

**10 Tips for obtaining census samples and documentation
from IPUMS-International <https://international.ipums.org/international>
IPUMSi is an initiative of the Minnesota Population Center (MPC)
with major funding by the University of Minnesota,
National Science Foundation, and National Institutes of Health (USA)**

June 2010

10 Tips—see next page.

Researchers: please make good use of the IPUMSi system, so National Statistical Offices will continue to entrust census microdata for the IPUMS-International database!

Goals:

1. preserve census microdata and documentation from the 1960s to the present
2. integrate anonymized, high-precision samples and corresponding documentation
3. disseminate census microdata and metadata to researchers world-wide, without cost

Accomplishments, 1999-present:

- 85 National Statistical Offices endorsed the IPUMSi memorandum of understanding.
- 270 sets of census microdata entrusted
- 158 samples (55 countries, 325 million persons) integrated and available to researchers
- many datasets recovered from old tapes, primitive disc drives and even floppy discs!
- over 3,300 registered researchers
- many conference papers, articles, theses and books are already in preparation

How does IPUMSi work?

- A vast body of census microdata and documentation is being assembled at the University of Minnesota Population Center, thanks to a global collaboratory of National Statistical Offices, international organizations, universities and researchers.
- Accredited researchers obtain free access to all documents and anonymized samples.
- Microdata and documents are integrated to facilitate comparative research.
- Each user extracts a selection of microdata, downloads and analyzes it using his/her own software and computer resources.
- Researchers provide copies of publications to IPUMSi for transmittal to National Statistical Office partners.

Why aren't IPUMSi datasets available on DVDs or by FTP?

- The database is expanding rapidly, with the inclusion of data and documents for more countries, censuses, and variables.
- The database is already too huge to ship on a CD or DVD like a book, video or other products obtained from the web
- The dynamic metadata system helps navigate both integrated and original source documentation and cannot be easily replicated on other media.
- New samples, variables, and enhancements are added frequently--the database is dynamic.
- Note where internet access is limited National Statistical Office partners are provided, upon request, a DVD containing an integrated set of samples, variables and documentation for their country to be disseminated to national researchers.

10 Tips: browse <https://international.ipums.org/international>:

Tip 1. Register to use the microdata--Click ""User Registration and Login" on the IPUMSi home page, then click "Apply for Access," submit email address, and complete the form. Check each box to show that you will protect the data against unauthorized use.

Tip 2. Protect statistical confidentiality.

- Faithfully abide by each of the conditions of use as stated on the registration form.
 - Do not "share" IPUMSi extracts with others, unless they are duly registered users (such as co-researchers on a project).
- Violations will lead to denial of access for the researcher and recall of all microdata. In cases of lax oversight, access may be denied to all users at the researcher's institution.
 - Please report suspected violations to ipumsi@pop.umn.edu

Tip 3. Study the documentation thoroughly.

- Comprehensive documentation is provided in the official language
Tip: click "Census Questionnaires" on the IPUMSi home page then image ("pdf") for census forms and instructions for the desired country and census year.
- English texts and translations are available for all documents:
Tip : click "html" on the "Census Questionnaires" page.
Tip: click "enumeration text" on any variable description page to invoke the dynamic metadata system (see "Features" below)
- IPUMSi metadata describe, for each country:
 - censuses - title, agency, population enumerated, universe, type, unit, field work period, forms used, etc.
 - samples - source, design, sample unit, fraction, size, weights, etc.
 - variables - availability, universe, codes, enumeration text, unharmonized codes, description, comparability, etc.

Tip 4. Construct extracts judiciously

- Request only microdata samples needed for a specific research project
 - Some samples contain 10 million or more person records, if 100,000 cases are sufficient, request that number (see "customize sample size" below)
Tip: extracts may be reviewed and revised (see "Download Data Extract" on the IPUMSi home page to see a menu of your previous extracts.)

Tip 5. Use weights (expansion factors) where necessary.

- **Tip:** not all samples have the same design. To avoid incorrect inferences, use weights.

Tip 6. Analyze carefully

- Use proper statistical techniques.
 - Take into account bias and errors in census operations as well as sampling errors.
- IPUMSi sample statistics may differ from official figures for many reasons (due to loss of original data, omission of certain classes of individuals such as residents of collective households, etc.).

Tip 7. Cite properly: Cite both IPUMS-International and the respective National Statistical Agency(ies). See list below and "International Partners" on the IPUMSi home page.

Tip 8. Share publications: Post citations on the IPUMSi home page (click "Bibliography"). Email copies to ipumsi@pop.umn.edu for forward to the respective National Statistical Agencies. Failure to do so may lead to denial of access for you & other users at your institution.

Tip 9. Click "FAQ" on the IPUMSi home page for "Frequently Asked Questions"

Tip 10. Email ipumsi@pop.umn.edu for help with questions, problems or errors.

Features of the IPUMS-International website.

<https://international.ipums.org/international>

A. IPUMS-International dynamic metadata system (open access; no registration required):

For a tutorial, click: www.hist.umn.edu/~rmccaa/IPUMSI then click "Dynamic Metadata"

- Note that while browsing, click any box (similar to the one that begins this line) to select that item (sample or variable) to include in an extract. Click any checked box to de-select.
- 1. Click "samples" for descriptions of the samples currently available, from Argentina to Vietnam, including microdata for Cambodia, Canada, China, Colombia, Costa Rica and Cuba--to name only the "C"s.
- 2. From the IPUMS-International home page, click "variables," for descriptions of available variables.
- 3. From the "variables menu" page:
 - a. click "select samples" to choose samples for the desired countries and census years.
 - b. click "person", "dwelling", or the alphabetic list to see a listing of variables.
 - c. if unharmonized variables are of interest, toggle "switch to unharmonized variables"Note: unharmonized variables are unique to a single census and are similar to the original source variables. They are useful for checking integrated variables. Note, too, that variables that are exclusive to a single census are not integrated, and consequently are available only in non-harmonized form.
- 4. From a list on the variables page (e.g., "Core Demographic"), click any variable name to view metadata.
- 5. From this or any variable description page:
 - o **Tip:** click "codes," then "case-count view" for counts (number of person or households) in each sample
 - o **Tip:** click "enumeration text" to compare the wording and instructions to enumerators for the corresponding variable and selected samples. **Compare questions for any combination of countries and censuses in the database. Use this tool to ascertain whether integrated variables and codes are appropriate for your research needs.**

B. Register for Access to the microdata (click "User Registration and Login").

- Note that with IPUMSi, the researcher downloads the microdata as "extracts".
- IPUMSi extracts are analyzed using the statistical software and computer resources familiar to the researcher.
- The advantage of extracts is that the researcher selects:
 - the desired countries, census years, sample sizes, sub-populations, and variables.
- To access data, first, click "apply for access", then complete the registration form, briefly describe the research plan, and agree to the terms and conditions (protect statistical confidentiality of the microdata, safeguard the microdata, provide a copy of publications, etc.). Read each condition of use and click each box to indicate your agreement to abide by each restriction. The registration form is quite detailed to restrict access to bona-fide researchers. Approval is usually forthcoming in 1-7 days.

C. Create an extract (restricted to bona-fide, registered researchers)

- The IPUMSi database is huge with hundreds of millions of records, thousands of variables, and more than one-hundred samples.
- 15-20 new samples are added each year.
- The IPUMSi extract engine makes it easy for each user to select precisely the samples, variables and records important to your research.

(continued)

Registered users obtain data by clicking "Create Data Extract" and completing 6 steps: <https://international.ipums.org/international>

Step 1. Login: enter email address and password.

Step 2. Select samples, countries and census years:

- a. click as many boxes on the sample selection menu as needed
- b. click "continue"

Step 3. Select variables

- a. Click "add more variables"
- b. Roll mouse over "household", "person", or "A-Z" button, and click a selection
- c. Click the box for each variable of interest
- d. Repeat b and c until all variables of interest are selected
 - **Tip:** to see all variables in a single menu, roll mouse over "Options and Help" and click "view one group"
- e. If unharmonized variables are of interest, click "switch to unharmonized variables"
 - **Tip:** a unique set of variables will be offered for each census (e.g., seemingly similar variables, "age", "sex" are, nevertheless, designated with unique variable names)
- f. Once all variables of interest are selected, review all by clicking "Review variable selection".
- g. Click "add more variables" (see "a" above) or "continue to next step"

Step 4. "Variable options": select cases for inclusion/omission and attach characteristics.

- **Tip: This step is optional.**
- a. Click "select cases" for any variable to be used to include/exclude cases in the extract (e.g., click "sex" to be able to select only females for inclusion in the extract)
 - b. Click "attach characteristics" of household head, mother, father and/or spouse for any selected variable.
 - c. click "continue to next step"
 - d. attach characteristics: of household head, mother, father or spouse; then click "continue to next step"
 - e. "case selection"
 1. choose the conditions to be included in the extract (e.g., click female to include only females in the extract)
 2. to include in the extract all co-resident persons of the selected case, click "include all persons in the household"
 3. click "continue to next step"

Step 5. Customize sample sizes. Tip: ask yourself the question: "Is the projected extract file size (in megabytes) too big for what I want to do?" .

- a. Click "continue to next step" (and skip this section), if the projected size of the dataset is OK
- b. Otherwise, click "customize sample sizes", if you wish to reduce the size of the samples:
 - **Tip:** to make all the samples the same size, enter a number in the first row. If the requested size is larger than the maximum, the maximum is automatically entered.
 - 1. enter a number in any of the boxes--either for the number of cases or % density of each sample
 - 2. to estimate the projected extract file size, click "recalculate"
 - 3. to restore the maximum projected size, click "reset selections"
 - 4. when finished, click "continue to next step"

(continued)

Step 6. Submit selections to construct the extract. Before submitting the selections, review:

- a. Review the selections of samples, variables, options, attached characteristics, case selection, and sample sizes
 - **Tip:** at this stage, as in all previous stages, click any high-lighted variable name to recall definitions, descriptions, etc.
- b. If any changes are desired, click the appropriate item:
 - Samples: to change the sample selections
 - Variables: to change the variable selections
 - Variable options: to remove/add case selections or attached characteristic variables
 - Attached characteristics: to change selections of head, spouse, mother, father
 - Sample size: to change customized sizes
- c. **Tip:** in the text box, type a brief description or name for this extract, as a reminder.
 - The extract will be registered with IPUMS
 - If you need to recall an extract to add samples, variables, etc., you may do so by clicking “Download Data Extract” on the home page. Log in (email and password) to see a list of your extracts, including date and description.
 - Thus, as 2010 round samples are added to IPUMS, it will be easy to update your own integrated database by revising a previous extract.
- d. Finally, "Submit extract" to complete the request.

When the extract is ready, an email will be sent to the registered address.

Click the link in the email, enter your email and password and download the data extract, codebook, and SAS, STATA, or SPSS dictionary.

Unzip the data and construct the database with the corresponding dictionary. **Tip:** for SPSS, enter in the command file (xxx.sps) the folder information for the exact location of the data file.

New in 2010:

- Earn a Red Wing Pottery® “Tulip”, “Obama” or classic “China” mug, with the IPUMS-International logo, by teaching 30 or more students to use IPUMS-International samples (students must register name of instructor in project description to be eligible).
- Click “Bibliography”: to search/add citations using IPUMS-International integrated samples
 - Click “Bibliography” from the home page (153 items as of Jan, 2010). To limit search, click “search by project” then check “IPUMS-International”.
 - Add citations to the IPUMS-International bibliography: click “add your publication” from the “MPC Bibliography” page.
- Click “GIS and other data files” on the IPUMS-International home page.
 - Boundary files for each country (Note: these files are unofficial and do not track changes in boundaries over time).
 - Mortality, migration and fertility records documenting demographic events for households for a period preceding a census.

More is coming to IPUMSi! Plans, 2011-

- expand archiving
 - backward to encompass all published census documents in addition to unpublished
 - forward to include scanned images and other forms of completed questionnaires
 - outward to provide easy access to census microdata and documentation, world-wide
- integrate 2010 round censuses on a top priority when the microdata become available
- continue integration of samples for participating countries (see list below)
- invite other national statistical offices and organizations to participate in IPUMS
- increase access to more researchers

- enhance documentation with modules on sampling errors, official publications, etc.
- implement a tabulator of integrated samples (to assist with extract requests)
- expand the MPC's Secure Data Laboratory for access to microdata that cannot be disseminated over the internet

Please email questions or comments to: ipumsi@pop.umn.edu or rmccaa@umn.edu Additional information is available at: www.hist.umn.edu/~rmccaa/IPUMSI

IPUMS International Partners

85 National Statistical Offices and 10 International Agencies

Argentina	<u>National Institute of Statistics and Censuses</u>
Armenia	<u>National Statistical Service</u>
Austria	<u>Statistics Austria</u>
Bangladesh	<u>Bureau of Statistics</u>
Belarus	<u>Ministry of Statistics and Analysis</u>
Bolivia	<u>National Institute of Statistics</u>
Botswana	<u>Central Statistics Office</u>
Brazil	<u>Institute of Geography and Statistics</u>
Bulgaria	<u>National Statistics Institute</u>
Cambodia	<u>National Institute of Statistics</u>
Canada	<u>Statistics Canada</u>
Chile	<u>National Institute of Statistics</u>
China	<u>National Bureau of Statistics</u>
Colombia	<u>National Administrative Department of Statistics</u>
Costa Rica	<u>National Institute of Statistics and Censuses</u>
Cuba	<u>National Statistics Office</u>
Czech Republic	<u>Czech Statistical Office</u>
Dominican Republic	<u>National Statistics Office</u>
Ecuador	<u>National Institute of Statistics and Censuses</u>
Egypt	<u>Central Agency for Public Mobilization and Statistics</u>
El Salvador	<u>General Directorate of Statistics and Censuses</u>
Ethiopia	<u>Central Statistical Agency</u>
Fiji Islands	<u>Bureau of Statistics</u>
France	<u>National Institute of Statistics and Economic Studies</u>
Germany	<u>Federal Statistical Office</u>
Ghana	<u>Ghana Statistical Services</u>
Greece	<u>National Statistical Office</u>
Guatemala	<u>National Institute of Statistics</u>
Guinea-Bissau	<u>National Statistics Institute</u>
Guinea (Conakry)	<u>National Statistics Directorate</u>
Haiti	<u>Institute of Statistics and Informatics</u>
Honduras	<u>National Institute of Statistics</u>
Hungary	<u>Central Statistical Office</u>
India	<u>Ministry of Statistics and Programme Implementation</u>
Indonesia	<u>Statistics Indonesia</u>
Iraq	<u>Central Organization for Statistics and Information Technology</u>
Ireland	<u>Central Statistical Office</u>
Israel	<u>Central Bureau of Statistics</u>
Italy	<u>National Institute of Statistics</u>
Jordan	<u>Department of Statistics</u>
Kenya	<u>National Bureau of Statistics</u>
Kyrgyz Republic	<u>National Statistical Committee</u>
Lesotho	<u>Bureau of Statistics</u>

Madagascar	<u>National Institute of Statistics</u>
Malawi	<u>National Statistical Office</u>
Malaysia	<u>Department of Statistics</u>
Mali	<u>National Directorate of Statistics and Informatics</u>
Mauritius	<u>Central Statistical Office</u>
Mexico	<u>National Institute of Statistics, Geography, and Informatics</u>
Mongolia	<u>National Statistical Office</u>
Morocco	<u>High Commission of Planning</u>
Mozambique	<u>National Institute of Information for Development</u>
Nepal	<u>Central Bureau of Statistics</u>
Netherlands	<u>Statistics Netherlands</u>
Nicaragua	<u>National Institute of Statistics and Censuses</u>
Niger	<u>National Statistics Institute</u>
Pakistan	<u>Statistics Division</u>
Palestine	<u>Central Bureau of Statistics</u>
Panama	<u>Census and Statistics Directorate</u>
Paraguay	<u>General Directorate of Statistics, Surveys, and Censuses</u>
Peru	<u>National Institute of Statistics and Informatics</u>
Philippines	<u>National Statistics Office</u>
Portugal	<u>National Institute of Statistics</u>
Romania	<u>National Institute of Statistics</u>
Rwanda	<u>National Institute of Statistics</u>
Saint Lucia	<u>Government Statistics Department</u>
Senegal	<u>Agence Nationale de la Statistique et de la Démographie</u>
Sierra Leone	<u>Statistics Sierra Leone</u>
Slovenia	<u>Statistical Office of the Republic of Slovenia</u>
South Africa	<u>Statistics South Africa</u>
Spain	<u>National Institute of Statistics</u>
Sudan	<u>Central Bureau of Statistics</u>
Switzerland	<u>Federal Statistical Office</u>
Tanzania	<u>National Bureau of Statistics</u>
Thailand	<u>National Statistical Office</u>
Turkey	<u>Turkish Statistical Institute</u>
Turkmenistan	<u>National Institute of Statistics and Information</u>
Uganda	<u>Bureau of Statistics</u>
Ukraine	<u>State Committee of Statistics</u>
United Kingdom	<u>Office of National Statistics</u>
United States	<u>Bureau of the Census</u>
Uruguay	<u>National Institute of Statistics</u>
Venezuela	<u>National Institute of Statistics</u>
Vietnam	<u>General Statistics Office</u>
Zambia	<u>Central Statistical Office</u>

International

[African Centre for Statistics \(ACS\)](#)

[Arab Institute for Training and Research in Statistics \(AITRS\)](#)

[Centre d'Estudis Demogràfics \(CED\)](#)

[Centre Population et Développement \(CEPED\)](#)

[Data First/Centre of Actuarial Research, University of Capetown](#)

[Demographic Center of Latin America and the Caribbean \(CELADE\)](#)

[DRS Data Services Ltd.](#)

[Institut National d'Études Démographiques \(INED\)](#)

[International Household Survey Network \(IHSN\)](#)

[Inter-American Statistical Institute \(IASI\)](#)