencionados anteriormente, compromete la calidad del resultado final. Una mala cartografía puede ser la causa fundamental de omisiones de áreas geográficas completas o de duplicación de empadronamientos por mala delimitación de los segmentos censales. Si el cuestionario está mal confeccionado y las preguntas no son de clara comprensión, es probable que las respuestas sean erróneas y por lo tanto se introduzcan importantes errores de contenido en la información. También cabe incluir bajo esta categoría, los errores originados en la aplicación de conceptos que en la comunidad estudiada tiene un significado distinto o simplemente se desconoce. Así, los ejemplos pueden multiplicarse para cada una de estas etapas preparatorias del censo y además extenderse a las otras fuentes de datos.

b) La comunicación entre el informante y el registrador. Se trata de los errores que se cometen al momento de recoger el dato, debido a la comunicación imperfecta entre el enumerador y el informante.

Errores debidos al enumerador:

i) Conceptuales. Cuando el que registra el dato -enumerador u oficial del registro civil-, quizás debido a una mala capacitación, interroga sobre un asunto que no corresponde a lo que se desea investigar, su origen reside principalmente en una capacitación defectuosa. Por ejemplo, si en lugar de averiguar en un censo el número de hijos nacidos vivos de las mujeres de 15 a 49 años de edad, solamente pregunta sobre los hijos vivos al momento del empadronamiento.Un ejemplo tomado de situaciones de este tipo se encuentra con frecuencia en los registros de estadísticas vitales donde para algunos conceptos se requiere una capacitación especial (determinación de causas de muerte).

ii) Mala interpretación de la respuesta. Este tipo de error es frecuente cuando se trata de preguntas con respuestas abiertas como son las que averiguan sobre las características económicas de la población. Es probable que se confunda un peón rural con un campesino por interpretarse equivocadamente la respuesta de un hombre que trabaja en el campo.

iii) Error al registrar. Cuando se anotan los datos en la boleta censal o en las actas de registro, pueden cometerse equivocaciones. En un cuestionario precodificado en que hay que marcar con una X en una celda determinado atributo, sucede a veces que por apresuramiento o descuido se anota en una que no corresponde.

Errores debidos al informante:

i) Conceptuales o mala interpretación de una pregunta. El informante puede no comprender correctamente qué dato se le está solicitando e involuntariamente brindar una respuesta incorrecta. Por ejemplo, ante la pregunta ¿qué año o grado de estudio más alto aprobó?, es posible que la respuesta se refiera al año que se está cursando, cuando en realidad se solicita el último grado ya aprobado que en este caso será el anterior.

ii) Ignorancia de la respuesta correcta. Para la mayoría de los casos el que informa es un tercero que no siempre tiene un conocimiento cabal de los datos que se solicitan. Ante la ignorancia de la edad de un familiar, el informante puede responder que la ignora, pero también puede intentar dar una estimación aproximada que se aleja en muchos casos de la verdadera. A esto cabe agregar la ignorancia no solamente respecto de datos de un tercero, sino algunos relativos al propio entrevistado, como es el caso de la edad en comunidades poco desarrolladas y de algunas relaciones con terceros (número de hijos nacidos vivos, sobrevivencia de padres o hermanos, etc.).

iii) Error deliberado. Otro tipo de error en la declaración de la información suele ocurrir porque se falsea deliberadamente el dato. El informante está consciente de que la información que está brindando no es la verdadera, pero recurre a ello por muy diversas razones, que en general son ajenas al procedimiento de recolección del dato. Si un censo se realiza simultáneamente con un período de reclutamiento para el servicio militar obligatorio, muchos jóvenes que ya han superado la edad mínima de ingreso al servicio, probablemente declaren una edad menor pensando que el censo está asociado al reclutamiento. Por supuesto este tipo de errores produce una modificación sistemática de la estructura por edades de la población masculina. Otro ejemplo se tiene cuando se investiga el nivel de ingresos en que, por temor a que la investigación estuviese ligada al cobro de impuestos, se declara generalmente una cifra mucho menor que la real.

c) La elaboración y publicación de los datos. Con posterioridad a la recolección de la información también pueden cometerse irregularidades que afecten la calidad de los datos obtenidos.

En la etapa de codificación y crítica se incurre en errores de diversa naturaleza. Algunos de éstos radican en la misma definición de los códigos o en los criterios usados para la crítica, así como en la aplicacion práctica de estos conceptos.

Actualmente es común que parte de la etapa de crítica se realice durante el procesamiento de la información censal, mediante el uso de computadores. Por ejemplo, existen programas que asignan personas a las viviendas cuyos moradores estaban ausentes al momento del empadronamiento y otros que asignan algunas características ignoradas para una persona que ha sido empadronada. Esto puede conducir a violar el principio de "rehabilitación", es decir, que con el ánimo de corregir la información introducimos nuevos errores debido a criterios equivocados del programa de asignaciones.

La publicación misma puede quedar afectada por errores al copiar las cifras de los resultados finales, aunque actualmente se ha reducido en gran medida este error mediante el sistema de copias fotográficas directas de las salidas del computador.

La importancia de poder detectar en qué etapa radica la fuente del error está relacionada no sólo con el hecho de corregir estas irregularidades en el futuro, sino además porque cada fuente de deficiencias define un patrón particular de error y por lo tanto una forma distinta de afrontar su corrección. Por ejemplo, si se produce sistemáticamente una mala declaración de la edad por parte del informante, debido a ignorancia del dato verdadero, lo más probable es que se ocasione lo que se ha dado en llamar "preferencia de dígitos", es decir que al estimar la edad se la redondee en cifras terminadas en "0" y "5".

De esta forma cuando el investigador se enfrenta a los datos demográficos definitivos, debe tener presente que ellos están afectados por deficiencias como las reseñadas que, en general, tienen un efecto acumulativo en perjuicio de su calidad. Es precisamente una función del demógrafo discernir sobre aquellas, identificándolas con miras a su corrección.