2008 ACS 1-Year PUMS File Readme

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

The Public Use Microdata Sample files, or PUMS, are a sample of the actual responses to the American Community Survey and include most population and housing characteristics. The PUMS will allow users to tabulate data to their own specifications. The 2008 ACS 1-Year PUMS currently contains data for the nation, states, Puerto Rico, and Public Use Micro Data Sample Areas (PUMAs). PUMAs are special non-overlapping areas that partition a state, and each PUMA contains a population of about 100,000. State governments drew the PUMA boundaries at the time of the 2000 Census.

II.) Public Use Microdata Area (PUMA)

The PUMA (Public Use Microdata Area) level geography lets you map and analyze data for your entire state. One of the things people do with data from the Census Bureau is to create thematic maps or summary reports that show spatial distributions of data within their state or region. To learn more, you can start with a set of PDF-file base maps accessible from the Bureau's web site at http://www.census.gov/geo/www/maps/puma5pct.htm. From this index page choose your 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 PUMA's you want. The PUMAs you do want are sometimes referred to a "5% PUMAs" because they were the geography used on the 5% Sample PUMS files in 2000, 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.

Census Geographic Equivalency Files, which will help you use the PUMA's with the ACS PUMS are located on:

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 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 URLs:

http://www.census.gov/acs/www/Products/PUMS/pumsdict.html

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

http://www.census.gov/acs/www/Products/PUMS/pumsaccuracy_archived.html

PUMS code lists are located on the following URL:

http://www.census.gov/acs/www/Products/PUMS/codelist_archived.html

Eleven housing variables and seven 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. You can ignore the percentiles and use the values under the variable names. The 2008 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/Products/PUMS/C2SS/minmaxval_archived.html

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 one of the following files located on the following URL:

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

You may want to try using the ACS 2008 PUMS with the Census Bureau tool DataFerrett, when trying to create the tabulations that you are interested in. The 2008 PUMS is available in DataFerrett. The DataFerrett is a data-mining tool that accesses data stored in The Data Web through the Internet. DataFerrett can be installed as an application or accessed through an Internet browser as a java applet.
See the following URL for more information:

http://www.census.gov/acs/www/Products/PUMS/acs_pums_download_via_ferrett.htm

IV.) Developing Estimates Using the PUMS

A. Instructions on merging the PUMS population and housing files:

To create the estimates using the PUMS, you should merge the population and housing PUMS files. Here are the 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.

B. Using PUMS Weights

When developing weighted estimates using the ACS PUMS, the following weights should be used:

PWGTP - Person's weight for generating person counts such as Age
WGTP - Housing Weight for generating household counts such as Number of Households

Each housing unit and person record contains 80 replicate weights. For any estimate X, 80 replicate estimates are also computed using the replicate weights. Replicate weights can be 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. The advantage of using replicate weights is that a single formula is used to calculate the standard error of many types of estimates.

To use our replicate weights, the instructions are to form the estimate first using the full PUMS weight, then form the estimate using each of the 80 replicate weights. That will give them the full PUMS estimate and 80 replicate estimates. Plug this into the formula in the accuracy document and that will provide an estimate of the standard error. See our PUMS Accuracy document for 2008 PUMS for the formula, pages 8-9 available at: http://www.census.gov/acs/www/Products/PUMS/pumsaccuracy_archived.html

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/dm1.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."

C. Special usage note on the National PUMS Files

Due to download issues reported by data users, the US ACS PUMS files have been split into two sets of files. One set has names ending with the letter “a”, the other set with the letter “b”.

The following states are in “a” files:

01 Alabama/AL
02 Alaska/AK
04 Arizona/AZ
05 Arkansas/AR
06 California/CA
08 Colorado/CO
09 Connecticut/CT
10 Delaware/DE
11 District of Columbia/DC
12 Florida/FL
13 Georgia/GA
15 Hawaii/HI
16 Idaho/ID
17 Illinois/IL
18 Indiana/IN
19 Iowa/IA
20 Kansas/KS
21 Kentucky/KY
22 Louisiana/LA
23 Maine/ME
24 Maryland/MD
25 Massachusetts/MA
26 Michigan/MI
27 Minnesota/MN
28 Mississippi/MS

The following states are in “b” files:

29 Missouri/MO
30 Montana/MT
31 Nebraska/NE
32 Nevada/NV
33 New Hampshire/NH
34 New Jersey/NJ
35 New Mexico/NM
36 New York/NY
37 North Carolina/NC
38 North Dakota/ND

39 Ohio/OH
40 Oklahoma/OK
41 Oregon/OR
42 Pennsylvania/PA
44 Rhode Island/RI
45 South Carolina/SC
46 South Dakota/SD
47 Tennessee/TN
48 Texas/TX
49 Utah/UT
50 Vermont/VT
51 Virginia/VA
53 Washington/WA
54 West Virginia/WV
55 Wisconsin/WI
56 Wyoming/WY

The SAS data sets have the following naming convention:

psam_husa.sas7bdat
psam_husb.sas7bdat
psam_pusa.sas7bdat
psam_pusb.sas7bdat

The CSV files have the following naming convention:

ss08husa.csv
ss08husb.csv
ss08pusa.csv
ss08pusb.csv

The UNIX files have the following naming convention:

unix_husa.zip
unix_husb.zip
unix_pusa.zip
unix_pusb.zip

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 2008 Accuracy of the PUMS document located on the following URL:
http://www.census.gov/acs/www/Products/PUMS/pumsaccuracy_archived.html

For more information on ACS data products, please reference the ACS 2008 Guide to the data products, located on the ACS Website on the following URL:
http://www.census.gov/acs/www/Products/users_guide/2008/index.htm

Throughout 2009 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

2008 ACS 1-Year PUMS Variable Changes

Variables that changed from 2007 (See 2008 ACS PUMS data dictionary)
Variables that have changes: BDSP, FS, RMSP, VALP, YBL, FFSP, REL, SCHG, YOE, DIS, IND, NAICS, and SCHL.

Deleted Variables (See 2008 ACS PUMS data dictionary)
ADJUST (person and housing unit record), MILY and FMILYP.

New variables for 2008 (See 2008 ACS PUMS data dictionary)
ADJHSG, ADJINC, BATH, CITWP, REFR, RWAT, SINK, STOV, TOIL, DDRS, DEAR, DEYE, DOUT, DPHY, DREM, DRAT, DRATX, HINS1, HINS2, HINS3, HINS4, HINS5, HINS6, HINS7, HICOV, KIT, MARHM, MARHW, MARHD, MARHT, MARHY, MARHDUR, PRIVCOV, PUBCOV, PLM, MULTG, FCITWP, FDDRSP, FDEARP, FDEYEP, FDOUTP, FDPHYP, FDRATP, FDRATXP, FDREMP, FHINS1P, FHINS2P, FHINS3P, FHINS4P, FHINS5P, FHINS6P, FHINS7P, FBATHP, FSTOVP, FREFRP, FRWATP, FSINKP, FTOILP, FMARHDURP, FMARHDP, FMARHMP, FMARHTP, FMARHWP, FMARHYP.