AMERICAN COMMUNITY SURVEY
2013 ACS 1-YEAR PUMS FILES

Prepared by
American Community Survey Office
U.S. Census Bureau
January 15, 2015[1]

I.) Overview of the Public Use Microdata Sample (PUMS)

The Public Use Microdata Sample (PUMS) contains a sample of actual responses to the American Community Survey (ACS). The PUMS dataset includes variables for nearly every question on the survey, as well as many new variables that were derived after the fact from multiple survey responses (such as poverty status). Each record in the file represents a single person, or--in the household-level dataset--a single housing unit. In the person-level file, individuals are organized into households, making possible the study of people within the contexts of their families and other household members. PUMS files for an individual year, such as 2013, contain records of data from approximately one percent of the United States population. As such, PUMS files covering a three-year period, such as 2011-2013, contain records of data from approximately three percent of the United States population, while PUMS files covering a five-year period, such as 2009-2013, contain records of data from approximately five percent of the United States population.

The PUMS files are much more flexible than the aggregate data available on American FactFinder, though the PUMS also tend to be more complicated to use. Working with PUMS data generally involves downloading large datasets onto a local computer and analyzing the data using statistical software such as R, SPSS, Stata, or SAS.

Since all ACS responses are strictly confidential, many variables in the PUMS files have been modified in order to protect the confidentiality of survey respondents. For instance, particularly high incomes are "top-coded", uncommon birthplace or ancestry responses are grouped into broader categories, and the PUMS files provide a very limited set of geographic variables (explained more below).

II.) Public Use Microdata Areas (PUMA)

While PUMS files contain cases from nearly every town and county in the country, towns and counties (and other low-level geography) are not identified by any variables in the PUMS datasets. The most detailed unit of geography contained in the PUMS files is the Public Use Microdata Area (PUMA). PUMAs are special non-overlapping areas that partition each state into contiguous geographic units containing no fewer than 100,000 people each.

The 2013 ACS PUMS files rely on PUMA boundaries that were drawn by state governments using data from the 2010 Census. An interactive mapping application, TIGERweb, can be used to view PUMA boundaries from 2010. TIGERweb is available from the Census Bureau's web site. To access the maps:
- Go to http://www.census.gov/acs/www/data_documentation/other_resources/.
- In the middle of the page, click on "2010 Census Public Use Microdata Area (PUMA) Maps".
- Click on the "TIGERweb Applications" tab.
- Click the "TIGERweb Decennial" link on the left side of the webpage. A new window will open.
- Find the "Layers" tab on the upper left. If it did not open automatically, click on the Layers icon and a drop down menu should open.

- In the "Layers" drop-down menu, find the "Select Vintage" box and select "Census 2010" to view PUMA boundaries used for ACS data years 2012 and 2013.
- Check the box next to the "PUMAs, UGAs, and ZCTAs" and expand this section to check only the box for "Public Use Microdata Areas." You may have to zoom in on the map before this choice is allowed.
- You can now zoom in on the map to view the PUMA boundaries and numbers.
- You will need to adjust the level of zoom until you see the blue numbers that are the PUMA codes.

There are two additional resources that may help PUMS users understand and use PUMAs. They are the software MABLE, developed by the Missouri Census Data Center, and static maps published by the Census Bureau.

The proportion of a PUMA's population that is within a county or other geography can be calculated by using the software MABLE. Found at http://mcdc.missouri.edu/websas/geocorr12.html, it allows you to enter the geography you are interested in and then supplies you with the PUMA codes.

Static maps for PUMAs can be found at http://www.census.gov/geo/maps-data/maps/reference.html. Use the zoom feature to read the fine print that identifies geography features.

III.) PUMS Documentation

The PUMS Documentation page (http://www.census.gov/acs/www/data_documentation/pums_documentation/) includes the following documents:
- **Subjects in the PUMS**
- **PUMS Code Lists**
- **PUMS Top Coded and Bottom Coded Values**
  This document contains tables that show the top code only or the top code and bottom code values for each of these housing and person variables by state.
- **PUMS Data Dictionary**
  Information on PUMS variables.
- **PUMS ReadMe**
- **PUMS Estimates for User Verification**
  PUMS estimates for selected housing and population characteristics are included on the ACS website to assist data users in determining that they are correctly using the weights to compute estimates. These estimates are referred to as PUMS Control Counts. When data users have doubts about the way they are computing estimates, they should attempt to reproduce the estimates that are provided in the files. The standard errors provided in this document were computed using the replicate weight method.
- **Accuracy of the PUMS**
  Detailed descriptions of the sampling methodology, weighting methodology, confidentiality, and standard errors for the PUMS.

IV.) Getting PUMS data

ACS Website

PUMS files can be accessed via the ACS website at
http://www.census.gov/acs/www/data_documentation/pums_data/

American FactFinder

PUMS Files are also accessible via American FactFinder at
http://factfinder2.census.gov/

DataFerrett

It is also possible to get PUMS data from the Census Bureau's DataFerrett, which has the additional feature of being able to make tables and perform basic analysis online. This tool is particularly useful for researchers who need a quick statistic or do not have access to statistical software. DataFerrett is available at
http://www.census.gov/acs/www/data_documentation/data_ferrett_for_pums/

V.) PUMS files structure

The ACS questionnaire contains "household" items that are the same for all members of the household (such as the number of rooms in the home) and "person" items that are unique for each household member (such as age, sex, and race). The ACS PUMS files are made available in this same structure. Researchers who are analyzing only household-level items can use the household files, whereas those using only person-level variables can use the person-level files.

Data users should note that PUMS files containing data for the entire United States (in contrast to individual state and state-equivalent files) are separated into multiple data files. These files must be concatenated in order to create a complete file. For example, users downloading the 2013 ACS 1-year PUMS files of United States Population Records will notice an "a" file and a "b" file. Each file contains about half the population records in the 2013 1-year PUMS dataset of the United States. Below are instructions for concatenating the two PUMS person-level files, in the form of an italicized SAS program and pseudo-code.

Concatenate the two **person-level** files using the set statement:

data population;
set psam_pusa psam_pusb;
run;

The 2013 ACS 1-year PUMS files of the United States Housing Records also contains an "a" file and a "b" file upon downloading. To create a complete housing-level file, the two files must be concatenated. Below are instructions for concatenating the two PUMS household-level files, in the form of an italicized SAS program and pseudo-code.

Concatenate the two **household-level** files using the set statement:

data housing;
set psam_husa psam_husb;
run;

Some data users will need to use household and person items together--for instance, to analyze how the number of rooms in a home varies by a person's age. This type of analysis will require the merging of the household and person files. This merger must rely on the SERIALNO variable, which is the same in the household and person files. Below are instructions for merging the housing and population PUMS files, in the form of an italicized SAS program and pseudo-code.

Use the variable SERIALNO to merge population and housing files.

1. First make sure the files are sorted by SERIALNO.
    proc sort data=population;
    by serialno;
    run;
    proc sort data=housing;
    by serialno;
    run;

2. Then merge the two files together using SERIALNO as a merge key.
    data combined;
    merge population (in=pop) housing;

/*In SAS, the 'in=' option will allow you to keep only those housing units that have people*/

    by serialno;

/*This SAS statement keeps only those housing units that were in the population file*/

    if pop;
    run;

You should not merge the files unless the estimates you want require a merge. Note that there are many estimates that can be tabulated from the person file and from the household file without any merging. The suggested merge will create a person level file, so that the estimate of persons can be tallied within categories from the household file and the person weights should be used for such tallies.

Please note that housing characteristics cannot be tallied from this merged file without extra steps to ensure that each housing weight is counted only once per household.

VI.) Weights in the PUMS

The ACS PUMS is a weighted sample, and weighting variables must be used to generate accurate estimates and standard errors. The PUMS files include both population weights and household

weights. Population weights should be used to generate statistics about individuals, and household weights should be used to generate statistics about housing units. The weighting variables are described briefly below.

PWGTP: Person's weight for generating statistics on individuals (such as age).

WGTP: Household weight for generating statistics on housing units and households (such as average household income).

WGTP1-WGTP80 and PWGTP1-PWGTP80: Replicate weighting variables, used for generating the most accurate standard errors for households or individuals.

PWGTP and WGTP can be used both to generate the point estimates and to generate standard errors when using a generalized formula. Replicate weights can be used just to calculate "direct standard errors." Direct standard errors are expected to be more accurate than generalized standard errors, although they may be more inconvenient for some users to calculate. Both generalized and direct standard errors are explained in more detail in the Accuracy of the PUMS document (http://www.census.gov/acs/www/data_documentation/pums_documentation/).

Each housing unit and person record contains 80 replicate weights. To use the replicate weights to calculate an estimate of the direct standard error, first form the estimate using the full PUMS weight, then form the estimate using each of the 80 replicate weights--providing both the full PUMS estimate and 80 replicate estimates. These should then be entered into the following formula, which is explained in more detail in the Accuracy of the PUMS document:

$$SE(X) = \sqrt{\frac{4}{80} \sum_{r=1}^{80} (X_r - X)^2}$$

Where X_r is a replicate estimate from X_1 to X_{80}, and X is the full PUMS weighted estimate.

The technical explanation of the ACS replicate weights is in Chapter 12 of the Design and Methodology document found at: http://www.census.gov/acs/www/methodology/methodology_main/. For more information on the theoretical basis, please reference Fay, R. and Train, G. (1995), "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties," Proceedings of the Section on Government Statistics, American Statistical Association, pp. 154-159, 1995.

Please note that many estimates generated with PUMS will be slightly different from estimates for the same characteristics published in American FactFinder. These differences are due to the fact that the PUMS files include only about two-thirds of the cases that were used to produce estimates on American FactFinder, as well as additional PUMS edits. More information on the PUMS sample design is available in the "Accuracy of the PUMS" document (http://www.census.gov/acs/www/data_documentation/pums_documentation/).

VII.) Variable changes in the 2013 1-year PUMS files

The 2013 1-year PUMS includes most of the variables that were included in the 2012 1-year PUMS. There were deleted variables, and some new variables, plus some variables with new codes, modified codes, or cosmetic changes to variable labels or value labels. See the 2013 1-year PUMS Data Dictionary for a complete listing of the variables and values contained in the 2013 1-year PUMS data file.

If PUMS users want to compare 2013 PUMS estimates to 2012 PUMS, or want to reuse a program written for 2012 PUMS data, the following table provides some information about variables with changes between 2012 and 2013. For more details on the changes, see both the 2012 and 2013 1-year PUMS Data Dictionaries and Code Lists at
http://www.census.gov/acs/www/data_documentation/pums_documentation/.

Variable	Explanation of changes from the 2012 ACS PUMS to 2013 ACS PUMS
MIL	The categories have changed from the 2012 PUMS.
MLPCD, MLPFG	MLPC, MLPD, MLPF and MLPG no longer exist. Instead, their time periods have been combined to create the new variables MLPCD and MLPFG.
PERNP, PINCP	The lowest possible value for PERNP is now -010000 instead of -009999. The lowest possible value for PINCP is now -019999 instead of -019998.
RESMODE	A code of 3 is included for Internet responses.
RWAT	This variable identifies housing units with hot and cold running water. The change made to 2013 data is that a code of 9 for 'not applicable' is applied to cases from Puerto Rico. RWATPR ('running water') replaces RWAT for cases in Puerto Rico.
SSMC	SSMC is a new variable identifying same-sex married couples. This variable replaces SSPA which was used in 2012 for the same-sex spouse recode.
FER, TEL	These variables were suppressed in a few select PUMAs in 2012 due to data collection errors. They are no longer suppressed in 2013.
INDP, NAICSP	These variables used with 2013 ACS data were recollapsed for PUMS based on the official 2012 classification codes for industry. This resulted in a few differences in the NAICSP and INDP variables on the 2012 ACS PUMS data which were based on the 2007 codes.
ADJINC	This variable is updated each year to account for changes in inflation. See the PUMS Data Dictionary for a list of the income variables which should be multiplied by ADJINC (after applying the decimal point) to adjust them to 2013 dollars.

In 2012, there was a need to collapse the categories of certain variables due to data disclosure requirements. The variables remain the same in 2013 as they were in 2012. For reference, those variables are:

- ANC1P
- ANC2P
- LANP
- OCCP

- SOCP
- CITWP
- MARHYP
- YOEP

- MIGSP
- POBP
- POWSP
- RAC2P

- RAC3P

Variables new to the 2013 PUMS 1-year data file are:

- RWATPR and FRWATPRP
- SSMC
- MLPCD, MLPFG
- All computer and internet variables, plus their flags:

  ACCESS, BROADBND, COMPOTHX, DIALUP, DSL, FIBEROP, HANDHELD, LAPTOP, MODEM, OTHSVCEX, SATELLITE.

  FACCESSP, FBROADBNDP, FCOMPOTHXP, FDIALUPP, FDSLP, FFIBEROPP, FHANDHELDP, FLAPTOPP, FMODEMP, FOTHSVCEXP, FSATELLITEP

VIII.) Additional Information

The Census Bureau occasionally provides corrections or updates to PUMS files. We notify users of these updates via the Census Bureau's E-mail Updates system (https://service.govdelivery.com/service/subscribe.html?code=USCENSUS_C12) and on the ACS errata page (http://www.census.gov/acs/www/data_documentation/errata/).

Please contact acso.users.support@census.gov with any PUMS-related questions.