

Capítulo 8.

Precisión de los Datos

INTRODUCCIÓN

Nota: Esta sección explicativa fue preparada e incluida aquí antes de decidir si se ajustarían los datos.

Durante todos los censos tomados en Puerto Rico, se han hecho mejoras constantemente al proceso, no sólo para mejorar su eficacia sino también para mejorar su calidad. De acuerdo con esta tradición, para el Censo 2000, los datos presentados en este informe se basan en la enumeración actual del censo. Sin embargo, estos datos se han modificado (ajustado) para reflejar los mejores estimados de las personas omitidas y las personas contadas más de una vez durante la enumeración original.

La modificación al recuento básico del censo se basó en una encuesta grande de muestreo que se utilizó para medir la cobertura neta del recuento básico del Censo 2000. Esta encuesta se conoce como la Evaluación de Precisión y Cobertura (EPC). A base de los resultados de esta encuesta, se hizo un estimado de las personas que fueron omitidas y de las que fueron enumeradas más de una vez durante el censo original utilizando un proceso que se conoce como el sistema dual de estimación. Los estimados de cobertura se usaron para calcular factores de corrección por ajuste o cobertura. Los factores de ajuste se aplicaron a todos los niveles de geografía censal para calcular estimados finales de población.

Por lo tanto, como las tabulaciones en este informe incorporan los resultados de la Evaluación de Precisión y Cobertura, éstas están sujetas a errores de muestreo medibles además de errores ajenos al muestreo de los cuales sufren todos los censos y encuestas. Esta sección presenta una descripción de las operaciones de la Evaluación de Precisión y Cobertura, el diseño de la muestra de la Evaluación de Precisión y Cobertura y la metodología de estimación usada en la Evaluación de Precisión y Cobertura, además de una explicación de los errores de muestreo y los errores ajenos al muestreo.

EL ARCHIVO MAESTRO DE DIRECCIONES Y LOS PROCEDIMIENTOS DE ENUMERACIÓN

El método de entrega de cuestionarios utilizado en Puerto Rico fue actualización/entrega. El listado de direcciones para las áreas de actualización/entrega se compiló durante una operación de campo del Negociado del Censo que se conoce como Listado de Direcciones. Los empleados del censo se mandaron al campo con mapas de su área de asignación y fueron instruidos a anotar las direcciones de tipo urbano, las direcciones de tipo no urbano o una descripción de la localización, o posiblemente alguna combinación de éstas, para cada unidad de vivienda. Además, la localización de la unidad se anotó en el mapa censal con lo que se conoce como un punto en el mapa. Esta operación se llevó a cabo en el otoño de 1998.

Al completar el procesamiento de datos sobre el listado de direcciones, fue posible tabular el número de unidades de vivienda en cada bloque. Como las unidades de vivienda en esas áreas podrían tener direcciones postales que no fueran estándares y que podrían ser anotadas en archivos censales sólo con una descripción de la localización, se le envió listas de recuentos de unidades de vivienda por bloque a las unidades gubernamentales que estaban participando en la operación de revisión local en esas áreas. Esta operación se conoció como LUCA 99. Cuando un participante de LUCA 99 no estaba de acuerdo con un recuento de bloques censales, se hizo un recorrido de nuevo de este bloque, durante el cual los empleados del censo fueron reasignados para actualizar el listado de direcciones. También, hubo un proceso de apelación para resolver discrepancias en la clasificación de unidades de vivienda, que podría tener como resultado añadir unidades de vivienda al listado de direcciones. El segundo recorrido de LUCA 99 y el proceso de apelación de LUCA 99 se llevaron a cabo varias veces durante la creación del Archivo Maestro Decenal de Direcciones. La mayoría de las entidades de LUCA 99 tenían procesados sus resultados

del segundo recorrido antes de crear el Archivo Maestro Decenal de Direcciones inicial, pero muchos no lo tenían. Se diseñaron muchas actualizaciones al Archivo Maestro Decenal de Direcciones específicamente para añadirle resultados que se recibieron después del segundo recorrido y la apelación a los archivos censales para que se entregaran los cuestionarios a tiempo.

La última operación de hacer listados de direcciones en las áreas de actualización/entrega fue la operación Actualización/Entrega. Durante esta operación, se entregó a mano en cada unidad de vivienda un cuestionario del censo. Durante el proceso, se actualizaron el Archivo Maestro de Direcciones y los mapas.

Las personas que no recibieron un cuestionario en su residencia pudieron entregar un cuestionario de Sea Usted Contado o llamar a Asistencia por Teléfono sobre Cuestionarios y tener su información recopilada por teléfono. Las direcciones de estas operaciones que no parearon con las que ya estaban listadas en el Archivo Maestro Decenal de Direcciones fueron visitadas en una operación de Verificación en el Campo para determinar si existían. Las direcciones verificadas fueron añadidas al listado de direcciones.

Una fuente adicional de información sobre unidades de vivienda listadas en el Archivo Maestro Decenal de Direcciones es la operación Seguimiento por Falta de Respuesta (NRFU, por sus siglas en inglés). Durante NRFU, los enumeradores dieron seguimiento a unidades que no habían devuelto un cuestionario predirigido del censo. Las unidades en NRFU podrían ser eliminadas o consideradas como desocupadas. A la vez, las unidades que no aparecieron en el listado de direcciones o los mapas pudieron ser añadidas y enumeradas simultáneamente. Esta operación se llevó a cabo en las áreas de actualización/entrega.

DISEÑO DE LA MUESTRA PARA LA EVALUACIÓN DE PRECISIÓN Y COBERTURA

Resumen

La Evaluación de Precisión y Cobertura (EPC) fue una muestra de múltiples fases diseñada para hacer un estimado de la cobertura neta del Censo 2000. Una muestra de “múltiples fases” es un diseño en el cual la información que se necesita para producir la x fase de la muestra no se puede obtener hasta que se complete la $x-1^{\text{a}}$ fase. La muestra asignada a los Estados Unidos, aproximadamente 300,000 unidades de vivienda en 11,303 áreas de muestra, fue repartida entre los 50 estados y el Distrito de Columbia, más o menos proporcional a su tamaño de población. La muestra de Puerto Rico se calculó a aproximadamente 14,000 unidades de vivienda en 499 áreas de muestra.

Para medir la cobertura neta, se necesitaron dos muestras. La primera fue una muestra de personas que debieron haberse enumerado en el censo. Comparando esta muestra, la muestra de población o muestra P, con las personas que en realidad fueron enumeradas en el censo, se puede obtener un estimado del número de personas omitidas del censo. La segunda muestra fue una muestra de personas que fueron enumeradas en el censo, conocida como la muestra de enumeración, o muestra E. La muestra E se usó para estimar los duplicados y otros tipos de enumeraciones imprecisas en el censo. El establecimiento de estas dos muestras conjuntas constituyó la Evaluación de Precisión y Cobertura.

La unidad principal de muestreo para la Evaluación de Precisión y Cobertura fue el grupo de bloques. Los mismos grupos de bloque fueron incluidos en la muestra E y P. La muestra E representa las enumeraciones del censo en grupos de la muestra. Para la muestra P, los enumeradores llevaron a cabo entrevistas en los grupos de bloques de la muestra. Utilizando un listado de direcciones independiente que se obtuvo de operaciones anteriores, los enumeradores intentaron identificar a todas las personas que vivían en el grupo de bloques durante las entrevistas además de las personas que vivían en los grupos de bloques de la muestra el Día del Censo (el 1 de abril de 2000) pero que se han mudado desde entonces. Esta operación se llevó a cabo durante el verano de 2000. Vea Tabla 8-1 a continuación para los tamaños de la muestra y los resultados de las entrevistas del Día del Censo.

Tabla 8-1. **Características de la Muestra Para EPC: Resultados del Día del Censo**

Característica	Estados Unidos	Puerto Rico
Total de unidades de vivienda	300,913	13,736
Unidades de vivienda ocupadas	261,969	11,309
Unidades de vivienda entrevistadas	254,175	11,007
Unidades de vivienda no entrevistadas	7,794	302
Unidades de vivienda desocupadas, o que no existen	38,944	2,427
Grupos de bloques	11,303	499

Durante la entrevista para la Evaluación de Precisión y Cobertura, que fue independiente de la entrevista del censo, los enumeradores intentaron recopilar los nombres y otras características requeridas para parear las personas de la muestra P a las que fueron enumeradas durante el censo. Además, se estableció la clasificación de residente el Día del Censo para las personas que vivían en el grupo de bloques. Se recopilaron datos por poder para las personas que vivían en el grupo de bloques el Día del Censo pero no durante la entrevista para la Evaluación de Precisión y Cobertura. Los datos de la entrevista y por poder se usaron para buscar la enumeración del censo para ver si las mismas personas ya fueron contadas en el censo. Este proceso de pareo permitió que se determinara el número de omisiones y enumeraciones imprecisas en el censo.

Las personas que vivían en áreas remotas de Alaska (como se definieron las operaciones de enumeración), al igual que aquellas personas que no vivían en hogares, fueron excluidas del grupo de muestreo de la Evaluación de Precisión y Cobertura. Se estimó poco realista intentar hacer una entrevista independiente de estos grupos. Esto significa que la Evaluación de Precisión y Cobertura no incluyó a personas que vivían en dormitorios de universidades, refugios, hogares para personas de edad avanzada, bases militares y otras unidades en alojamientos de grupo. Los recuentos de individuos en estos alojamientos de grupo se añadieron a los totales de los estimados calculados para la Evaluación de Precisión y Cobertura para lograr los totales finales corregidos.

Procedimiento de muestreo

La muestra para la Evaluación de Precisión y Cobertura fue una selección de múltiples fases, estratificada y sistemática de grupos de unidades de vivienda contiguas geográficamente. Se consiguió una muestra separada de cada uno de los 50 estados, el Distrito de Columbia y Puerto Rico.

La unidad principal de muestreo fue el grupo de bloques, que es un grupo de bloques adyacentes para la recopilación del Censo 2000. Los grupos de bloques se formaron utilizando los recuentos de unidades de vivienda del listado preliminar de direcciones censales para crear unidades de muestra que fueron aproximadamente el mismo tamaño y para facilitar esfuerzos eficaces para la recopilación de datos. Un bloque con 80 ó más unidades de vivienda generalmente formó un grupo por sí solo. Se agruparon los otros bloques para que los bloques de tamaño mediano tuvieran un promedio de 30 unidades de vivienda.

Había tres fases de muestreo. Al principio, se seleccionó una muestra grande de listado. La muestra de listado de aproximadamente 30,000 grupos de bloques fue visitado por personal de campo que desarrollaron un listado de direcciones independiente de las unidades de vivienda. Utilizando el resultado de esta operación, se seleccionó una submuestra de grupos de bloques para ser parte de la muestra de la Evaluación de Precisión y Cobertura. Al final, se seleccionó en grupos grandes una submuestra de unidades de vivienda dentro del grupo.

En la primera fase de muestreo, se clasificaron los grupos de bloques a base de su tamaño en cuatro grupos homogéneos mutuamente exclusivos conocidos como estratos de muestreo. Habían estratos para grupos con 0 a 2 unidades de vivienda (grupos pequeños), para grupos con 3 a 79 unidades de vivienda (grupos de tamaño mediano), para grupos con 80 ó más unidades de vivienda (grupos grandes) y para grupos en una reservación de indios americanos con 3 ó más unidades de vivienda. Los grupos grandes fueron seleccionados con mayor probabilidad que los grupos de tamaño mediano en esta fase porque sus ponderaciones aumentarían más tarde durante el submuestreo al nivel de grupos de bloque de unidades de vivienda. Dentro de cada estrato de muestreo, se clasificaron los grupos según sus características geográficas y demográficas, y se seleccionó una muestra sistemática con igual probabilidad dentro de cada estado de los EE.UU.

En la segunda fase de muestreo, se presentó algún muestreo diferencial. Primero, estratos de reducción se formaron dentro de los estratos de muestreo medianos y grandes a base de la composición demográfica estimada de los grupos de bloques y la relación entre el recuento de unidades de vivienda del listado independiente y el listado de direcciones censales de enero de 2000. Dentro de cada estrato de reducción, se clasificaron los grupos y se seleccionó una muestra sistemática con igual probabilidad. En esta fase, los grupos de tamaño mediano y grande en Puerto Rico no fueron parte de la submuestra.

La segunda fase también incluyó una reducción de grupos pequeños. Usando los recuentos de unidades de vivienda del listado independiente y el listado de direcciones censales de enero de 2000, los grupos de bloques pequeños dentro de cada estado y Puerto Rico fueron estratificados por tamaño. Después, se seleccionó una muestra sistemática de cada estrato con igual probabilidad. Se retuvieron todos los grupos con diez o más unidades de vivienda en el listado de direcciones censales o el listado de direcciones independiente. También se retuvieron los grupos en reservaciones de indios americanos tanto como los grupos en áreas de listado/enumeración.

Durante un paso individual de muestreo, se seleccionó una submuestra de unidades de vivienda dentro de los grupos grandes para ser entrevistadas, donde se utilizó una medida actualizada del tamaño del grupo. Si un grupo consistió en 79 ó menos unidades de vivienda, se incluyeron todas las unidades de vivienda en la muestra final de la Evaluación de Precisión y Cobertura. Para los grupos con 80 ó más unidades de vivienda, se seleccionó una submuestra de estas unidades de vivienda para reducir los efectos de diseño y facilitar la recopilación de datos en el campo. El submuestreo se llevó a cabo formando grupos de igual tamaño de unidades de vivienda adyacentes, conocidos como segmentos. Se seleccionó una muestra sistemática de segmentos dentro de cada grupo. Todas las unidades de vivienda en los segmentos seleccionados se incluyeron en la muestra final de la Evaluación de Precisión y Cobertura.

La muestra P consistió en las entrevistas de la Evaluación de Precisión y Cobertura que se llevaron a cabo en esos grupos de bloques y segmentos de grupos de bloques seleccionados. La muestra E consistió en las enumeraciones del censo que ocurrieron en estos mismos grupos de bloques y segmentos de grupos de bloque.

LAS OPERACIONES DE LA EVALUACIÓN DE PRECISIÓN Y COBERTURA

Los procedimientos principales de la Evaluación de Precisión y Cobertura (EPC) incluyeron el pareo de unidades de vivienda, las entrevistas y el pareo de personas. Representantes de campo recorrieron sistemáticamente cada grupo de bloques de la muestra y anotaron las direcciones para todos los alojamientos, creando el listado independiente de unidades de vivienda de EPC. Las unidades de vivienda de EPC se parearon al inventario del censo de unidades de vivienda del Archivo Maestro de Direcciones del Censo Decenal de enero de 2000 utilizando un algoritmo computadorizado con una revisión en la oficina. La computadora sólo usó la dirección para el pareo pero los oficinistas usaron otros materiales para ayudarles en parear las unidades de vivienda. No todas las direcciones fueron de tipo urbano, lo cual consistía en un número de casa y nombre de calle. Los números de ruta y buzón se usaron en algunas áreas rurales. Algunas direcciones sólo consistieron en una descripción de la localización. Los mapas que identifican la localización de unidades de EPC y unidades de vivienda censales sirvieron de ayuda en el pareo de unidades de vivienda. Después de conciliar las que no parearon, se preparó una lista de las unidades de vivienda de EPC que se confirmaron haber existido dentro de los grupos de bloques y se llevaron a cabo las entrevistas a personas en esas unidades de vivienda. La entrevista a personas en cada unidad de vivienda se hizo usando una computadora portátil.

Se identificaron tres tipos de personas en estas entrevistas: aquéllas que vivían en la dirección de muestra al momento de la entrevista y el Día del Censo, aquéllas que se habían mudado a la dirección de muestra desde el Día del Censo y aquéllas que vivían en la dirección de muestra el Día del Censo pero se mudaron antes de la entrevista de EPC. Al principio, las entrevistas se llevaron a cabo sólo con los miembros del hogar. Se permitieron hacer entrevistas con personas que no eran miembros del hogar después de que todos los esfuerzos hechos para comunicarse con un miembro del hogar fueron infructuosos. Fue necesario llevar a cabo entrevistas con respondedores por poder en los hogares donde los residentes se habían mudado después del Día del Censo.

El procedimiento de estimación necesita la tasa de pareo para las personas de la muestra P y la tasa de enumeración precisa para las personas de la muestra E. Las personas que vivían en una unidad de vivienda el Día del Censo fueron pareadas con las personas enumeradas en el censo para lograr un estimado de la tasa de pareo para la Evaluación de Precisión y Cobertura. Esta operación de pareo se llevó a cabo con una computadora y revisión en la oficina. El nombre, dirección, fecha de nacimiento, edad, sexo, raza, origen hispano o latino y parentesco se usaron para identificar pareos entre las personas de la Evaluación de Precisión y Cobertura y las enumeraciones del censo. La tasa de enumeración precisa se estimó del pareo y las entrevistas para las personas enumeradas en el censo. Las personas del censo que fueron pareadas en la Evaluación de Precisión y Cobertura fueron enumeradas correctamente. Las personas del censo que no parearon con la Evaluación de Precisión y Cobertura fueron entrevistadas durante operaciones de seguimiento en el campo para clasificarlas como personas enumeradas correctamente o incorrectamente. Los resultados de la operación de pareo fueron preparados y utilizados en la fase de estimación de la Evaluación de Precisión y Cobertura.

En algunos casos, la clasificación de enumeración precisa y la clasificación de pareo no se pudo resolver por el pareo por computadora, en la oficina o por seguimientos en el campo. En estos casos, las tasas de pareo y enumeración precisa fueron imputadas a base de la distribución de los resultados de las entrevistas resueltas por seguimiento. También como en el censo, algunos respondedores no contestaron todas las preguntas en la entrevista de la Evaluación de Precisión y Cobertura necesarias para hacer el estimado. La imputación de información que faltó en la muestra E se resolvió durante el procesamiento censal. Sin embargo, si las variables de tenencia, sexo, raza, origen hispano o latino o edad faltaban para los individuos de la muestra P, se imputó esa información dependiendo de la variable según la distribución de la variable dentro del hogar, la distribución de la variable nacionalmente o el uso de métodos de “hot-deck,” el procedimiento de imputar entradas omitidas. Finalmente, se hicieron ajustes para tomar en cuenta las ponderaciones de hogares que debían haber sido entrevistados en la Evaluación de Precisión y Cobertura.

BÚSQUEDA DE EXTENSIÓN DIRIGIDA

La Búsqueda de Extensión Dirigida (TES, por sus siglas en inglés) se diseñó para mejorar la precisión del sistema dual de estimación (DSE, por sus siglas en inglés). La operación se estableció en un subconjunto de grupos de bloques de la Evaluación de Precisión y Cobertura seleccionado por una combinación de muestreo de certeza y probabilidad.

Hay errores de codificación geográfica de exclusión e inclusión en los grupos de bloques de la Evaluación de Precisión y Cobertura. Los errores de codificación geográfica por exclusión afectan la tasa de pareo de la muestra P y los errores de codificación geográfica por inclusión afectan la tasa errónea de enumeración en el censo o la muestra E. Si la unidad de vivienda censal se omite del grupo de bloques de la muestra, el hogar de la muestra P no se pareará. Esto resultará en una tasa de pareo más baja. Para la muestra E, si una unidad de vivienda se incluye en el grupo de bloques de la muestra debido a un error de codificación geográfica, las personas de la muestra E serán enumeradas incorrectamente. La motivación para usar un área de búsqueda extendida es permitir que más personas de la muestra P sean pareadas y que más personas de la muestra E sean convertidas a una enumeración correcta. Evaluaciones de censos anteriores han demostrado que el error de codificación geográfica es muy agrupado. La operación de TES se diseñó para aprovechar la distribución de errores de codificación geográfica al enfocar esos grupos que contienen la posibilidad mayor de errores de codificación geográfica. El resultado del sistema dual de estimación no se afecta mientras se tratan las dos muestras igualmente con respecto al área de búsqueda. El establecimiento de esta operación resultará en un sistema dual de estimación más preciso.

En la Encuesta de Pos Enumeración (PES, por sus siglas en inglés) de 1990, la operación de búsqueda extendida se llevó a cabo para todos los grupos de bloques. La Búsqueda de Extensión Dirigida para la Evaluación de Precisión y Cobertura de 2000 fue un proceso de dos etapas. Primero, los grupos se identificaron a base del provecho potencial de extender el área de búsqueda a los bloques circundantes; o sea, los grupos de bloques con un gran número de errores posibles de codificación geográfica. Segundo, se dirigió la búsqueda a los bloques dentro de los grupos seleccionados. En resumen, la dirección y el muestreo se diseñaron para cumplir con estos requisitos:

-
- Incluir con certeza en la Búsqueda de Extensión Dirigida el 5 por ciento de grupos con el mayor número total de errores posibles de codificación geográfica; sin y con ponderaciones por sus respectivas ponderaciones de muestreo de la Evaluación de Precisión y Cobertura.
 - Asignar todos los grupos listados de nuevo y de listado/enumeración a ser incluidos con certeza en la muestra de la Búsqueda de Extensión Dirigida.
 - Sacar una muestra que consiste en el 10 por ciento de grupos entre esos grupos con por lo menos un error posible de codificación geográfica.

La muestra de la Búsqueda de Extensión Dirigida se obtuvo de los resultados de la primera operación de pareo de unidades de vivienda. Los cambios en el inventario de unidades de vivienda censales después de enero de 2000 no fueron reflejados en las operaciones de pareo de unidades de vivienda utilizadas para identificar los grupos de bloques de la muestra de la Búsqueda de Extensión Dirigida. Las unidades de vivienda del censo que no fueron listadas en el Archivo Maestro de Direcciones del Censo Decenal de enero—las adiciones censales que se identificaron como errores de codificación geográfica en el seguimiento de personas durante la Evaluación de Precisión y Cobertura—se imputaron como probabilidad de una enumeración precisa. Las personas en hogares completos de la muestra P que no parearon y que vivían en unidades de vivienda pareadas con las unidades censales que se eliminaron entre enero y septiembre también fueron parte de la búsqueda extendida.

PROCEDIMIENTOS DE LA EVALUACIÓN DE PRECISIÓN Y COBERTURA

El propósito de la Evaluación de Precisión y Cobertura fue contar a las personas que no fueron enumeradas utilizando procedimientos tradicionales del censo o que fueron incluidas equivocadamente en el censo. Estudios de censos anteriores muestran evidencia convincente de que ha sido más difícil contar a personas en grupos demográficos específicos, especialmente grupos minoritarios e inquilinos utilizando métodos tradicionales del censo. La Evaluación de Precisión y Cobertura se diseñó para permitir estimados válidos estadísticamente de la proporción de la población omitida. Las estimaciones de la Evaluación de Precisión y Cobertura que aparecen en esta publicación se obtuvieron mediante un proceso de dos pasos. Los dos pasos son el sistema dual de estimación y la estimación sintética, o estimación de áreas pequeñas. Los dos pasos se explican a continuación.

Sistema Dual de Estimación

El Sistema Dual de Estimación (DSE, por sus siglas en inglés) se usó para hacer un estimado de la proporción de la población de hogares omitidos o incluidos erróneamente en el censo, incluyendo récords inputados. Estimaciones de las personas omitidas o contadas dos veces no se hicieron para las personas en alojamientos de grupo.

El término sistema dual de estimación se derivó del uso de dos encuestas independientes para medir la misma población en las áreas de muestra de la Evaluación de Precisión y Cobertura. El sistema dual de estimación ha sido utilizado por muchos años en estudios de poblaciones difíciles de contar. El recuento tradicional del censo se llevó a cabo como el primer sistema. El segundo sistema, la encuesta independiente—la Evaluación de Precisión y Cobertura del Censo 2000—se llevó a cabo solamente en las áreas de muestra de la Evaluación de Precisión y Cobertura. Después, se parearon los resultados de la Evaluación de Precisión y Cobertura a los resultados del censo. Se hizo un estimado del porcentaje de personas encontradas en la segunda muestra que fueron contadas en el primer recuento. Por ejemplo, si el 95 por ciento de las personas encontradas en el segundo sistema también fueron encontradas en el primer sistema, se presumió que el primer sistema encontró el 95 por ciento de la población total y omitió el último 5 por ciento de la población total. En este caso, el recuento del censo sería multiplicado por $1.00/0.95$ ó aproximadamente 1.0526 para obtener un estimado de la población total. También se diseñó la Evaluación de Precisión y Cobertura para estimar la proporción de personas en el censo que no debían haber sido incluidas.

Pos estratificación

El sistema dual de estimación calcula la proporción de personas omitidas en cada uno de los grupos de población relativamente homogéneos, conocidos como pos estratos. Los pos estratos

para la Evaluación de Precisión y Cobertura del Censo 2000 fueron definidos por estas variables: dominio de raza/origen hispano, edad/sexo y tenencia, región censal, área estadística metropolitana/tipo de área de enumeración y tasa de devolución. Los siete dominios de raza/origen hispano fueron:

- Dominio 1. En reservaciones de indios americanos o nativos de Alaska
- Dominio 2. Fuera de reservaciones de indios americanos o nativos de Alaska
- Dominio 3. Hispano
- Dominio 4. Negro o africano americano no hispano
- Dominio 5. Nativo de Hawaii o de las Islas del Pacífico
- Dominio 6. Asiático no hispano
- Dominio 7. Blanco no hispano o “Alguna otra raza”

Las personas se incluyeron en el dominio del número más bajo para el cual todas las condiciones de inclusión fueron satisfechas. Así, las personas que marcaron indio americano o nativo de Alaska fueron incluidas en uno de los dominios de indio americano o nativo de Alaska. Las personas que marcaron hispano fueron incluidas en el dominio de hispano, y así sucesivamente. Para el Censo 2000, las personas pudieron marcar más de una raza. Las personas no hispanas que marcaron dos grupos de raza fueron incluidas en el dominio para uno de los dos dominios marcados según las reglas de inclusión especificadas de antemano. Las personas no hispanas que marcaron tres o más grupos de raza (excluyendo indio americano o nativo de Alaska en territorio de indios) fueron incluidas en el dominio de blanco no hispano o alguna otra raza. Por ejemplo, si un respondedor marcó no hispano, blanco y negro o africano americano, fue incluido en el dominio de negro o africano americano no hispano. Todas las combinaciones de una sola raza y múltiples razas se tabularon separadamente para el censo.

Dentro de cada dominio de raza/origen hispano, siete grupos de edad/sexo fueron definidos al combinar valores de edad y sexo de la siguiente manera:

1. Todas las personas 0 - 17 años
2. Varones 18 - 29 años
3. Hembras 18 - 29 años
4. Varones 30 - 49 años
5. Hembras 30 - 49 años
6. Varones 50 años o más
7. Hembras 50 años o más

Todas las personas en unidades de vivienda ocupadas por el dueño, con hipotecas o libres y sin deuda (sin una hipoteca o préstamo), se consideraron propietarios. Todas las personas en unidades de vivienda alquiladas u ocupadas sin pagar alquiler en efectivo no se consideraron propietarios.

Los propietarios blancos no hispanos o de alguna otra raza en los siete grupos de edad/sexo fueron pos estratificados, además, por las cuatro regiones del Negociado del Censo: el Noroeste, la Región Central, el Sur y el Oeste. Esto tuvo como resultado 28 pos estratos para propietarios blancos no hispanos o de alguna otra raza. Los blancos no hispanos o de alguna otra raza que no eran propietarios no fueron subdivididos por región, lo que tuvo como resultado siete pos estratos de edad/sexo para personas blancas no hispanas o de alguna otra raza que no eran propietarios. Estos 35 grupos combinados de personas blancas no hispanas o de alguna otra raza se dividieron más en 140 grupos clasificando cada pos estrato con cuatro categorías de área estadística metropolitana/tipo de área de enumeración (ÁEM/Tipo de área de enumeración). Las categorías para ÁEM/Tipo de área de enumeración fueron (1) áreas de “mailout/mailback” (envío y entrega de cuestionarios por correo) de las 10 áreas estadísticas metropolitanas/áreas estadísticas metropolitanas consolidadas más grandes basado en recuentos no ajustados de población total del Censo 2000, incluyendo personas en alojamientos de grupo y personas en áreas remotas de Alaska, (2) áreas de “mailout/mailback” (envío y entrega de cuestionarios por correo) de todas las

otras ÁEM/ÁEMC con recuentos no ajustados de población total del Censo 2000, incluyendo personas en alojamientos de grupo y personas en áreas remotas de Alaska, de por lo menos 500,000 personas, (3) todas las otras áreas de “mailout/mailback” (envío y entrega de cuestionarios por correo), y (4) todos los otros tipos de áreas de enumeración.

Las personas en estos 140 grupos de personas blancas no hispanas o de alguna otra raza se dividieron en 280 pos estratos por la variable de la tasa de devolución. La tasa de devolución fue una variable del nivel de sector censal que midió la proporción de unidades de vivienda ocupadas en el universo de envío (“mailback”) que devolvieron un cuestionario del censo. Sectores de tasas bajas (altas) de devolución fueron los sectores cuya tasa de devolución fue menos que o igual al (más que) el valor límite de 25 percentil de la tasa de devolución. Valores límites separados del veinticinco percentil se formaron para los dos dominios aplicables de raza/origen hispano por grupos de tenencia.

Antes de ser agrupadas, las categorías de la pos estratificación para los dominios de negro no hispano e hispano eran idénticas. Cada dominio contenía siete grupos de edad/sexo dentro de cada dos grupos de tenencia, resultando en 14 pos estratos. Estos 14 pos estratos se clasificaron aún más en 56 grupos por la variable de ÁEM/Tipo de área de enumeración que tenían las siguientes categorías: (1) áreas de “mailout/mailback” (envío y entrega de cuestionarios por correo) con cifras totales no ajustadas de población del Censo 2000, incluyendo personas en alojamientos de grupo y áreas remotas de Alaska, de por lo menos 500,000 personas, y (2) todas las otras áreas de “mailout/mailback” (envío y entrega de cuestionarios por correo) y los otros tipos de áreas de enumeración. Los 28 pos estratos se dividieron aún más en 56 pos estratos basado en la variable de la tasa de devolución. La tasa de devolución fue una variable del nivel de sector censal que medía la proporción de unidades de vivienda ocupadas en el universo de envío y entrega que devolvieron un cuestionario del censo. Sectores de tasas bajas (altas) de devolución fueron estos sectores cuyos tasas de devolución fue menos que o igual al (más que) el valor límite de 25 percentil de la tasa de devolución. Valores límites separados del veinticinco percentil se formaron para los cuatro dominios de raza/origen hispano por grupos de tenencia.

Los dominios restantes de raza/origen hispano, que incluyen nativos de Hawaii o de las Islas del Pacífico, asiáticos no hispanos, indios americanos en reservaciones o nativos de Alaska e indios americanos no en reservaciones o nativos de Alaska, consistieron en dos categorías de tenencia y siete categorías de edad/sexo. Por lo tanto, cada uno de estos cuatro dominios de raza/origen hispano tenía 14 pos estratos.

Para los Estados Unidos, había 448 pos estratos potenciales que se pueden resumir en los dominios siguientes de raza/origen hispano: indio americano en una reservación o nativo de Alaska (14), indio americano no en una reservación o nativo de Alaska (14), hispano (56), negro no hispano (56), nativo de Hawaii o de las Islas del Pacífico (14), asiático no hispano (14) y blanco no hispano o de alguna otra raza (280). Alguna agrupación de los pos estratos se llevó a cabo para evitar aumentos en la variante debido a tamaños pequeños de la muestra.

Para Puerto Rico, había 84 pos estratos potenciales sacados de la clasificación cruzada completa de siete categorías para edad/sexo, dos categorías para tenencia, dos categorías para la tasa de devolución y tres categorías de las ÁEM. Valores límites separados del veinticinco percentil de la tasa de devolución se establecieron para propietarios y no propietarios. Los tres grupos de ÁEM para Puerto Rico fueron: (1) ÁEMC de San Juan, (2) Otro ÁEM y (3) No ÁEM.

Sistema Dual de Estimación

El sistema dual de estimación para cada pos estrato se definió por:

$$DSE = (CEN - II) \times \frac{CE}{E} \times \frac{P}{M}$$

donde CEN fue el recuento del censo para el pos estrato, excluyendo las adiciones tardías, II fue el número de personas en el censo cuyos datos fueron imputados por completo o casi por completo, CE fue el estimado ponderado del número de personas enumeradas correctamente en el censo, E fue el estimado ponderado del número de personas en el censo, P fue el estimado ponderado del número de personas encontradas por los procedimientos independientes de recopilación para la Evaluación de Precisión y Cobertura, y M fue el estimado ponderado del número de personas

encontradas por los procedimientos independientes de la Evaluación de Precisión y Cobertura que se parearon a las personas enumeradas en el censo. Todos los estimados ponderados se basaron en los resultados de los bloques de muestra para la Evaluación de Precisión y Cobertura.

Las personas que se mudaron entre el Día del Censo y la entrevista independiente de la Evaluación de Precisión y Cobertura requirieron tratamiento especial en el tercer componente del DSE que se menciona anteriormente. Para el estimado de P, fue mejor hacer un estimado del número de personas que se mudaron a las áreas de muestra de la Evaluación de Precisión y Cobertura que determinar el número de personas que se mudaron de estas áreas. Así, el número de personas que se mudaron fue estimado de las personas que se mudaron a las áreas de muestra de EPC. Sin embargo, para el estimado de M, fue mejor parear a las personas enumeradas en el censo utilizando el número de personas que se mudaron de las áreas de muestra de EPC. Como resultado, el número estimado de personas que se mudaron y que se podían parear a las personas enumeradas en el censo se calculó multiplicando el número estimado de personas que se mudaron a las áreas de muestra de EPC por la tasa estimada de pareo para las personas que se mudaron de las áreas de muestra de EPC.

Los factores de corrección de cobertura se obtuvieron dividiendo los DSE por los recuentos del censo, incluyendo las adiciones tardías. Los factores de corrección de cobertura para grupos de población que generalmente tenían buena cobertura se calcularon cerca de 1.00. Los grupos de población con poca cobertura tenían factores de corrección de cobertura un poco más alto que aquéllos con buena cobertura.

Estimación sintética (de áreas pequeñas)

Nota: Esta sección explicativa fue preparada e incluida aquí antes de decidir si se ajustarían los datos.

Un factor de corrección de cobertura se calculó para todos los pos estratos utilizando el método descrito anteriormente. Si se estimó que un pos estrato tenía más personas que el recuento censal, se seleccionó al azar una muestra del tamaño apropiado de personas censales dentro de cada bloque en este pos estrato. Se duplicaron los datos de las personas seleccionadas en sus bloques con una ponderación de +1. Si se estimó que un pos estrato tenía menos personas que el recuento censal, se seleccionó al azar una muestra de la cantidad apropiada de personas censales dentro de cada bloque en este pos estrato. Se duplicaron los datos de las personas seleccionadas en sus bloques con una ponderación de -1. **Bajo este procedimiento, no se quitaron de los archivos de datos del Censo 2000 ningunos datos sobre cualquier individuo.**

Las estimaciones se calculan a todo nivel de agregación añadiendo los datos de los bloques y/o pos estratos apropiados.

Cómo aplicar los factores de corrección de cobertura - un ejemplo

Nota: Esta sección explicativa fue preparada e incluida aquí antes de decidir si se ajustarían los datos.

Los factores de corrección de cobertura se usaron para calcular el estimado final de población por pos estratos para cualquier bloque. Esto se logró multiplicando el recuento del censo para un pos estrato determinado en un bloque por el factor de corrección de cobertura de ese pos estrato. Por ejemplo, si el factor de corrección de cobertura para propietarios varones negros no hispanos, de 18 a 29 años, fue 1.02, se añadieron dos nuevas personas por cada 100 personas en el censo con estas características. Muy pocos bloques serán tan grandes que tendrán 100 personas en cualquier pos estrato. Si un bloque tenía un recuento censal de 25 personas así, multiplicando por el factor de corrección de cobertura resultó en la necesidad de añadir a una media persona. Para lograr esto, se añadió una persona una mitad del tiempo. Si no había personas con estas características que vivían en el bloque, no se añadió ninguna.

ENUMERACIÓN DE LUGARES QUE PROVEEN SERVICIOS

Nota: Esta sección explicativa contiene referencias a la pos estimación de EPC de la población total y fue preparada antes de publicar los datos y decidir si se ajustarían los datos.

La Enumeración de Lugares que Proveen Servicios (SBE, por sus siglas en inglés) fue el programa estadístico diseñado para contar a las personas sin una residencia habitual que usan facilidades de servicios (como por ejemplo, refugios, comedores de beneficencia, servicios rodantes de comida). Para hacer un estimado de esta población, se usó un estimador de multiplicidad. Se utilizó estimación de multiplicidad cuando personas sin una residencia habitual se pueden relacionar con una o más unidades enumeradas. En el contexto censal, la unidad de enumeración fue el día de la Enumeración de Lugares que Proveen Servicios. Los clientes del lugar que proveían servicios se pudieron relacionar con una o más unidades de enumeración o días utilizando información obtenida mediante la pregunta sobre el uso de servicio. Se asignó a cada persona una ponderación de encuesta a base de la frecuencia del uso. Las ponderaciones de encuesta se usaron para producir un estimado de este componente de la población total. Además, las personas enumeradas en localizaciones específicas al aire libre y personas sin una residencia habitual que devolvieron cuestionarios de Sea Usted Contado aumentaron la estimación. Este estimado final no está sometida a la corrección de cobertura por la Evaluación de Precisión y Cobertura; el estimado de la población de SBE simplemente se añadió a la pos estimación de la población total de la Evaluación de Precisión y Cobertura. El error debido a la estimación de multiplicidad no se incorporó a los valores del error estándar que se indican a continuación. Este componente de la enumeración no se debe interpretar como un estimado de la población de personas sin hogar.

CONFIDENCIALIDAD DE LOS DATOS

El Negociado del Censo ha modificado algunos datos en esta difusión de datos para proteger la confidencialidad. La Sección 9 del título 13 del Código de los Estados Unidos prohíbe que el Negociado del Censo haga público los resultados que pueden identificar datos sobre un individuo.

La comisión interna de revisión de supresión de datos (“Internal Disclosure Review Board”) del Negociado del Censo observa el proceso de revisión de la supresión de datos y establece las reglas de confidencialidad para todas las emisiones de datos. El método de usar listas de cotejo se usa para asegurar que se consideran y se manejan todos los riesgos posibles. Se hace una lista de problemas posibles y la comisión de revisión de la supresión de datos asegura que se toman los pasos apropiados para asegurar la confidencialidad de los datos.

Título 13 del Código de los Estados Unidos

El título 13 del Código de los Estados Unidos autoriza que el Negociado del Censo lleve a cabo censos y encuestas. La Sección 9 del mismo título requiere que cualquier información que se recoPILE del público bajo la autoridad del título 13 se mantenga confidencial. La Sección 214 del título 13 y las Secciones 3559 y 3571 del título 18 del Código de los Estados Unidos estipula que se apliquen penalidades de hasta 5 años de cárcel y hasta \$250,000 en multas por la divulgación intencional de información confidencial del censo.

Límite a la divulgación de datos

El límite a la divulgación de datos es el proceso de disfrazar los datos para proteger la confidencialidad. Una divulgación de datos ocurre cuando alguien puede usar información estadística publicada para identificar a un individuo que provee información bajo una promesa de confidencialidad. Utilizando la limitación a la divulgación, el Negociado del Censo modifica o elimina todas las características que podrán exponerse al riesgo de revelación. Aunque pueda parecer que una tabla muestra información sobre un individuo específico, el Negociado del Censo ha tomado pasos para disfrazar los datos originales mientras asegura que los resultados son útiles.

Intercambio de datos

El intercambio de datos es un método de limitar la divulgación de datos que se diseñó para proteger la confidencialidad en tablas con datos de frecuencia (el número o porcentaje de la población con ciertas características). El intercambio de datos se hace editando las fuentes de datos o intercambiando los registros por una muestra de casos al crear una tabla. Una muestra de hogares se selecciona y se para usando un conjunto escogido de variables claves con hogares en áreas geográficas vecinas (bloques o sectores censales) que tienen características semejantes

(como por ejemplo, el mismo número de adultos y el mismo número de hijos). Como el intercambio ocurre frecuentemente dentro de un sector, no afecta los totales marginales para el sector o para los totales que incluyen datos de sectores múltiples. Debido al intercambio, los usuarios no deben suponer que las tablas con celdas con un valor de 1 ó 2 revelen información sobre individuos específicos.

CÁLCULO DE ERRORES ESTÁNDARES

Nota: Esta sección explicativa fue preparada e incluida aquí antes de decidir si se ajustarían los datos.

Tipo de error

La variabilidad surge en todas las muestras, como la muestra de la Evaluación de Precisión y Cobertura establecida para el Censo 2000. Los estimados diferirían si se hubiera escogido a distintas personas y unidades de vivienda para la muestra de la Evaluación de Precisión y Cobertura. El error estándar y la variación (el cuadrado del error estándar) son medidas de la variación entre los estimados de todas las muestras posibles y, por lo tanto, son medidas de la precisión con la cual un estimado de una muestra particular se aproxima al resultado promedio de todas las muestras posibles.

Además de la variabilidad que surge de los procedimientos de muestreo, los estimados están sujetas a errores ajenos al muestreo (errores humanos y de computadoras), que se pueden introducir durante cualquiera de las varias operaciones complejas utilizadas para recopilar, procesar y tabular datos.

El error ajeno al muestreo puede afectar los datos en dos formas. Errores que se introducen al azar aumentarán la variabilidad de los datos y por lo tanto deben reflejarse en el error estándar. Errores que tienden a ser uniformes en una dirección harán que los datos estén sesgados en esa dirección. Por ejemplo, si los respondedores constantemente tienden a informar menos años de edad, las cifras para edad tendrán la tendencia a ser sesgadas en las categorías de edades menores. Entonces, el estimado de la categoría de personas por edad que resulta será más bajo que las cifras actuales. Tales sesgos no son reflejados en el error estándar.

Cálculo del error

El cálculo del total de la variación pretendió capturar la variación debido a los procedimientos de muestreo de la Evaluación de Precisión y Cobertura, los procedimientos para datos omitidos que se utilizaron con las muestras E y P, al igual que con la variación creada al aplicar los factores de corrección sobre la cobertura del sistema dual de estimación a la población en conjunto. Estos estimados de variación no capturan la variación debido al estimado de población de la enumeración en lugares que proveen servicios. Se usaron métodos de duplicación para calcular el total de la variación.

El proceso de calcular las variaciones fue complicado porque el diseño de muestreo de la Evaluación de Precisión y Cobertura fue una muestra de múltiples fases. El muestreo de múltiples fases difiere de un muestreo de múltiples etapas en la siguiente manera: en un diseño de múltiples etapas, la información que se necesita para incluir todas las etapas de la muestra se conoce antes de que empiece el muestreo; en un diseño de múltiples fases, la información que se necesita para hacer la n fase de la muestra no se puede obtener hasta que se complete la $n-1^{\text{a}}$ fase de la muestra. Se desarrolló una nueva metodología para las muestras de múltiples fases, la que permitió que se calcule la variación por una modificación del procedimiento de estimación de la variación de Rao-Shao.

El componente básico del nuevo procedimiento fue el método jackknife. En el método jackknife, se dió a cada grupo un conjunto de aproximadamente 30,000 ponderaciones duplicadas. Para cada duplicación, los grupos en estratos específicos de muestreo fueron ponderados de nuevo y esas nuevas ponderaciones duplicadas se usaron en recalcular la estimación del sistema dual de estimación. Se basa la variación en las diferencias cuadradas entre las aproximadamente 30,000 estimados del DSE duplicados y la producción oficial de estimados del DSE.

Debido al gran número de estimados a niveles más bajos de geografía, no sería posible proveer tablas que listan el error estándar de cada estimado publicado. En vez, se provee un parámetro, el coeficiente de variación generalizado (GVC, por sus siglas en inglés) que permite al usuario aproximar el error estándar para cualquier estimado deseado. El coeficiente de variación generalizado de un estimado es simplemente la proporción del error estándar del estimado y el estimado en sí.

Se diseñó el GVC para un subconjunto de datos sobre la redistribución de distritos electorales (Ley Pública 94-171) que se categorizan separadamente por población total, raza, edad y origen hispano latino para cada estado a niveles de geografía más bajo que estado. La estimación de interés puede sustituirse en la fórmula de la función de la variación generalizada utilizando los GVC computarizados para calcular un estimado del error estándar.

Para calcular el error estándar de un estimado o una proporción estimada, el primer paso es escoger el GVC apropiado de las tablas de GVC en American FactFinder™, basado en una combinación de edad/raza/origen hispano o latino que es lo más apropiado para la estimación. Si se puede aplicar más de un GVC al estimado deseado, se recomienda que el usuario calcule los errores estándares utilizando el GVC aplicable más grande.

El error estándar de un estimado, \hat{x} , se calcula utilizando la fórmula

$$SE(\hat{x}) = GCV \times \hat{x} \quad (1)$$

donde \hat{x} , es el número estimado de personas, y CVG es el parámetro estimado que se encuentra en las tablas en American FactFinder™.

En algunos casos, el error ajeno al muestreo es un componente grande del error total (de muestreo y ajeno al muestreo). Especialmente, para bloques y grupos de bloques, esperamos que el error ajeno al muestreo sea, en promedio, grande en comparación con el error de muestreo. Por lo tanto, sugerimos que no se calculen las variaciones de muestreo para obtener estimados para bloques y grupos de bloque porque considerablemente se subestimaría el error total en estos cálculos. Para áreas geográficas que son más grandes que bloques y grupos de bloques, las estimaciones se sujetan menos a este problema. Aún, como el error ajeno al muestreo está presente en las estimaciones, sugerimos que los usuarios tengan cuidado al interpretar variaciones de muestreo como medidas de errores generales, especialmente para áreas pequeñas.

Para cualquier estimado que es la suma o diferencia de dos o más estimados dados, el error estándar es la raíz cuadrada de las sumas de los errores estándares al cuadrado para los estimados dados:

$$SE(\hat{x}_1 \pm \hat{x}_2 \pm \dots) = \sqrt{(SE(\hat{x}_1))^2 + (SE(\hat{x}_2))^2 + \dots} \quad (2)$$

Sin embargo, este método subestimaré (sobrestimaré) el error estándar si existe una alta correspondencia positiva (o negativa) entre las dos partidas en una suma o si existe una alta correspondencia negativa (o positiva) entre las dos partidas en una diferencia.

Como no hay un error de muestreo asociado con los recuentos no ajustados del censo, el error de muestreo que está asociado con una estimación de la cobertura neta del error—la diferencia entre la estimación ajustada de EPC y el recuento no ajustado del censo—es igual al error de muestreo de la estimación ajustada de EPC utilizando la fórmula (2).

Una estimación de la tasa de cobertura neta,

An estimate of the net coverage rate,

$$\text{Tasa de cobertura neta} = 100\% \times \frac{\text{Estimación de EPC} - \text{Recuento del censo}}{\text{Estimación de EPC}} \quad (3)$$

y una aproximación de su error estándar:

$$SE(\text{Tasa de cobertura neta}) = (100\% - \text{Tasa de cobertura neta}) \times GVC \quad (4)$$

Si la estimación de interés es una proporción entre dos variables (\hat{x} y \hat{y}) la fórmula a continuación debe utilizarse para aproximar los errores estándares:

$$SE(\hat{x}/\hat{y}) = \frac{\hat{x}}{\hat{y}} \sqrt{\frac{(SE(\hat{x}))^2}{\hat{x}^2} + \frac{(SE(\hat{y}))^2}{\hat{y}^2}} \quad (5)$$

Sin embargo, este método sobreestimaré (subestimaré) el error estándar si existe una alta correspondencia positiva (negativa) entre las dos partidas en la proporción.

Intervalos de Confianza

Un estimado de la muestra y su error estándar estimado se pueden utilizar para construir intervalos de confianza de la estimación. Estos intervalos son gamas que contienen el valor promedio de la característica que resulta de todas las posibles muestras, con una probabilidad conocida. Por ejemplo, si todas las muestras posibles que podrían resultar del diseño de la muestra de la Evaluación de Precisión y Cobertura del Censo 2000 se seleccionaron y se examinaron independientemente bajo las mismas condiciones, y si la estimación y su error estándar se calcularon para cada una de estas muestras, entonces:

1. Aproximadamente el 68 por ciento de los intervalos entre un error estándar estimado por debajo de la estimación y un error estándar estimado por encima de la estimación contendrían el resultado promedio de todas las posibles muestras.
2. Aproximadamente el 90 por ciento de los intervalos de 1.645 veces el error estándar estimado por debajo de la estimación hasta 1.645 veces el error estándar estimado por encima de la estimación contendrían el resultado promedio de todas las muestras posibles.
3. Aproximadamente el 95 por ciento de los intervalos dentro de los errores estándares estimados por debajo de la estimación y los errores estándares estimados por encima de la estimación contendrían el resultado promedio de todas las posibles muestras.

Los intervalos se conocen como intervalos de confianza de 68 por ciento, 90 por ciento, y 95 por ciento, respectivamente.

Puede que el valor promedio de la característica estimada que se podría derivar de todas las muestras posibles esté o no contenido dentro de cualquier intervalo particular calculado. Por lo tanto, no podemos aseverar que el valor promedio tiene cierta probabilidad de caer entre los límites del intervalo de confianza calculado. En su lugar, podemos decir con cierta probabilidad o confianza, que el intervalo de confianza calculado incluye el estimado promedio de todas las muestras posibles.

Para calcular los límites de un intervalo de confianza de 90 por ciento, utilice:

$$\begin{aligned} \text{El límite inferior del intervalo de confianza de 90 por ciento} &= \text{estimado} - 1.645 \times \text{SE} \\ &(\text{estimación}) \\ \text{El límite superior del intervalo de confianza de 90 por ciento} &= \text{estimado} + 1.645 \times \text{SE} \\ &(\text{estimación}) \end{aligned} \quad (6)$$

Para calcular los límites de un intervalo de confianza de 68 por ciento o 95 por ciento, sustituya 1.0 ó 1.96 (respectivamente) por 1.645 en la fórmula (6).

También se pueden construir intervalos de confianza para la proporción, suma de, o diferencia entre dos cifras de la muestra. Esto se hace primero calculando la proporción, suma, diferencia, y obteniendo el error estándar de la razón, suma, o diferencia (utilizando las fórmulas presentadas anteriormente), y finalmente formulando un intervalo de confianza para esta proporción, suma, o diferencia estimada según se explica anteriormente. Entonces podemos decir con cierta confianza que este intervalo incluye la proporción, suma, o diferencia que se hubiera obtenido al promediar los resultados de todas las muestras posibles.

Estos errores estándares estimados no incluyen todas las porciones de la variabilidad debido a los errores ajenos al muestreo que puedan estar presentes en los datos. Los errores estándares reflejan el efecto de variaciones simples en las respuestas, pero no el efecto de errores correlacionados introducidos por enumeradores, codificadores u otro personal de campo de procesamiento.

Por lo tanto, los errores estándares calculados representan un límite inferior del error total. Como resultado, los intervalos de confianza formados utilizando estos errores estándares estimados puede que no alcancen los niveles de confianza establecidos (por ejemplo, 68, 90 ó 95 por ciento). Por esto, en este producto de datos se debe ejercer cierta cautela en la interpretación de los datos basados en los errores estándares estimados.

Si el usuario necesita más información sobre intervalos de confianza y errores ajenos al muestreo, refiérase a un texto estándar sobre teoría de muestreo.

Ejemplos de cálculos de errores estándares

Suponga, por ejemplo, que un usuario tiene interés en saber el error estándar del estimado de la población de asiáticos en un sector censal. Una de las partidas de datos sobre la redistribución de distritos electorales es “Todas las personas, raza asiática solamente.” Suponga que los parámetros asociados de GCV es 0.0133.

Suponga que el estimado de la población para esta categoría de redistribución de distritos electorales es 370; entonces, de la fórmula (1),

$$\begin{aligned} SE(\hat{x}) &= GCV \times \hat{x} \\ &\doteq 0.0133 \times 370 \\ &\doteq 4.921 \approx 5 \end{aligned}$$

Utilizando la fórmula (6), un intervalo de confianza de 90 por ciento para el número total en el sector de aquellas personas que informaron su raza como asiática solamente sería

$$\begin{aligned} \text{límite inferior} &= \hat{x} - 1.645 \times SE(\hat{x}) \doteq 370 - 1.645 \times 4.921 \doteq 362 \\ \text{límite superior} &= \hat{x} + 1.645 \times SE(\hat{x}) \doteq 370 + 1.645 \times 4.921 \doteq 378 \\ \text{el intervalo de confianza de 90 por ciento es} & \text{ (362,378)} \end{aligned}$$

La fórmula (1) no se puede utilizar directamente para aproximar el error estándar de un estimado de personas no asiáticas menor de 18 años porque los parámetros de GCV sólo se publican para “Población total” y “18 años o más.” Para calcular el error estándar de ese estimado, se necesita utilizar la fórmula (2). Suponga que el número estimado de asiáticas de 18 años o más es 310. Así,

$$\begin{aligned} \text{Población menor de 18 años} &= \text{Población total} - \text{Población de 18 años o más} \\ &= 370 - 310 \\ &= 60. \end{aligned}$$

Del cálculo anterior, el error estándar del número total de “Todas las personas, raza asiática solamente” es 4.921. Utilizando la fórmula (1) otra vez y suponiendo que los parámetros para “Personas de 18 años o más, raza asiática solamente” es

$$GVC = 0.0108 \qquad \hat{x}=310$$

Este error estándar se puede calcular como 3.348. Utilizando estos valores y la fórmula (2)

$$\begin{aligned} SE(\text{menor de 18 años}) &= SE(\text{total} - 18 \text{ años o más}) = \sqrt{SE(\text{total})^2 + (SE(18 \text{ años o más}))^2} \\ &\doteq \sqrt{4.921^2 + 3.348^2} \\ &\doteq 5.952 \approx 6 \end{aligned}$$

Para calcular el error estándar de la proporción de asiáticos que tienen 18 años o más, podemos utilizar la fórmula (5):

$$\begin{aligned} \hat{x} &= \text{estimado de “Personas de 18 años o más, raza asiática solamente”} = 310 \\ \hat{y} &= \text{estimado de “Toda las personas, raza asiática solamente”} = 370 \end{aligned}$$

$$\begin{aligned}
 SE(\hat{p}) &= \frac{\hat{x}}{\hat{y}} \sqrt{\frac{SE(\hat{x})^2}{\hat{x}^2} + \frac{SE(\hat{y})^2}{\hat{y}^2}} \\
 &\doteq \frac{310}{370} \sqrt{\frac{3.348^2}{310^2} + \frac{4.921^2}{370^2}} \\
 &\doteq 0.014
 \end{aligned}$$

ERRORES AJENOS AL MUESTREO

En cualquier operación estadística de gran escala, como el Censo 2000, ocurren errores humanos y de computadoras. Estos errores frecuentemente se conocen como errores ajenos al muestreo. Tales errores incluyen el no enumerar cada hogar o a cada persona en la población, no obtener toda la información requerida de los respondedores, obtener información incorrecta o imprecisa y anotar información incorrectamente. Además, pueden ocurrir errores durante la revisión en el campo del trabajo de los enumeradores, durante el manejo de los cuestionarios en la oficina o durante el procesamiento electrónico de los cuestionarios.

Como es imposible eliminar completamente los errores ajenos al muestreo de una operación tan grande y compleja como el censo decenal, el Negociado del Censo intenta controlar las fuentes de tales errores durante las operaciones de recopilación y procesamiento. A continuación, se describen las fuentes principales de errores ajenos al muestreo y los programas establecidos para controlar este tipo de error durante el Censo 2000. Sin embargo, el éxito de estos programas dependió de cuán bien se siguieron las instrucciones en la realidad durante el censo. Como parte del programa de evaluación del Censo 2000, se evaluarán tanto los efectos de estos programas como la cantidad de errores que permanecieron después de su aplicación.

TIPOS DE ERRORES AJENOS AL MUESTREO

Falta de respuesta. La falta de respuesta a ciertas preguntas en el cuestionario del censo o no obtener cualquier información sobre una unidad de vivienda da base a que se introduzca sesgo en los datos, ya que no se han observado las características de las personas que no han respondido y pueden diferir de aquéllas provistas por los respondedores. Como resultado, cualquier procedimiento de imputación utilizando los datos del respondedor puede que no refleje estas diferencias ya sea al nivel elemental (persona o unidad de vivienda individual) o como promedio. El minimizar la falta de respuesta ofrece alguna protección contra la introducción de grandes sesgos en los datos. Las características para las preguntas sin respuestas fueron imputadas utilizando datos reportados para una persona o unidad de vivienda con características similares.

Errores del respondedor y del enumerador. La persona que contestó el cuestionario y lo devolvió por correo o respondió a las preguntas hechas por un enumerador para un hogar podría ser una fuente de error. Aunque la manera en que se frasearon las preguntas fue probada en varios estudios experimentales antes del censo, el respondedor que contesta por correo puede dejar pasar o malinterpretar una pregunta, o contestar una pregunta en una manera que no se puede interpretar correctamente por el sistema de la captura de datos. El enumerador también puede malinterpretar o anotar incorrectamente la información provista por el respondedor; o puede que no recopile alguna de la información para una persona u hogar, o puede recopilar datos para hogares que no habían sido designados como parte de la muestra. Para controlar problemas como estos durante la enumeración en el campo, se verificó el trabajo de los enumeradores cuidadosamente. El personal de campo se preparó para sus tareas utilizando adiestramientos estándares que incluían experiencia práctica en el uso de los materiales del censo. Una muestra de los hogares entrevistados por enumeradores fue entrevistada nuevamente para controlar la posibilidad de que los enumeradores estuvieran sometiendo datos falsificados.

Errores de procesamiento. Las muchas etapas en el procesamiento de los datos censales representan fuentes potenciales para la introducción de errores ajenos al muestreo. El procesamiento de los cuestionarios censales completados por los enumeradores incluyó la revisión en el campo por el líder de grupo, la verificación de recibo de los cuestionarios y el envío de cuestionarios completados. No hubo revisiones en el campo para los cuestionarios devueltos

por correo en este censo. También se puede introducir errores por la malinterpretación de datos por el sistema de la captura de datos o por no capturar toda la información que los respondedores o enumeradores proveyeron en las formas. Las respuestas escritas fueron procesados en operaciones de codificación, que también pueden ser una fuente de errores de procesamiento en los datos. Muchas de las operaciones de campo, codificación y computadora pasan por un número de revisiones para el control de la calidad para asegurar que fueron aplicadas correctamente.

Reducción de errores ajenos al muestreo

Para reducir los varios tipos de errores ajenos al muestreo, se utilizaron un número de técnicas durante la planificación, el desarrollo del listado de direcciones, la recopilación de datos y las actividades de procesamiento de datos. Se usaron métodos de verificar la calidad durante las fases de recopilación y procesamiento de datos para mejorar la calidad de los datos. Un programa de entrevistas suplementarias fue establecido para minimizar los errores en la fase de recopilación de datos para los cuestionarios completados por enumeradores.

Se establecieron varios programas para mejorar la cobertura durante el desarrollo del listado de direcciones y la enumeración y procesamiento del censo para minimizar la subcobertura de la población y unidades de vivienda. Estos programas se desarrollaron a base de experiencias del censo de 1990 y los resultados de la prueba del Censo 2000.

- Cuestionarios de Sea Usted Contado, formas sin direcciones que pidieron partidas de la forma corta del cuestionario, además de unas partidas adicionales, estaban disponibles en lugares públicos para personas que creían que no fueron contadas en el censo.
- Una carta de aviso se entregó a todas las direcciones en las áreas de actualización/entrega antes de que se entregaran los cuestionarios del censo. Una tarjeta postal recordatoria también se envió a estas direcciones.
- Se enviaron por correo cuestionarios en inglés a aquellas personas que los pidieron.
- Se le dió mucha publicidad a un número de teléfono libre de cargos que estaba disponible para contestar preguntas sobre los cuestionarios. Además, se podían aceptar contestaciones por teléfono de hogares que habían recibido la forma corta del cuestionario.
- En el programa de Actualización Local de Direcciones Censales (LUCA, por sus siglas en inglés), muchos gobiernos locales tuvieron la oportunidad de expresar sus preocupaciones específicas sobre la precisión e integridad del Archivo Maestro de Direcciones antes de que entregaran los cuestionarios.

Resolución al recibo de múltiples respuestas

Con las múltiples maneras en que personas podían empezar su enumeración y las operaciones de seguimiento en el campo, fue muy probable que algunas personas fueran enumeradas más de una vez. Se estableció un proceso especial de computadoras para controlar la extensión de estos tipos de errores ajenos al muestreo al resolver situaciones en las cuales más de un cuestionario se recibió de una dirección. El proceso consistió en varias etapas. Las direcciones que devolvieron más de una forma fueron analizadas. Los datos sobre hogares de un cuestionario fueron escogidos como los datos sobre hogares que se utilizarían en procesamientos subsiguientes del censo. Dentro de cada una de estas direcciones, se compararon los registros de persona en cada cuestionario devuelto con los registros de persona en los otros cuestionarios devueltos para la misma dirección. Las personas que se incluyeron en dos o más cuestionarios devueltos fueron marcadas como tal y sólo uno de los registros de persona se utilizó en procesamientos subsiguientes.

Imputación de la clasificación de unidades de vivienda y recuentos de población

Después de completar todas las actividades de recopilación de datos para el Censo 2000, se preparó un archivo de computadora de las unidades de vivienda del censo. Para algunas unidades de vivienda, no había información disponible sobre la clasificación de la unidad de vivienda, si estaba ocupada, desocupada o no existía. Se definieron estas unidades de vivienda como “no clasificadas.” Las unidades de vivienda no clasificadas fueron asignadas una clasificación de unidad de vivienda - ocupada, desocupada o no existente asignándoles la clasificación de una

unidad de vivienda cercana a la unidad no clasificada. Además, no se sabía el número de personas que vivían en algunas unidades de vivienda ocupadas. Se asignó el recuento de población de una unidad de vivienda ocupada cercana a las unidades de vivienda con una población desconocida. Todos los otros datos para estas unidades de vivienda fueron asignados por medio de sustitución o asignación durante la revisión de datos inaceptables que se explica en la sección a continuación.

REVISIÓN DE DATOS INACEPTABLES

El objetivo de la operación de procesamiento fue producir un conjunto de datos que describe la población lo más claro y preciso posible. En un cambio importante de prácticas anteriores, la información en los cuestionarios del Censo 2000 generalmente no se revisó durante la recopilación de datos en el campo ni durante operaciones de la captura de datos para asegurar su uniformidad, integridad y aceptabilidad. Líderes de grupo del censo y oficinistas locales revisaron los cuestionarios completados por enumeradores para verificar que se siguieron procedimientos específicos. No se hicieron revisiones en la oficina de los cuestionarios devueltos por correo para asegurar que la información en el cuestionario se podría capturar. Además, no se comunicaron con hogares para recopilar datos que faltaron de los cuestionarios devueltos como se hizo en censos anteriores.

La mayoría de cuestionarios recibidos por correo de respondedores, al igual que los completados por enumeradores, fueron procesados por un nuevo sistema de escáner creado por contratistas que utilizó reconocimiento óptico de marcas y caracteres para convertir las respuestas en archivos de computadora. El reconocimiento óptico de caracteres, o OCR, utilizó varias verificaciones de diseño y contexto para estimar los límites de precisión para cada campo con respuestas escritas. El sistema también permitió “soft edits” en la mayoría de respuestas numéricas que fueron escritas por una persona que no fuera el respondedor para decidir si los valores de los campos leídos por la máquina eran aceptables. Si el valor leído tenía un límite de precisión más bajo que el límite aceptado o si estaba fuera de la gama de “soft edit,” se mostraba la imagen de la partida a un oficinista que entonces entraba la respuesta.

Para controlar la creación de personas posiblemente erróneas de cuestionarios completados incorrectamente o que contenían marcas fuera de los encasillados, se hizo una revisión del número de personas indicadas en cada forma devuelta por correo y cada cuestionario completado por un enumerador como parte del sistema de la captura de datos. Si no pasaban esta revisión, un operador revisaba la imagen del cuestionario en su estación de trabajo e identificaba registros de personas erróneas y corregía errores de interpretación del reconocimiento óptico de caracteres en el campo de recuento de población.

En la oficina central del Negociado del Censo, los registros de datos sobre envíos por correo fueron sujetos a una revisión por computadora que identificó hogares que indicaron un problema posible de cobertura y aquéllos con más de seis miembros del hogar, el número máximo de personas que podían ser enumeradas en un cuestionario enviado por correo. Se trató de comunicarse con estos hogares por teléfono para corregir la falta de uniformidad del recuento y para recopilar los datos censales de aquellas personas para quienes no había espacio en el cuestionario.

Para información incompleta o no uniforme en los registros de datos de cuestionarios se asignó valores aceptables utilizando procedimientos de imputación durante la revisión automatizada final de los datos recopilados. Con frecuencia, se necesitan imputaciones, o asignaciones computarizadas de códigos aceptables en lugar de anotaciones inaceptables o en blanco, cuando falta una anotación para una partida dada o cuando la información reportada para una persona en esa partida no concuerda con otra información para esa misma persona. Este proceso se conoce como asignación. Al igual que en censos anteriores, el procedimiento general para cambiar anotaciones inaceptables fue asignar una anotación para una persona que concordara con anotaciones para personas con características similares. La asignación de códigos aceptables en lugar de anotaciones en blanco o inaceptables aumenta la utilidad de los datos. Las tasas de asignación para partidas censales están disponibles con los datos censales publicados.

Otra manera en que se hicieron correcciones durante el proceso de revisión automatizada fue a través de la sustitución; es decir, la asignación de un conjunto completo de características para las personas en un hogar. Cuando había indicación de que una unidad de vivienda estaba ocupada por un número de personas especificado pero el cuestionario no contenía la información para las

personas dentro del hogar o cuando los ocupantes no estaban listados en el cuestionario, se seleccionó como sustituto un hogar del mismo tamaño y se duplicó el conjunto completo de características del sustituto.