

ACS 2009 1-Year PUMS File Readme

I.) Overview of the Public Use Microdata Sample files (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 household. In the person-level file, individuals are organized into households, making possible the study people within the contexts of their families and other household members. The PUMS contains data on approximately one 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 file have been modified in small ways 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 file provides a very limited set of geographic variables (explained more below).

II.) Public Use Microdata Area (PUMA)

While PUMS files contain cases from nearly every town and county in the country, most 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. ACS PUMS files from 2005-2009 rely on PUMA boundaries that were drawn by state governments at the time of the 2000 Census. PDF-format maps of PUMA boundaries are available from the Census Bureau's web site at <http://www.census.gov/geo/www/maps/puma5pct.htm>. From this index page choose a state. When you get to the PDF document be sure to note that the first page is an index page that displays entities called "Super PUMAs". These are not the PUMAs available in the ACS PUMS files. The PUMAs in the ACS PUMS are sometimes referred to a "5% PUMAs" because they were also used on the 5% Sample PUMS files from the 2000 decennial census, whereas the Super-PUMAs (also known as "1% PUMAs") were the ones used on the 1% PUMS files in 2000. The key to using these maps is to understand that the PUMAs nest within the Super-PUMAs and these PDF files have, following the initial state-level Super-PUMA overview map, 1 or more inset maps showing more detail

for metropolitan areas within the state, and then one page for each super-puma showing the boundaries of the PUMAs. The maps also show relevant place and county boundaries to help you see what geographic areas correspond to the PUMAs.

A listing of the detailed components of each PUMA is available within the directories at http://www2.census.gov/census_2000/datasets/PUMS/FivePercent/.

The Missouri Census Data Center created a tool that allows you to enter the geography you're interested in and then it supplies you with the PUMA codes. For more information, go to the following URL: <http://mcdc2.missouri.edu/websas/geocorr2k.html> There is a power point slide set (.ppt) at the top of the page of the MABLE/Geocorr page MABLE/Geocorr tutorial that is current with a few examples. The link to the earlier versions of power point slides, with 1990 vintage geographies, also at the top of the page, gives the best instructions.

III.) PUMS Documentation

Information on PUMS variables can be obtained from the PUMS data dictionary located on the following URL:

http://www.census.gov/acs/www/Downloads/data_documentation/pums/DataDict/PUMSDataDict09.pdf

Detailed descriptions of the sampling methodology for the PUMS can be found with the following link:

http://www.census.gov/acs/www/Downloads/data_documentation/pums/Accuracy/2009AccuracyPUMS.pdf

PUMS code lists are located on the following URL:

http://www.census.gov/acs/www/Downloads/data_documentation/pums/CodeLists/ACSPUMS2009CodeLists.pdf

Fourteen housing variables and thirteen person variables are top coded and bottom coded in the Public Use Microdata Sample files. The ACS website 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. The 2009 PUMS Top Coded and Bottom Coded Values are available in HTML and Excel formats and are located on the following URL:

http://www.census.gov/acs/www/Downloads/data_documentation/pums/TopCodedValues/2009PUMS_top_coded_values.pdf

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 should attempt to reproduce the estimates that are provided in the files located on the following URL:

<http://www.census.gov/acs/www/Products/PUMS/>

IV.) Getting PUMS data

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

It is also possible to get PUMS data from the Census Bureau's DataFerret, 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. DataFerret is available at http://www.census.gov/acs/www/Products/PUMS/acs_pums_download_via_ferrett.htm

V.) Analyzing PUMS data

A. PUMS file 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.

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 the race of the household. 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 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 file includes both population weights and household weights. Population weights should be used 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 - Housing Weight for generating statistics on households (such as average household income).

WGTP1-WGTP80 and PWGTP1-PWGTP80 - replicate weighting variables, used for generating statistics on individuals or households with the most accurate standard errors available.

While PWGTP and WGTP can be used to generate both the point estimates and error for the characteristic variables, replicate weights can be used to create a even more accurate

estimates of just the standard error. Replicate weights are used to calculate what we refer to as direct standard errors. Direct standard errors will often be more accurate than generalized standard errors, although they may be more inconvenient for some users to calculate. Each housing unit and person record contains 80 replicate weights. For any estimate X, 80 replicate estimates are also computed using the replicate weights. Using replicate weights leads to a better estimate of standard error than simply using the adjustments due to the person's or household 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 plugged into the following formula, which is explained in more detail in the accuracy document (found at: http://www.census.gov/acs/www/Downloads/data_documentation/pums/Accuracy/2009AccuracyPUMS.pdf):

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

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

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/Downloads/survey_methodology/acs_design_methodology_ch12.pdf. 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."

V.) Additional Information

Some of the PUMS estimates will be different from the estimates for the same characteristics published in the American FactFinder and for Census 2000. For an explanation of these differences, see the 2009 Accuracy of the PUMS document located on the following URL:

http://www.census.gov/acs/www/Downloads/data_documentation/pums/Accuracy/2009AccuracyPUMS.pdf

After the release of the 2009 file, we may issue updates and corrections to the 2008 version of the PUMS. We will keep users aware of these updates via the ACS Alert from the ACS website and on the ACS errata page located on:
<http://www.census.gov/acs/www/UseData/Errata.htm>

2009 ACS 1-Year PUMS Variable Changes

Variables that changed from 2008 (See 2009 ACS PUMS data dictionary)

Variables that have substantive changes: ADJINC, YBL, CITWP, MARHYP, YOEP, INDP, NAICSP, VPS

Variables that have cosmetic changes: ADJHSG, AGS, CONP, FULP, MHP, MRGI, MRGP, MRGT, MRGX, RNTM, SMP, VALP, GRNTP, OCPIP, SMOCP, SMX, SVAL, TAXP, FACRP, FAGSP, FBATHP, FBDSP, FBLDP, FBUSP, FCONP, FELEP, FFSP, FFULP, FGASP, FHFLP, FINSP, FKITP, FMHP, FMRGIP, FMRGP, FMRGTP, FMRGXP, FMVP, FPLMP, FREFRP, FRMSP, FRNTMP, FRNTP, FRWATP, FSINKP, FSMP, FSMXHP, FSMXSP, FSTOVP, FTAXP, FTELP, FTENP, FTOILP, FVACSP, FVALP, FVEHP, FWATP, FYBLP, CIT, COW, JWRIP, LANP, OCCP, POBP, POVPIP, SOCP

Deleted Variables (See 2009 ACS PUMS data dictionary)

UWRK

New variables for 2009 (See 2009 ACS PUMS data dictionary)

FOD1P, FOD2P, FHINS3C, FHINS4C, FHINS5C, WRK