Preguntas y respuestas
IPUMS International: Una Introducción
Abril 25, 2019 (11:00 a.m.-12:00 p.m. CT)

Características de los datos

1. **¿Se puede acceder a los metadatos a través de IPUMS? Por ejemplo, me gustaría conocer el *sampling design* de las encuestas.**

El sistema de diseminación de datos en línea de IPUMS permite acceder a los metadatos y documentación del censo, tales como las etiquetas de variables y categorías, los cuestionarios, o las instrucciones para los enumeradores. El diseño de las muestras de los censos en IPUMS International también se encuentra disponible en línea y se puede acceder en: https://international.ipums.org/international-action/sample_details

2. **¿A qué se refiere con muestra? ¿No están los censos completos? Pero en cuyo caso, ¿se tienen los factores de expansión?**
3. **El uso de 10% de los datos censales, ¿implica que el uso de datos desagregados a nivel individual requiere ponderar el manejo de información?**

Los microdatos de censos en IPUMS International son muestras (no el censo completo), por razones de confidencialidad estadística. En general, se realiza una selección sistemática simple de 1 de cada 10 hogares (la muestra incluye a los hogares seleccionados y a todas las personas en dichos hogares). En aquellos datos en que se aplicó este método de muestreo, el factor de expansión es constante e igual a 10. Por otro lado, las muestras de algunos censos cuentan con un diseño complejo, en cuyo caso el método de muestreo y los pesos vienen definidos por el Instituto Nacional de Estadística correspondiente. La información sobre el diseño de las muestras y los factores de expansión aplicables se encuentra disponible en el sistema de diseminación en línea.

4. **En las muestras de los países donde se realizó una selección sistemática, ¿qué variables se tuvieron en cuenta en el orden de las unidades previo a la selección?**

El orden de los datos en los cuales se aplicó una selección sistemática de hogares corresponde a las variables de geografía censal y viene definido por cada Instituto Nacional de Estadística. De este modo, el muestreo sistemático simple permite representar a todos los hogares en un censo/país específico.

5. **¿Podrían explicar mejor lo que significa "*density*"?**

Density se refiere a la fracción de muestreo para el censo específico. Por ejemplo, los datos de las muestras censales en las que se seleccionó 1 de cada 10 hogares tienen una densidad ("*density*") de 10%.

6. **Las muestras están al 10% del censo, ¿son paneles de los mismos individuos dentro del mismo país/región a través de los años, o son diferentes muestras de individuos cada año?**

Los microdatos son anonimizados y no permiten la identificación de hogares o personas, por razones de confidencialidad estadística. Más aún, la selección de la muestra se realiza de forma independiente para cada censo. En la práctica, nada garantiza que los hogares o personas incluidos en dos muestras de un mismo país tengan un alto nivel de coincidencia. En ese sentido, los microdatos en IPUMS son de corte transversal.

Los microdatos permiten, sin embargo, realizar agregaciones para incorporar un componente longitudinal en el análisis. Por ejemplo, es posible hacer seguimiento a indicadores calculados para unidades administrativas en el primer o segundo nivel de geografía a lo largo del tiempo.

7. **¿En el caso de bases de datos de encuestas también se dispone sólo del 10% de los casos, o eso aplica sólo para bases de censos?**

Las encuestas de hogares en IPUMS International cuentan con la totalidad de la muestra correspondiente. El muestreo se aplica sólo a las bases de censos. Cabe resaltar que, en la actualidad, la colección de datos del proyecto incluye mayoritariamente a censos de población.

8. **¿Cómo conozco el nivel de error muestral con el que trabajo? Dado que estoy trabajando con una muestra y no el total del universo.**

IPUMS International pone a disposición de los usuarios la información relevante sobre el diseño muestral, incluido el tipo de selección muestral realizada y el número de unidades en cada muestra. Las decisiones sobre los métodos aplicables para el cálculo de estadísticos muestrales a partir de los datos (v.g. intervalos de confianza para un estimador muestral) deben basarse en dicha información. Los usuarios también pueden consultar los reportes censales producidos por los Institutos Nacionales de Estadística, si se desea realizar comparaciones con respecto a características de la población total (universo) correspondiente.

9. **¿Estamos hablando solo de Censos Nacionales de Población o también, por ejemplo, de Censos Nacionales Agropecuarios?**

Actualmente la colección de datos incluye únicamente censos de población. No se ha trabajado, por el momento, con censos agropecuarios.

Disponibilidad de países

10. ¿Cuáles son los países latinoamericanos que se encuentran (o que no se encuentran) en las bases de datos?

La colección de datos en IPUMS International cuenta actualmente con 23 países de América Latina y el Caribe. El listado detallado de países disponible se puede encontrar en: https://international.ipums.org/international-action/sample_details

11. ¿Puerto Rico está incluido como país independiente? ¿O debo buscar sus datos bajo US?

Los datos de Puerto Rico se han procesado de forma independiente. La información sobre las muestras censales de Puerto Rico está disponible en: https://international.ipums.org/international-action/sample_details/country/pr#tab_pr1970a

12. Tratando de replicar la demostración en búsqueda de datos, sólo me aparecen datos de USA y Puerto Rico, ¿cómo puedo cambiar este *display*?

El sistema de diseminación de datos permite seleccionar muestras específicas y mostrar información únicamente para los datos seleccionados. Para cambiar la selección activa e incluir países/años diferentes, se debe ir a "*change samples*".

13. ¿Tienen relación con el Instituto de Estadística de Perú? ¿Se cuenta con microdatos de todos los censos realizados?

El Instituto Nacional de Estadística de Perú (INEI) es uno de los institutos asociados a IPUMS International. Los datos en línea incluyen muestras de los censos de 1993 y 2007.

14. ¿Está España incluida entre los países participantes?

El Instituto Nacional de Estadística de España (INE) es también un instituto asociado a IPUMS International. Se puede encontrar en línea muestras de los censos de 1981, 1991, 2001, y 2011.

15. En algunos casos (Argentina, por ejemplo) se toman censos con una parte del cuestionario aplicado a través de una muestra. Los casos de ocupación provienen de la muestra. ¿Podrían indicarnos si esa información está disponible en la muestra de Argentina?

Los censos de Argentina contaron con un cuestionario básico y un cuestionario ampliado en 1980, 1991, y 2010. Los datos en IPUMS International incluyen la información del cuestionario ampliado en 1980 y 1991, en tanto que la muestra del censo de 2010 corresponde al cuestionario básico.

Características de las variables

16. ¿Qué temas son los que tienen más variables homologadas en las encuestas?

La disponibilidad de variables en IPUMS International corresponde a las preguntas que esperaríamos formen parte de un cuestionario censal; por ejemplo, características demográficas básicas, empleo, educación, fertilidad, mortalidad, migración, entre otras. IPUMS produce versiones homologadas para la mayoría de variables que se encuentren disponibles en más de una muestra, para un mismo país o varios países.

El listado detallado de variables homologadas se puede encontrar en el sistema de diseminación de datos, organizado según grupo temático:
https://international.ipums.org/international-action/variables/group

17. ¿Puedo hacer cruces de información entre variables de diferentes categorías a partir de estos datos? Es decir, matrimonios por nivel educativo, por ejemplo.

IPUMS International realiza diseminación de microdatos de muestras censales, lo cual permite hacer cruces de información entre variables. Por ejemplo, se puede producir una tabla cruzada del estado civil y el nivel educativo para un censo específico a partir de un extracto de datos en el que se incluya ambas variables y usando el paquete estadístico de su preferencia (o, alternativamente, con la herramienta de análisis de datos en línea).

18. ¿Incluyen las variables de pertenencia étnica y/o racial para aquellos países que las incluyen en sus censos? ¿Existe información de pueblos indígenas y lenguas propias para cada país?

Las variables de identificación según grupo étnico y/o racial se incluyen en IPUMS International si se encuentran en el cuestionario del censo específico de interés y siempre que hayan sido incluidas en la muestra por el Instituto Nacional de Estadística correspondiente. IPUMS realiza también la homologación de datos para dichas variables.

Para mayor detalle sobre censos específicos, se puede consultar el grupo de variables "*ethnicity and language*" en el sistema de diseminación de datos:
https://international.ipums.org/international-action/variables/group?id=ethnic

19. Mi campo es la movilidad social (clase, ocupacional, educativa) por lo cual se precisa datos del encuestado y del progenitor, ¿están disponibles los datos del progenitor?

Los usuarios pueden realizar análisis de las características del padre de un miembro del hogar, siempre que éste sea co-residente. IPUMS genera variables "*pointer*" que permiten identificar el número de orden en el roster del hogar para personas específicas que sean co-residentes. Los usuarios tienen a su disposición variables "*pointer*" para el padre (POPLOC), la madre (MOMLOC), y la pareja (SPLOC). IPUMS construye dichas variables "*pointer*" a partir de

características observadas de los miembros del hogar, tales como la relación con el jefe del hogar, el sexo, o la edad. En algunos censos, la información sobre la posición de la madre o del padre en el roster del hogar se pregunta directamente en el cuestionario.

Por ejemplo, si estamos interesados en el nivel educativo, un usuario puede trabajar con la variable "*pointer*" para el padre (POPLOC) y crear un indicador para el nivel educativo del padre, siempre que ambos individuos residan en el mismo hogar.

20. ¿Si elijo RELATE para *Spouse*, entonces te da la relación con respecto a la esposa?

Entre las opciones para personalizar un extracto de datos, el sistema de diseminación ofrece la opción de vincular características ("*attach characteristics*") para ciertos miembros del hogar. En efecto, es posible identificar la relación con el jefe del hogar para una persona identificada como "pareja" (en SPLOC). Sin embargo, las variables "*pointer*" identifican no solamente a la pareja del jefe, sino también de otros miembros del hogar.

Por ejemplo, consideremos un hogar en el que co-residen el jefe, su pareja, un hijo, y la pareja de dicho hijo. En esta situación, las variables *pointer* identifican a dos parejas en el hogar, la pareja del jefe y del hijo del jefe. Entonces, sería posible determinar la relación de cada una de dichas parejas con respecto del jefe del hogar.

21. ¿Publican documentos sobre la metodología de homologación?

El proceso de homologación de variables sigue la mecánica general presentada en el *webinar*. Para mayor información sobre variables homologadas específicas, los usuarios tienen a su disposición los metadatos en línea, que incluyen la estructura de códigos homologada y el texto que discute la comparabilidad de los datos entre muestras. Además, se pueden comparar los códigos de las variables fuente ("*source variables*") y las variables homologadas ("*harmonized variables*") en una tabla cruzada, con el fin de determinar qué decisiones de homologación se tomaron para códigos específicos de la variable original del censo.

22. ¿Los umbrales para las diferentes preguntas están homologados?

Las variables homologadas son construidas a partir de una codificación consistente entre países y años censales. En ese sentido, los "umbrales" para las diferentes preguntas se encuentran homologados, si entendemos los mismos como la definición de categorías específicas entre los censos.

23. ¿Qué tratamiento se le dio a los ignorados en cada variable? ¿Se respeta el dato enviado por cada INE?

Los datos son homologados respetando la información confiada a IPUMS por cada Instituto Nacional de Estadística. IPUMS no realiza ninguna imputación de datos para los casos de información ignorada.

Durante el procesamiento de datos, únicamente se realiza una distinción en los casos de no respuesta entre las personas/hogares fuera del universo (que no deberían consignar una respuesta) y aquellas dentro del universo de la pregunta censal (que constituyen auténticos casos de información ignorada). Por ejemplo, si la pregunta censal sobre estado civil estaba destinada a personas de 12 años o más, se distingue entre personas menores de esa edad (que se codifican como "fuera del universo" o "*not in universe*") y los casos realmente ignorados (que se codifican como "ignorado" o "*unknown*").

Variables de geografía

24. En algunas ocasiones las divisiones geopolíticas de los países han cambiado a lo largo del tiempo, y esto afecta la comparación entre regiones entre un censo y otro. ¿las divisiones territoriales en IPUMS corrigen por esto?

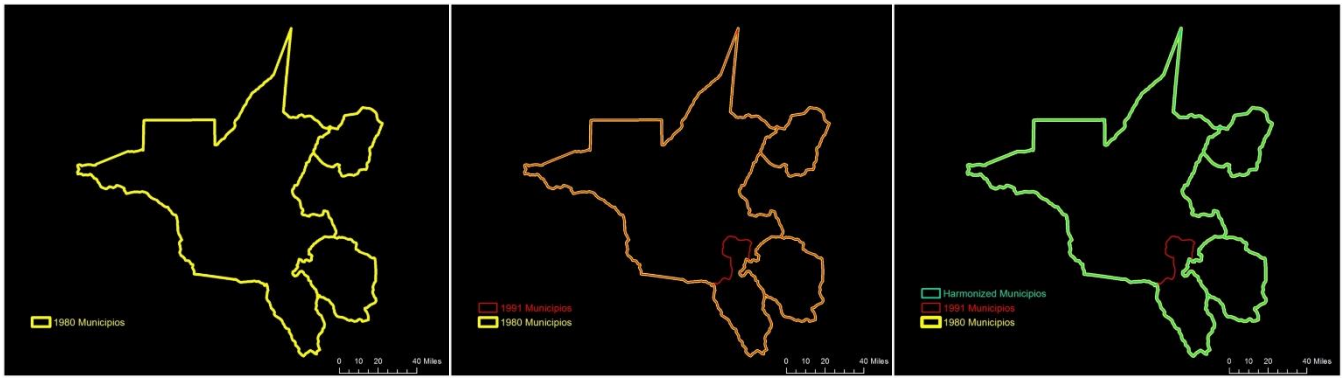
IPUMS realiza la homologación de divisiones administrativas a través del tiempo para los países representados en la colección de microdatos. La homologación de variables de geografía identifica unidades administrativas consistentes en el tiempo, a partir de la agrupación de unidades similares y que son contiguas entre sí. De este modo, es posible corregir las distorsiones que generan los cambios en las divisiones administrativas a lo largo del tiempo. El siguiente documento de trabajo contiene mayor detalle sobre el proceso de homologación de geografía:
https://pop.umn.edu/sites/pop.umn.edu/files/3.workingpaperall.pdf

Por ejemplo, se puede ver el resultado de la homologación del segundo nivel de geografía para Sudáfrica en el video de la presentación del *webinar*. En dicho ejemplo, se observa 45 unidades administrativas en 1996, mientras que existen 52 en 2007 y en 2011. La geografía homologada en el segundo nivel administrativo para Sudáfrica cuenta únicamente con 26 unidades administrativas, que son aquellas que se pueden identificar de forma consistente en 1996, 2007, y 2011.

25. ¿La homologación geográfica crea una unidad nueva o simplemente se agrupa en la unidad superior?

La homologación geográfica agrupa unidades contiguas hasta encontrar un espacio consistente y que pueda ser identificado a lo largo del tiempo. En ese sentido, las unidades geográficas homologadas no son estrictamente "nuevas" (no se traza nuevos límites) sino se trata de unidades más pequeñas que son agrupadas entre sí. Con este fin, se empieza por identificar los cambios en las divisiones administrativas del país en el tiempo, a partir de la geografía representada en cada operación censal.

Por ejemplo, en el gráfico incluido a continuación, observamos que un municipio se divide entre los censos de 1980 y 1991. En la geografía homologada, se requiere agrupar nuevamente dichas unidades (que se han separado) para identificar un espacio consistente en el tiempo.



26. ¿Entonces es posible que se pierdan algunos municipios al momento de hacer un comparativo entre bases de datos históricas y actuales? ¿Toman en cuenta el cambio de nombre de los municipios?

En la medida que el proceso de homologación de geografía implica agrupar unidades hasta encontrar un espacio consistente en el tiempo, es posible que algunas unidades administrativas específicas no puedan ser identificadas de forma independiente en la geografía homologada y sólo se puedan analizar como parte de una unidad agrupada. El proceso de homologación considera los diferentes cambios en las unidades administrativas de un país entre diferentes censos, incluidos los cambios de nombre de cada unidad.

No obstante, los usuarios tienen a su disposición no sólo las variables de geografía homologadas, sino también las específicas a cada censo. En ese sentido, las variables específicas a un censo permiten analizar ciertos municipios que no se pueden identificar de forma independiente en la versión homologada. Por ejemplo, en el caso de Argentina, se cuenta con provincias (primer nivel) y departamentos (segundo nivel) consistentes en el tiempo, pero también con las unidades existentes para ambos niveles de geografía en los censos de 1970, 1980, 1991, 2001, y 2010.

27. Para la muestra sobre Argentina, ¿el nivel de detalle de los microdatos incluye a municipios o barrios?

Las muestras de censos de Argentina permiten identificar únicamente provincias (primer nivel) y departamentos (segundo nivel).

28. Si se quiere comparar un dato en el tiempo, ¿es mejor trabajar con el segundo nivel de agrupación? Yo trabajo con datos para México, en el caso de datos climáticos me interesan en la mínima unidad sin embargo la serie que trabajo viene desde 1970 a 2013, ¿cómo se manejan los datos cuando no hay dato, para algún municipio en uno de los años ¿ponen ceros, como NA?

Las muestras de México permiten identificar estados (primer nivel) y municipalidades (segundo nivel). Si un usuario desea trabajar con datos complementarios (como información sobre clima), el nivel administrativo relevante para el análisis dependerá de la disponibilidad de los datos complementarios desagregados para dichos niveles (estado o municipalidad).

En caso algún municipio no esté representado en los microdatos, simplemente no encontraremos hogares o personas que declaren residir en dicho municipio. Si algún hogar o persona no cuenta con información sobre su municipio de residencia, recibiría el código correspondiente a una respuesta ignorada ("*unknown*").

29. Un censo no es una encuesta estadística. Usando los datos de IPUMS, ¿puedo hacer inferencias a nivel de distrito, provincia, estado o solo país? ¿Cuál es la mínima unidad política-administrativa a la que las muestras son representativos?

Los microdatos en IPUMS International son de alta densidad y permiten realizar desagregaciones más finas. En la mayoría de casos, las muestras de censos corresponden a un 10% de la población y pueden contar inclusive con varios millones de registros de personas. El nivel mínimo de la división político-administrativa disponible en los datos depende del país, pero se puede acceder generalmente a 2 niveles de geografía. Los usuarios deberían realizar sus propios cálculos para determinar la representatividad estadística en cada uno de los niveles disponibles, pero el tamaño de muestra debería ser apropiado, en general, para trabajar inclusive con unidades en el segundo nivel de geografía administrativa.

30. Entonces, ¿todos los países que tienen en la colección de datos también tienen *shapefiles*/geografía?

La información detallada sobre geografía se encuentra descrita en la sección de "*Geography and GIS*", que incluye una mirada general al proceso de homologación de geografía, los *shapefiles* disponibles para descarga, un listado de todas las variables de geografía para cada censo, y las equivalencias con unidades de geografía en las *Demographic and Health Surveys* (DHS). En general, los usuarios encontrarán disponibles hasta 2 niveles de geografía administrativa, variables de geografía homologadas y variables de geografía específicas a cada censo, así como los *shapefiles* correspondientes para trabajar con cada una de estas variables.

La sección de "*Geography and GIS*" se puede consultar en:
https://international.ipums.org/international/geography_gis.shtml

31. ¿Es posible realizar *merge* entre los microdatos de la base de datos de mujeres de las DHS de Macro International?

Los microdatos en IPUMS International se pueden analizar conjuntamente con los datos disponibles a través de IPUMS DHS. El proyecto IPUMS DHS realiza diseminación de datos homologados de las *Demographic and Health Surveys* (DHS).

Para mayor información sobre cómo conectar la geografía representada en ambas fuentes de datos, se puede consultar el *crosswalk* de IPUMS International e IPUMS DHS en: https://international.ipums.org/international/gis_crosswalk.shtml

32. ¿Existe información geográfica (GIS, por ejemplo)? ¿Uso de suelo o similares?

El proyecto IPUMS Terra cuenta con información de población y medio ambiente. Para mayor detalle sobre las variables específicas disponibles, los usuarios pueden consultar: https://terra.ipums.org/

Sistema de extractos

33. ¿Hay un listado agregado de todas las variables y *codebook*, o hay que entrar en el sistema para buscar las variables?

Los usuarios pueden consultar la disponibilidad de variables específicas para las muestras de censos de su interés en el sistema de diseminación de datos, ya sea navegando los grupos temáticos de variables o a través de la opción de búsqueda por palabras clave. Los códigos, etiquetas, universos, y el texto del cuestionario se encuentran disponibles en la página de la variable correspondiente en el sistema de diseminación en línea.

En ese sentido, no es necesario revisar un listado de variables o libro de códigos, dado que toda la información relevante es fácilmente accesible en línea. No obstante, si lo desean, los usuarios pueden acceder a un libro de códigos para un extracto de datos específico entre las opciones de descarga.

34. Aparece en la pantalla una opción para el paquete R, ¿vienen los scripts?

Los datos de IPUMS requieren instalar el paquete ipumsr para que se puedan leer en el software estadístico R. El usuario simplemente debe seguir las instrucciones que aparecerán en su pantalla una vez creado un extracto de datos (hacer *click* en “R”).

Para mayor información sobre el paquete ipumsr, se puede consultar: https://cran.r-project.org/web/packages/ipumsr/index.html

35. ¿Cómo funciona el análisis en línea?

La herramienta de análisis de datos en línea (SDA) permite realizar una variedad de análisis estadísticos completamente en línea, sin la necesidad de descargar datos. La herramienta permite analizar ya sea un único censo, todos los censos de un único país, o todos los censos de una región del mundo en la colección de IPUMS. Por ejemplo, se puede tabular variables, comparar medias, calcular correlaciones, o correr regresiones. Para trabajar con la herramienta de análisis de datos en línea se requiere ser usuario registrado de IPUMS.

Para mayor información, la herramienta de análisis de datos en línea se puede encontrar en la siguiente página:
https://international.ipums.org/international/sda.shtml

Otros

36. ¿Esta presentación se guarda automáticamente y la puedo volver a ver más tarde?

La presentación se encuentra disponible en *Youtube* en: https://youtu.be/fl5w3rm7BLc

37. ¿Los datos se pueden usar en una investigación periodística o es considerado comercial?

Los datos se pueden usar en una investigación periodística, siempre que no sea con fines de lucro. El uso específico que se dará a los datos debe ser descrito con precisión en el formato de inscripción como usuario de IPUMS. Para mayor información, se puede consultar las condiciones de uso en el formato de inscripción para los usuarios de IPUMS International:
https://uma.pop.umn.edu/ipumsi/user/new?return_url=https%3A%2F%2Finternational.ipums.org%2Finternational-action%2Fmenu

38. ¿Se pueden hacer aportes públicos a IPUMS de las investigaciones que se hagan de forma personal?

Las investigaciones realizadas con datos en IPUMS se pueden encontrar en la bibliografía en línea. Si los usuarios realizan investigaciones usando IPUMS, les pedimos que envíen dicha información para que pueda ser agregada a la bibliografía en línea. Las citas se pueden agregar directamente a la bibliografía en línea (https://bibliography.ipums.org/) o se puede enviar copias de sus investigaciones a ipums@umn.edu.