AMERICAN COMMUNITY SURVEY
2016-2020 5-YEAR PUMS File

ReadMe

Prepared by
American Community Survey Office
U.S. Census Bureau
March 31, 2022

Table of Contents

I. Purpose of This Document 4

II. Overview of the Public Use Microdata Sample (PUMS) 4

III. PUMS Geographies 5

A. Geographies Available in PUMS 5
B. Overview of Public Use Microdata Areas (PUMA) 5
C. Note on Dual PUMAs 5
D. Interactive Maps Using TIGERweb 5
E. Static Maps for PUMAs 6
F. Crosswalking Dual PUMA Codes Using GEOCORR and MABLE 7

IV. PUMS Documentation 7

A. PUMS Webinar 7
B. PUMS Documentation 7
C. PUMS Handbook 7

V. PUMS Data Dictionary 8

A. PUMS Data Dictionary Overview 8
B. Explanation of Variables in Data Dictionary PDF File 8
C. Explanation of Variables in Data Dictionary CSV File 8
D. Organization of PUMS Variables in the Data Dictionary 9

VI. Obtaining PUMS Data 10

A. ACS Website 10
B. ACS FTP Site 10
C. Microdata Analysis Tool (MDAT) 10

VII. PUMS File Structure 11

A. Basic Example of Combining PUMS Person and Housing Files 11

VIII. PUMS Weights and Notes on calculating variances 12

A. PUMS Weighting Variables 12
B. Successive Difference Replication (SDR) Formula for Calculating Uncertainty 13
C. Successive Difference Replicate (SDR) Documentation 14

IX. Changes to PUMS Variables for the 2016-2020 PUMS 5-year Files 14

X. Suppression of Variable Values for Select Geographies 14

A. Suppression for Ancestry Variables (ANC, ANC1P, ANC2P) and Divorced (MARHD) 14
B. Suppression for Telephone (TEL) 15

XI. Additional Notes and Useful Information 15

A. Note on 2020 PUMS 1-year Data 15
B. Definition of the Top- and Bottom-Coded Variables 15
C. Rounding Rules for Income Variables 16
D. Note on the PUMS Design Factors 16
E. Note on Income and Earnings Inflation Factor (ADJINC) 17
F. Note on Housing Dollar Inflation Factor (ADJHSG) 17
G. Issue with Own Child (OC) and Related Child (RC) 17

H. Note on Standard Occupational Classification codes (SOCP) 18
I. Note on Selected Values for Industry and Occupation (INDP, NAICSP, OCCP, and SOCP) 18
J. Codes to Identify North American Industry Classification System (NAICS) Equivalents 18
K. Additional Information on PUMS Industry and Occupation Codes 19
L. Note on PUMS Data Dictionary and Blank Values 19
M. Note on PUMS File Names for CSV Files 19
N. Additional Notes: 20

I. PURPOSE OF THIS DOCUMENT

This document is intended to provide resources and guidance for data users on how to use the American Community Survey (ACS) Public Use Microdata Sample (PUMS) files.

II. OVERVIEW OF THE PUBLIC USE MICRODATA SAMPLE (PUMS)

The Public Use Microdata Sample (PUMS) files allow data users to create estimates for user-defined characteristics. The files contain a sample of the responses to the American Community Survey (ACS). The PUMS files include variables for nearly every question on the ACS survey. Additional variables are also created from other recoded PUMS variables to provide data users with useful derived variables (such as poverty status) while protecting confidentiality and providing consistency within the PUMS files.

Please note that many estimates generated with PUMS may be different from estimates for the same characteristics published on data.census.gov. These differences are due to the fact that the PUMS microdata is a sample of the full ACS microdata, and includes only about two-thirds of the records that were used to produce ACS estimates. Additional edits appropriate for PUMS were also made for confidentiality reasons.

There are two types of PUMS files, one for Person records and one for Housing Unit records. Each record in the Person file represents a single person. Individuals are organized into households, making possible the study of people within the contexts of their families and other household members. In addition, the files contain people who reside in group quarters (GQ), such as nursing homes or college dormitories. The Housing Unit files contain records for individual housing units, including vacant housing units. In addition, GQ records are also on the Housing Unit file. However, they are placeholder records that may be used solely to obtain GQ information for the variable called FS ("Yearly food stamp/Supplemental Nutrition Assistance Program (SNAP) recipiency").

PUMS files for an individual year contain data on approximately one percent of the United States population. PUMS files covering a five-year period contain data on approximately five percent of the United States population.

The PUMS files are much more flexible than the aggregate data provided in tables on data.census.gov, 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 replaced with a top-code value and uncommon birthplace or ancestry responses are grouped into broader categories. The PUMS files also limits the geographic detail below the state level. The only substate geography provided is the Public Use Microdata Area, or PUMA.

III. PUMS GEOGRAPHIES

The following sections provide an overview on geographies available in the PUMS files.

A. Geographies Available in PUMS

In order to protect confidentiality, a limited number of geographic summary levels are available on the PUMS files. They include region, division, state and Public Use Microdata Area (PUMA).

Division is a subdivision of the region summary level. An example of region is "Northeast", while a division is "New England". The PUMS variable for regional division is called "DIVISION".

In addition to the 50 states, there are also two state equivalents: the District of Columbia and Puerto Rico. Records for the District of Columbia are included in the PUMS files for the nation. Puerto Rico data is only available as a state-level file.

B. Overview of Public Use Microdata Areas (PUMA)

While PUMS files contain records from across the nation, 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 PUMA.

PUMAs are special non-overlapping areas that partition each state into contiguous geographic units containing roughly 100,000 people at the time of their creation. They are created after each Decennial Census. The current PUMS files use the PUMA definitions created after Census 2010. PUMAs are expected to be updated using the Census 2020 data at a later date.

PUMAs are identified by a 5-digit code. Note that you must use the state variable (ST) along with the PUMA code to uniquely identify an individual PUMA. PUMA codes are unique within a state, but not between states. For example, the PUMA code "00100" is used in both Connecticut and North Dakota.

C. Note on Dual PUMAs

The current PUMA boundaries were first used beginning in 2012. For PUMS data for 2011 and earlier, the boundaries created after Census 2000 are used. Older multi-year records will contain both PUMA codes. For those files, PUMA codes from Census 2000 are called PUMA00, while the newer PUMA codes from Census 2010 are called PUMA10. Data users will need to crosswalk their data to obtain a single PUMA geography. They may do so using allocation rates using GEOCORR (see below).

D. Interactive Maps Using TIGERweb

The Census Bureau provides an interactive mapping application, called TIGERweb. Data users can view PUMA boundaries from 2010. TIGERweb is available at: https://tigerweb.geo.census.gov/tigerwebmain/tigerweb_main.html.

To access the maps:

- Click on "TIGERweb Applications" on the upper left.
- Click "TIGERweb" on the left column. This will take you to a new page.
- On the upper left, you should see choices: "Layers", "Legend", or "Task Results". "Layers" should by default be selected. If not, then select it.
- Under "Select Vintage:" choose "ACS 2019"[1].
- Select "PUMAs, UGAs, and ZCTAs" on the left.
- Click on the map to zoom, or move the zoom scale bar to zoom in closer to the map.
- Expand the "PUMAs, UGAs, and ZCTAs" box to see the choices: "2010 Census Public Use Microdata Areas" and "2010 Census ZIP Code Tabulation Areas".
- When the check boxes next to "2010 Census ZIP Code Tabulation Areas" and "2010 Census Public Use Microdata Areas" become active, deselect "2010 Census ZIP Code Tabulation Areas" to simplify the map.
- Alternately, you may check other boxes to add or remove geographic summary levels and other features, such as "Hydrography". Use the plus sign ("+") to see more detail for a particular selection.

E. Static Maps for PUMAs

Data users may be interested in static maps of PUMA boundaries. These may be found at: https://www.census.gov/programs-surveys/geography.html.

To access the static maps:

- Click on "Geographies" on the left hand side.
- Click on "Geography Reference Maps".
- Click on "More" to the right of the list of years. Change the year to "2010".
- Scroll down and click on "2010 Census Public Use Microdata Area (PUMA) Reference Maps".
- Choose the state you are interested in from the drop down menu.
- A list of state and PUMA codes with their names will appear. Click on the link at the top under "Access Public Use Microdata Area (PUMA) Maps". The link will have the state name followed by "Maps". For example, "Idaho Maps".
- You will be redirected to the relevant Census ftp site. Use the list of state and PUMA codes to select the relevant folder. For example, the state and PUMA code for "16 00300" has the description "Central Idaho--Lewiston City & Nez Perce Reservation PUMA". Thus, you would click on the folder called "puma1600300/" to access these maps.
- Select the pdf that you are interested in.

[1] Note that the vintage normally matches the data year. However, due to the fact that PUMA boundaries are updated after the completion and release of Decennial Census, the vintage for data year 2020 will say "ACS 2019".

F. Crosswalking Dual PUMA Codes Using GEOCORR and MABLE

The software MABLE was developed by the Missouri Census Data Center (MCDC), which is part of the Census Bureau's State Data Center program (https://www.census.gov/about/partners/sdc.html).

GEOCORR stands for "Geographic Correspondence Engine". According to the MCDC site it is an application that accesses the MABLE geographic database. More information may be found at: http://mcdc.missouri.edu/applications/geocorr.html.

Among other things, the software allows data users to calculate the proportion of a PUMA code from Census 2000 that lies within the new PUMA codes from Census 2010. It also provides the data user with an allocation factor so that they may crosswalk old PUMA codes to new PUMA codes.

Data users may be interested in a Data Gem on how to use GEOCORR to crosswalk geographies. It may be found here: https://www.census.gov/data/academy/data-gems/2021/how-to-use-the-geocorr-to-identify-the-geographies-that-make-up-your-area.html.

IV. PUMS DOCUMENTATION

The list below provide links to documentation that are useful for PUMS users.

A. PUMS Webinar

Data users new to PUMS may find the "Introduction to the Public Use Microdata Sample (PUMS) File" webinar to be a useful reference. The webinar may be found at: https://www.census.gov/data/academy/webinars/2020/introduction-to-american-community-survey-public-use-microdata-sample-pums-files.html.

Other training, such as how to use the online Microdata Analysis Tool (MDAT), may be found at https://www.census.gov/programs-surveys/acs/microdata/mdat.html.

B. PUMS Documentation

The PUMS documentation has been reorganized. Data users may go to https://www.census.gov/programs-surveys/acs/microdata.html to find a central location for PUMS documentation. The page provides links to the PUMS technical documentation, such as the PUMS data dictionary as well as provide links to webinars on PUMS. In addition, there are links to access the data either through the FTP site or through the new online tool, called MDAT.

C. PUMS Handbook

A series of Handbooks are available that provide an overview of various aspects of the ACS. They are located at: https://www.census.gov/programs-surveys/acs/guidance/handbooks.html. The PUMS Handbook is called "Understanding and Using the American Community Survey Public Use Microdata Sample Files: What Data Users Need to Know".

V. PUMS DATA DICTIONARY

A. PUMS Data Dictionary Overview

The PUMS Data Dictionary provides the values for each PUMS variable, as well as labels for each value. For example, on the PUMS files, if the PUMS variable "REGION" has a value of "3", a data user may use the Data Dictionary to see that "3" means "South".

The PUMS Data Dictionary is published in three different formats. There is a text version, a pdf version, and a comma-separated values (CSV) version. The information in each version is equivalent to one another.

B. Explanation of Variables in Data Dictionary PDF File

Below is an example of the PUMS variable for Record Type (RT). The first line shows the PUMS variable name (RT), followed by "Character" to indicate it is a character variable and the number 1 to indicate that the length of the variable is one. The next line provides the descriptive title for the variable ("Record Type"). The remaining two lines provide the PUMS values for the variable and their appropriate labels. For example, when RT = "H", this stands for "Housing Record or Group Quarters Unit".

Example of PUMS Data Dictionary (PDF Version)

```
RT          Character   1
      Record Type
            H .Housing Record or Group Quarters Unit
            P .Person Record
```

C. Explanation of Variables in Data Dictionary CSV File

Below is an example of how the Data Dictionary appears in the CSV version.

Example of Data Dictionary from CSV file

```
NAME,RT,C,1,"Record Type"
VAL,RT,C,1,"H","H","Housing Record or Group Quarters Unit"
VAL,RT,C,1,"P","P","Person Record"
```

The position of the variables (from left to right in the file) is provided in the table below. The file itself contains no variable names. That is, the first line of the file represents records for the data. Note that lines starting with "NAME" contain five variables and is equivalent to the first two lines in the PDF example (above). Lines starting with "VAL" have seven variables and include the variable starting and ending values and labels.

PUMS Variables in Data Dictionary CSV File

Position	Variable	Description
1	Identifying Flag (ID flag)	"NAME" for information about the variable "VAL" for values of the variable
2	PUMS Variable Name	PUMS variable name (e.g. RT, SERIALNO, AGEP, etc.)
3	Variable Type	"C" for Character variable "N" for Numeric variable Note: most variables are character variables
4	Length	Length of PUMS variable
5	Descriptive Title / Starting Legal Value	Descriptive Title (ID Flag = "NAME") Starting value for variable value range (ID Flag = "VAL")
6	Ending Legal Value	Ending value for value variable range (ID Flag = "VAL")
7	Description	Descriptive name (ID Flag = "VAL")

The text and CSV versions may be read in to statistical programs to create formats for the PUMS variables. Note that the CSV version was first published for 2017 PUMS data. For 2016 and earlier data, only the text version is available.

D. Organization of PUMS Variables in the Data Dictionary

The PUMS variables are placed into groups within the Data Dictionary. The Housing variables come first, followed by the Person variables. The variables are further divided into categories which are listed below. These sections and subsections are provided in the pdf and text versions of the Data Dictionary, but not the CSV version. The CSV version is intended to be machine-readable. Therefore, only variables and their values are present in that file.

Major Variable Organizational Categories

Data Dictionary Section	Description
HOUSING RECORD	
BASIC VARIABLES	Basic variables, such as geographic variables and inflation adjustment variables
HOUSING UNIT VARIABLES	Housing variables pertaining to the Housing Unit
HOUSEHOLD VARIABLES	Housing variables pertaining to the Household
ALLOCATION FLAGS	Housing allocation flag variables
REPLICATE WEIGHTS	Housing replicate weight variables used for variance calculation
PERSON RECORD	
BASIC VARIABLES	Basic variables, such as geographic variables and inflation adjustment variables
PERSON VARIABLES	Person Variables
RECODED PERSON VARIABLES	PUMS Person Variables created from other Variables.
ALLOCATION FLAGS	Person allocation flag variables
REPLICATE WEIGHTS	Person replicate weight variables used for variance calculation

VI. OBTAINING PUMS DATA

PUMS data may be obtained in multiple ways.

A. ACS Website

PUMS files can be accessed via the ACS website at https://www.census.gov/programs-surveys/acs/microdata/access.html. There are two columns shown. The column on the left links the FTP site. The column on the right links to the Microdata Analysis Tool (MDAT).

B. ACS FTP Site

The PUMS Files are also available through the file transfer protocol (FTP) site at: https://www2.census.gov/programs-surveys/acs/data/pums/. Choose the relevant year that you are interested in. Then choose the folder for the time period. For example, choose the folder marked "5-Year" to access 5-year PUMS files.

Data users may find a list of state names and abbreviations useful. The information may be obtained here: https://www.census.gov/library/reference/code-lists/ansi/ansi-codes-for-states.html. Click on "FIPS Codes for the States and District of Columbia" to obtain the state abbreviations.

C. Microdata Analysis Tool (MDAT)

The Microdata Analysis Tool (MDAT) may be found at: https://data.census.gov/mdat/.

The tool may be used to create estimates online without the use of statistical software. Note that the tool may change in the future. It is still under development and in beta form. As a note, the tool cannot currently inflation adjust PUMS variables. This capability will be added as a future enhancement.

Data users may find the webinar "Custom Tabulation Capabilities on data.census.gov" useful in learning how to use MDAT. It may be found here: https://www.census.gov/data/academy/webinars/2021/custom-tabulation-capabilities.html.

VII. 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 may use the housing unit files, whereas those using only person-level variables may use the person files.

The person files also contain records for persons in group quarter facilities (such as nursing homes or college dorms). The housing unit files contain placeholder records for group quarters. The majority of the variables for housing unit records for group quarters are blank. The weights and replicate weights are zero. The group quarter place holder records exist so that data users may obtain values for the variable FS (Yearly food stamp/Supplemental Nutrition Assistance Program recipiency).

PUMS files containing data for the entire United States are separated into several files due to their size. For 1-year data, there are two files, and "a" and a "b" file. For 5-year PUMS data, there are four files ("a" through "d"). More information is provided at the end of this document.

A. Basic Example of Combining PUMS Person and Housing Files

Below are instructions for concatenating the two 1-year "a" and "b" PUMS files to create a single national file. The code is in italics and uses SAS programming code[2].

Concatenate the person-level files using the set statement:

```
data population;
   set psam_pusa psam_pusb;
run;
```

Concatenate the household-level files using the set statement:

```
data housing;
   set psam_husa psam_husb;
run;
```

[2] The code presented here uses the SAS software. Copyright © 2021 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

As mentioned above, if the data user is using 5-year PUMS files they will need to concatenate four files (psam_husa through psam_husd) together.

Some data users will need to use household and person items together. For instance, in order to analyze how the number of rooms in a home varies by a person's age, merge the household and person files together using the serial number variable (SERIALNO).

Note that vacant housing units are included on the housing files. There are no person-level records for these records.

First make sure the files are sorted by SERIALNO.

```
proc sort data=population;
   by serialno;
run;
proc sort data=housing;
   by serialno;
run;
```

Then merge the two files together using SERIALNO as a merge key. Note that in SAS, the 'in=' option will allow you to identify records from a specific file. The line 'if pop' retains only records from the population file.

```
data combined;
   merge population (in=pop) housing;
   by serialno;
   if pop;
run;
```

You do not need to merge the files unless the estimates you wish to create require a merge. Note that there are many estimates that may 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 may be tallied within categories from the household file and the person weights should be used for such tallies.

Note also that the housing unit record files contain vacant housing units. There are no population records for these housing units.

VIII. PUMS WEIGHTS AND NOTES ON CALCULATING VARIANCES

A. PUMS Weighting Variables

The ACS PUMS is a weighted sample. Weighting variables must be used in order to calculate estimates which represent the actual population. Weighting variables are also needed to generate accurate measures of uncertainty, such as the standard error or margin of error.

The PUMS files include both population weights (in the Person files) and household weights (located in the Housing files). Population weights should be used to generate statistics about individuals, and household weights should be used to generate statistics about housing units or households. The weighting variables are described briefly below.

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

PWGTP1-PWGTP80: Replicate Person weighting variables, used for generating the standard error and margin of error for person characteristics.

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

WGTP1-WGTP80: Replicate Housing Unit weighting variables, used for generating the standard error and margin of error for housing unit and household characteristics.

The PUMS Weighting variables (PWGTP and WGTP) may both be used to generate PUMS estimates. They are also used in the generalized variance formulas (GVF) method for calculating standard errors using the design factors. Replicate weights may only be used to calculate standard errors and margins of error using the successive difference replication (SDR) method. The SDR method may also be referred to as direct standard errors.

B. Successive Difference Replication (SDR) Formula for Calculating Uncertainty

The ACS uses the SDR methodology to calculate margins of error for published data products. The SDR method is discussed in the Accuracy of the PUMS document, located at: https://www.census.gov/programs-surveys/acs/microdata/documentation.html.

As previously mentioned, each housing unit and person record contains 80 replicate weights. To use the replicate weights to calculate an estimate of the SDR standard error, first calculate the PUMS estimate using the PUMS weight (either PWGTP or WGTP).

Then calculate 80 replicate estimates, using each of the 80 replicate weights. For example, for the first replicate estimate, use the first replicate weight (e.g. PWGTP1 instead of PWGTP, or WGTP1 instead of WGTP).

Calculate the variance by first taking the difference between each replicate estimate and the PUMS estimate. Square each of these differences, and then sum the 80 squared differences. Multiply this sum by the quantity 4/80. The 4 is required to remove bias, while 80 is present due to the 80 replicate estimates. The equation is below:

$$VAR(x) = \frac{4}{80}\sum_{r=1}^{80}(x_r - x)^2$$

In the equation, x_r is a r^{th} replicate estimate, and x is the full PUMS weighted estimate.

To obtain the standard error (SE), take the square root of the variance. To obtain a 90% confidence level margin of error, multiply the SE by 1.645.

The variance estimates for some 2016-2020 PUMS 5-year estimates may be larger than expected when compared to the equivalent variance estimates from previous years. This is due to a reduction in the number of 2020 records which was caused by collection issues due to the pandemic.

C. Successive Difference Replicate (SDR) Documentation

The webinar called "Calculating Margins of Error the ACS Way" provides an overview on how to calculate variance, standard errors, and margins of error using the SDR formula. It provides a worked example using PUMS data. It is located at: https://www.census.gov/data/academy/webinars/2020/calculating-margins-of-error-acs.html.

Another reference for how to use the SDR formula is provided in the Variance Replicate Estimate (VRE) Tables Documentation, located at: https://www.census.gov/programs-surveys/acs/data/variance-tables.html. Although the VRE documentation pertains to ACS data, the concepts provided in the documentation may be adopted for use with PUMS data. This document presents worked examples using the ACS VRE tables.

The technical explanation of the creation of the ACS replicate weights may be found in Chapter 12 of the Design and Methodology document located at: https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html.

IX. CHANGES TO PUMS VARIABLES FOR THE 2016-2020 PUMS 5-YEAR FILES

Changes to variables from last year are noted in a new document called "2016-2020 ACS 5-year PUMS Variable Changes and Explanations". It is located at: https://www.census.gov/programs-surveys/acs/microdata/documentation.html.

X. SUPPRESSION OF VARIABLE VALUES FOR SELECT GEOGRAPHIES

A. Suppression for Ancestry Variables (ANC, ANC1P, ANC2P) and Divorced (MARHD)

Due to issues found with the ACS data, Ancestry variables (ANC, ANC1P, and ANC2P) and the variable for Divorced in the Past 12 Months (MARHD) for selected geographies are suppressed. Records which were suppressed for ANC and MARHD were assigned a value of '8'. Records which were suppressed for ANC1P and ANC2P were assigned a value of "000".

More information may be found on the ACS Errata Notes page for 2018, located at: https://www.census.gov/programs-surveys/acs/technical-documentation/errata.html. The relevant notes are called "Data Collection Error in Philadelphia County, PA" and "Data Collection Error in Delaware".

B. Suppression for Telephone (TEL)

Problems in the collection of data on the availability of telephone service (TEL) led to suppressing this variable in various PUMAs in 2016 and 2019. Records which were suppressed were assigned a value of '8' for the Telephone variable (TEL).

For the affected geographies, please see the errata note located at : https://www.census.gov/programs-surveys/acs/technical-documentation/errata/118.html

XI. ADDITIONAL NOTES AND USEFUL INFORMATION

A. Note on 2020 PUMS 1-year Data

Due to data collection issues caused by COVID19, the 2020 PUMS 1-year files were released with experimental weights. In addition, the Puerto Rico data was not released. For more information, please see the 2020 PUMS 1-year ReadMe documentation, located at https://www.census.gov/programs-surveys/acs/data/experimental-data/2020-1-year-pums.html.

B. Definition of the Top- and Bottom-Coded Variables

To protect confidentiality, selected variables are top-coded. A process identifies records which meet or exceed the specific top-code threshold value. These records are replaced with the mean value. Note that the distribution from the full ACS sample is used to calculate the mean value. There are separate top-codes for each state.

The threshold value is calculated by identifying the top half percent or top three percent value using the distribution of the full microdata.

The following variables use the half percent threshold:
AGEP, BDSP, ELEP, GASP, INSP, JWMNP, MRGP, RMSP, RNTP, SSP, TAXAMT, VALP, WAGP, WATP

The following variables use the three percent threshold:
CONP, FULP, INTP, MHP, OIP, PAP, RETP, SEMP, SMP, SSIP

The "PUMS Top Coded & Bottom Coded Values" document contains the threshold value and mean value for select variables by state. The top-code threshold variables end in "TPCT". Beginning in 2017, the top-code value variables begin with "T_" (e.g. "T_ELE"). For 2009 through 2016, the top-code value variables had the same name as the PUMS variable that was top-coded (e.g. "ELEP"). Prior to 2009, the P at the end of the variable was not used (e.g. "ELE").

In addition, two variables are bottom-coded. Bottom-coding is similar to top-coding. It identifies records which are at or below the bottom-coded threshold value and replaces them with the mean value of all records less than or equal to the bottom-coded threshold.

The threshold value below which variables are bottom-coded end in "BPCT". Similarly to the top-codes, beginning in 2017 the variables for the bottom-coded values begin with "B_". For 2016 and earlier, the bottom-coded variables are named "BINT" and "BSEM".

Please note that the variable names for the top- and bottom-code threshold and mean variables are not on the PUMS microdata files. For a given state, if a record in the PUMS file has a value at or above the top-code threshold, the value is replaced with the appropriate top-code value.

C. Rounding Rules for Income Variables

PUMS income variables are subject to rounding rules, displayed in the table below.

Range	Rounding Rule	Unrounded Example	Rounded Example
0	0	No Rounding	0
$0 < X \leq 7$	4	6	4
$7 < X \leq 999$	Nearest ten	12	10
$999 < X \leq 49{,}999$	Nearest hundred	5,234	5,200
$X > 49{,}999$	Nearest thousand	54,123	54,000
$-7 \leq X < 0$	-4	-6	-4
$-999 \leq X < -7$	Nearest negative ten	-12	-10
$-49{,}999 \leq X < -999$	Nearest negative one hundred	-5,234	-5,200
$X \leq -49{,}999$	Nearest negative thousand	-54,123	-54,000

D. Note on the PUMS Design Factors

The PUMS design factors are periodically updated as the need arises. For example, if new variables are included on the PUMS file, additional design factors may be added. The design factors are not updated every year. The 2016-2020 PUMS 5-year design factors are the same as the 2015-2019 PUMS 5-year design factors.

For 2016 and earlier, the design factors were published in the PUMS Accuracy document. The values were given in tables at the end of the document. Beginning in 2017, the design factors were published as a csv file.

A description of the design factor variables in the csv file is provided in the table below.

PUMS Design Factor Variables in CSV File

Variable	Description
YEAR	4-digit year
PERIOD	Time period (1-year or 5-year)
STATE	State Name
ST	State FIPS Code
CHARTYP	Characteristic Type (either "POPULATION" or "HOUSING")
CHARACTERISTIC	Description of PUMS Design Factor Characteristic Group
DESIGN_FACTOR	Design Factor

E. Note on Income and Earnings Inflation Factor (ADJINC)

Divide ADJINC by 1,000,000 to obtain the inflation adjustment factor and multiply it to the PUMS variable value to adjust it to 2020 dollars. Variables requiring ADJINC on the Housing Unit file are FINCP and HINCP. Variables requiring ADJINC on the Person files are: INTP, OIP, PAP, PERNP, PINCP, RETP, SEMP, SSIP, SSP, and WAGP.

F. Note on Housing Dollar Inflation Factor (ADJHSG)

Divide ADJHSG by 1,000,000 to obtain the inflation adjustment factor and multiply it to the PUMS variable value to adjust it to 2020 dollars. Variables requiring ADJHSG on the Housing Unit files are: CONP, ELEP, FULP, GASP, GRNTP, INSP, MHP, MRGP, SMOCP, RNTP, SMP, TAXAMT, and WATP.

For PUMS 1-year data, ADJHSG has a value of 1,000,000. That is, a housing dollar inflation factor of 1. Consult the PUMS Data Dictionary for the values for the 5-year files.

Note that TAXAMT is inflation adjusted. In the past TAXP was not due to it being a categorical variable. In addition, the Housing Dollar Inflation factor is not applied to Property Value (VALP). Data users may apply the inflation adjustment to VALP if they wish to compare VALP and TAXAMT using the same criteria.

Note that ADJHSG does not apply to AGS because it is a categorical variable. If data users convert the categories in AGS to a numeric value (for example, using the midpoint of the range of each category), then they may apply the inflation factor.

G. Issue with Own Child (OC) and Related Child (RC)

For records from 2016 and earlier, group quarters records assigned the variables Own Child (OC) and Related Child (RC) a value of zero ('0'). For 2017 and later years, GQ records have a blank value for OC and RC. These variables are housing variables and therefore group quarters records should not be included in the universe of OC or RC. There is currently no plan to reissue data files.

H. Note on Standard Occupational Classification codes (SOCP)

In cases where the Standard Occupational Classification (SOCP) codes ends in X(s) or Y(s), two or more SOC occupation codes were aggregated to correspond to a specific PUMS SOCP code. In these cases, the PUMS occupation description is used for the SOC occupation title.

Additional information on Occupation groupings within major categories may be found at: https://www.census.gov/topics/employment/industry-occupation/guidance/indexes.html.

I. Note on Selected Values for Industry and Occupation (INDP, NAICSP, OCCP, and SOCP)

Some codes are pseudo-codes developed by the Census Bureau and are not official NAICS, industry, or occupation codes.

Pseudo-Codes Values for Select Variables

Variable	Value	Description
SOCP	999920	Unemployed, With No Work Experience In The Last 5 Years Or Earlier Or Never Worked
NAICSP	999920	Unemployed, With No Work Experience In The Last 5 Years Or Earlier Or Never Worked
SOCP	559830	MIL-Military, Rank Not Specified
OCCP	9920	Unemployed, With No Work Experience In The Last 5 Years Or Earlier Or Never Worked
INDP	9920	Unemployed, With No Work Experience In The Last 5 Years Or Earlier Or Never Worked

J. Codes to Identify North American Industry Classification System (NAICS) Equivalents

Data users may notice that some values of the PUMS variable NAICSP contain letters in addition to numbers. The table below provides an explanation of these letters.

Description of Special Letters in NAICSP Variable

Code	Description
M	Multiple NAICS codes
P	Part of a NAICS code - NAICS code split between two or more Census codes
S	Not specified Industry in NAICS sector - Specific to Census codes only
Z	Exception to NAICS code - Part of NAICS industry but has a unique Census code

Additional information on NAICS may be found at: https://www.census.gov/topics/employment/industry-occupation/guidance/indexes.html.

Note that NAICS is pronounced “nakes”.

K. Additional Information on PUMS Industry and Occupation Codes

Data users may wish to consult the Code Lists at https://www.census.gov/programs-surveys/acs/microdata/documentation.html for more information on how industry and occupation codes are mapped to PUMS industry and occupation codes.

For additional information on NAICS and SOC groupings within major categories see the Industry and Occupation page, located at: https://www.census.gov/topics/employment/industry-occupation.html.

L. Note on PUMS Data Dictionary and Blank Values

Records in PUMS that are not within the universe for a variable are given blank values. For example, for the PUMS variable Educational Attainment (SCHL), the universe is people age 3 or older. Person records with an age less than 3 have a blank value for SCHL.

The PUMS Data Dictionary represents blank values as a series of b's. For example, Educational Attainment has a length of 2. In the PUMS Data Dictionary, blank values for SCHL are displayed as "bb". The PUMS files do not use b's to denote blanks. Instead, they are either a numeric blank (for numeric variables) or a character blank value (for character variables).

M. Note on PUMS File Names for CSV Files

Data users may download PUMS data in either a CSV file or as a SAS file. Beginning with 2017 data, the CSV file will have the same name as the SAS file. For Person-level files, the name is "PSAM_P<ST>" and for Housing-level files, the name is "PSAM_H<ST>". Here, <ST> is the State FIPS code.

State names, abbreviations and FIPS codes may be found here: https://www.census.gov/library/reference/code-lists/ansi.html. Choose the "State and State Equivalents" link. FIPS Codes are 2-digit codes. For example, for Connecticut, <ST> is "09".

Note that for the National files, there are multiple files. For PUMS 1-year data there is an "A" and "B" file. For Person-level data the names are "PSAM_PUSA" and "PSAM_PUSB". The Housing-level files are "PSAM_HUSA" and "PSAM_HUSB".

States Contained in PUMS 1-year National Files

File	First State	First State FIPS Code	Last State	Last State FIPS Code
A	Alabama	01	Mississippi	28
B	Missouri	29	Wyoming	56

For the 5-year data, there are four files, an "A", "B", "C", and "D" file. For Person-level data the names are "PSAM_PUSA", "PSAM_PUSB", "PSAM_PUSC", and "PSAM_PUSD". The Housing-level files are "PSAM_HUSA", "PSAM_HUSB", "PSAM_HUSC", and "PSAM_HUSD".

States in PUMS 5-year National Files

File	First State	First State FIPS Code	Last State	Last State FIPS Code
A	Alabama	01	Hawaii	15
B	Idaho	16	Mississippi	28
C	Missouri	29	Oregon	41
D	Pennsylvania	42	Wyoming	56

Puerto Rico data is not included in the national files. It is published as a state equivalent and has a State FIPS code of "72".

N. Additional Notes:

The Census Bureau occasionally provides corrections or updates to PUMS files. Data users may sign up for notifications and updates via the Census Bureau's E-mail Updates system at: https://service.govdelivery.com/accounts/USCENSUS/subscriber/new?category_id=USCENSUS_C12.

PUMS errata notes are located at: https://www.census.gov/programs-surveys/acs/technical-documentation/errata.html.

User notes are located at: https://www.census.gov/programs-surveys/acs/technical-documentation/user-notes.html.

Data users may also email acso.users.support@census.gov with any PUMS-related questions.