

**AMERICAN COMMUNITY SURVEY**  
**2018 ACS 1-YEAR PUMS FILES**  
*ReadMe*

Prepared by  
American Community Survey Office  
U.S. Census Bureau  
November 14, 2019

## Table of Contents

I.	Purpose of This Document .....	3
II.	Overview of the Public Use Microdata Sample (PUMS).....	3
III.	PUMS Geographies .....	4
A.	Geographies Available in PUMS.....	4
B.	Overview of Public Use Microdata Areas (PUMA).....	4
C.	Note on Dual PUMAs.....	4
D.	Interactive Maps Using TIGERweb.....	5
E.	Static Maps for PUMAs.....	5
F.	Crosswalking Dual PUMA Codes Using GEOCORR and MABLE.....	6
IV.	PUMS Documentation.....	6
A.	PUMS Introduction to the Public Use Microdata Sample (PUMS) File Webinar .....	6
B.	PUMS Documentation.....	6
C.	PUMS Technical Documentation .....	7
V.	PUMS Data Dictionary.....	9
A.	PUMS Data Dictionary Overview .....	9
B.	Explanation of Variables in Data Dictionary PDF File .....	9
C.	Explanation of Variables in Data Dictionary CSV File.....	10
D.	Organization of PUMS Variables in the Data Dictionary .....	11
VI.	Obtaining PUMS Data.....	11
A.	ACS Website.....	11
B.	American FactFinder .....	11
C.	DataFerrett and MDAT.....	12
D.	ACS FTP Site.....	12
VII.	PUMS File Structure.....	12
A.	Basic Example of Combining PUMS Person and Housing Files .....	13
VIII.	PUMS Weights and Notes on calculating variances.....	14
A.	PUMS Weighting Variables .....	14
B.	Successive Difference Replication (SDR) Formula for Calculating Uncertainty .....	14
C.	ACS Variance Replicate Table Documentation .....	15
IX.	Changes to PUMS Variables for the 2018 PUMS 1-year Files.....	16
X.	Explanation of Variable Changes .....	18
A.	Note on New Utility Cost Flag Variables (ELEFP, FULFP, GASFP, and WATFP).....	18
B.	Note on Changes to Utility Cost variables (ELEP, FULP, GASP, and WATP) .....	18
C.	Note on Change to Property Tax (TAXP and TAXAMT) .....	18
D.	Note on Collapsing Year of Entry (YOEP) Categories .....	18
E.	Suppression for Ancestry Variables (ANC, ANC1P, ANC2P) and Divorced (MARHD) .....	19
F.	Note on Condo Fee (CONP).....	19
G.	Note on Language variables (LANP and HHLANP) .....	19

H.	Information on new Industry and Occupation Codes (INDP, NAICSP, SOCP, and OCCP)	19
I.	Note on New Format for Serial Numbers (SERIALNO).....	20
J.	Note on Removing “\$0 or Break Even” Category for Income Variables.....	20
K.	Rounding Rules for Income Variables.....	20
L.	Note on Universe for Total Person’s Earnings (PERNP).....	20
XI.	Additional Notes and Useful Information.....	21
A.	Note on the PUMS Design Factors.....	21
B.	Note on Income and Earnings Inflation Factor (ADJINC).....	21
C.	Note on Housing Dollar Inflation Factor (ADJHSG).....	21
D.	Note on Public Use Microdata Area (PUMA) Variable.....	21
E.	Note on Standard Occupational Classification codes (SOCP).....	21
F.	Note on Selected Values for Industry and Occupation (INDP, NAICSP, and OCCP).....	22
G.	Codes to Identify North American Industry Classification System (NAICS) Equivalents	22
H.	Additional Information on PUMS Industry and Occupation Codes.....	23
I.	Note on Change to World Area of Birth (WAOB).....	23
J.	Note on PUMS Data Dictionary and Blank Values.....	23
K.	Note on Top-coded and Bottom-coded Variables.....	23
L.	Note on PUMS File Names for CSV Files.....	24
M.	Location of PUMS Design Factors.....	24
N.	Additional Notes:.....	25

## **I. PURPOSE OF THIS DOCUMENT**

This document is intended to provide resources and guidance for data users on how to use the 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 includes variables for nearly every question on the ACS survey. Additional variables are also created from PUMS responses to provide data users with useful variables (such as poverty status) while protecting confidentiality and providing consistency within the PUMS files.

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, such as nursing homes or college dormitories. The Housing Unit files contain records for single housing units, including vacant housing units.

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, regional division, state and Public Use Microdata Area (PUMA).

Regional division is a subdivision of the region summary level. An example of region is “Northeast”, while a regional 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. They are 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 cases from nearly every town and county in 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 the Public Use Microdata Area (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 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 from Mable (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](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”. Select “Layers.”
- Under “Select Vintage:” choose “ACS 2018”.
- 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” on the left.
- 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.

#### **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.

### **IV. PUMS DOCUMENTATION**

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

#### **A. PUMS Introduction to the Public Use Microdata Sample (PUMS) File Webinar**

Data users new to PUMS may find the Introduction to PUMS webinar to be a useful reference. The webinar may be found at: <https://www.census.gov/programs-surveys/acs/guidance/training-presentations/acs-intro-pums.html>.

#### **B. PUMS Documentation**

PUMS provides documentation for data user, located at: <https://www.census.gov/programs-surveys/acs/technical-documentation/pums.html>. The documentation provides several relevant links.

##### **1. Confidentiality of PUMS**

This link provides a brief explanation on how respondent’s confidentiality is protected.

##### **2. What's Available and How Can I Access PUMS?**

This section provides links to a new beta version of the Microdata Analysis System (MDAT). This tool replaces DataFerrett and allows data users to calculate PUMS estimates online without the use of statistical software.

In addition, a link to the PUMS data is also provided. This links to the Census Bureau's FTP server where data users can download the PUMS files directly.

### **3. Need Help with PUMS?**

This section provides links to training resources, including the PUMS Handbook (also known as the PUMS Compass Product), the Introduction to PUMS Webinar, and training videos on how to use DataFerrett. These videos will be replaced with training videos for how to use MDAT as they become available.

### **4. PUMS Technical Documentation**

This section provides a link to the PUMS Technical documentation (discussed below).

### **5. Geographic Areas Available**

This section provides a brief overview on the Public Use Microdata Area (PUMA).

## **C. PUMS Technical Documentation**

The PUMS Technical Documentation is available at: <https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html>. This page includes the following documents:

### **1. Subjects in the PUMS**

This document provides a list of subjects available in the PUMS data. PUMS data users may also consult the ACS Subject Definitions located on the ACS Technical Documentation page, located at: <https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>. The ACS Subject Definitions provide relevant information on how variables were collected for the ACS.

### **2. PUMS Data Dictionary**

The PUMS Data Dictionary provides PUMS variables, their values and the meaning of each of the values. Data users may access the Data Dictionary in three formats. There is a pdf version, a text version, and a comma-separated value (CSV) version.

The text version has been reorganized and formatted to make it easier for data users to read it into a program. The CSV version is also designed to be machine-readable. The Census Bureau welcomes data users' comments and suggestions on how to improve format and layout of the Data Dictionary.

### **3. PUMS Code Lists**

The PUMS Code Lists provide categories for ACS variables with a large number of coded responses such as ancestry or occupation. They provide the relevant

PUMS categories for the detailed codes. This allows data users to more clearly understand how each ACS code maps to the broad categories used in PUMS.

#### **4. PUMS Top Coded and Bottom Coded Values**

To protect confidentiality, selected variables are top-coded. A process identifies records which exceed the specific top-code threshold value. These records are replaced with the mean value of all records that are equal to or greater than the threshold value.

The PUMS Top Coded and Bottom Coded Values document contains the threshold value and top-code value for select variables by state. Top-code threshold variables end in “TPCT”. Beginning in 2017, the top-code value variables begin with “T\_”. Prior to 2017, the top-code value variables had the same name as the PUMS variable that was top-coded.

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 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”.

Note that for 2018 PUMS data, the top-coding for property tax value is now top-coded using state-level thresholds. The previous PUMS variable for property tax was a categorical variable called TAXP. For 2018, it is replaced with a numeric variable, TAXAMT.

In addition, three additional variables were top-coded for 2018. They are Public assistance income past 12 months (PAP), Social Security income past 12 months (SSP), and Supplementary Security Income past 12 months (SSIP).

#### **5. Accuracy of the PUMS**

The Accuracy of the PUMS document provides descriptions of the sampling methodology, weighting methodology, confidentiality, and how to calculate standard errors.

For 2016 and earlier the PUMS design factors are located in appendices at the end of the PUMS Accuracy document. Beginning in 2017, the design factors are released in a comma-separated value (CSV) file. The design factors are used in calculating standard errors and margins of errors using a generalized variance function (GVF) method. A description of the layout of the design factor file may be found at the end of this document.



## 6. PUMS Estimates for User Verification

PUMS estimates for selected housing unit and population characteristics are included so that data users may verify they are using PUMS weights correctly. The estimates are weighted to represent the entire population. In addition, standard errors (SE) and margins of error (MOE) are also included for verification purposes. The SE and MOE provided in this document were computed using the successive difference replication (SDR) methodology, which uses the PUMS replicate weights. See the section on PUMS Weights and Notes on Calculating Variances below for more information.

Note that the text version of the file (ending in “[LST]”) has been discontinued. The last version of the text version was published in 2016. The CSV and SAS versions are still available.

## 7. PUMS Record Count

Data users may confirm that they have all records for PUMS by examining the unweighted count of PUMS records. Previously, this information was provided at the national level in the introduction to the PUMS Accuracy of the Data document. The PUMS Record Count file was first published in 2017.

# V. PUMS DATA DICTIONARY

## A. PUMS Data Dictionary Overview

The PUMS Data Dictionary provides the values for each PUMS variables, 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”. The equivalent layout for the CSV version of the Data Dictionary is provided further on in the text.

**Example of PUMS Data Dictionary (PDF Version)**

<b>RT</b>	<b>Character</b>	<b>1</b>
	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.

**PUMS Variables in Data Dictionary CSV File**

Position	Variable	Description
1	Identifying Flag	“NAME” for information about the variable “VAL” for values of the variable
2	PUMS Variable Name	PUMS variable name (e.g. RT, SERIALNO, AGE, 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	Starting Legal Value	Starting value for variable value range.
6	Ending Legal Value	Ending value for value variable range.
7	Description	Descriptive name

For values that are not a range, the starting and ending legal values will be identical.

Note that the first line of a variable (Identifying Flag is “NAME”) will have the variable’s descriptive name. Data users who read in the file expecting each record to contain seven variables will thus have the variable’s descriptive name in place of the Starting Legal Value and the Ending Legal Value or Description variables will be blank.

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/data/pums.html>.

Currently, the link above sends data users to American FactFinder (AFF). In the near future, it is expected to link directly with the data on the Census Bureau's FTP site. PUMS files for 2018 and later are not published on AFF.

### B. American FactFinder

PUMS Files are currently accessible via American FactFinder (AFF) at <https://factfinder.census.gov>. Data users should note that the Census Bureau is now disseminating data through its new dissemination platform (data.census.gov). American

Factfinder (AFF) is available as an archive system for accessing 2017 and earlier data through the spring of 2020.

### **C. DataFerrett and MDAT**

DataFerrett is no longer available for accessing PUMS data. The Census Bureau currently has a beta version of the replacement tool, which may be found at: <https://data.census.gov/mdat/>.

The new 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.

### **D. 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 “1-Year” to access 1-year PUMS files.

The document called “PUMS\_file\_naming\_convention.pdf” explains the file naming convention system. The file names include state abbreviations. 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.

## **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 place holder 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 reciprocity) for them.

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.

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;
```

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'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:** Replicate Household weighting variables, used for generating the standard error and margins of error for households.

**PWGTP1-PWGTP80:** Replicate Person weighting variables, used for generating the standard error and margins of error for individuals.

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**

SDR standard errors and margins of error are expected to be more accurate than GVF standard errors and margins of error, although they may be more inconvenient for some users to calculate. Both methods are explained in more detail in the Accuracy of the PUMS document, located on the PUMS Technical Documentation page:

<https://www.census.gov/programs-surveys/acs/technical-documentation/pums/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.

### **C. ACS Variance Replicate Table Documentation**

Data users who wish to learn more may consult the Variance Replicate Estimate (VRE) tables Documentation, located at: <https://www.census.gov/programs-surveys/acs/technical-documentation/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>.

For more information on the theoretical basis for using the Successive Difference Replicate method, please see the paper entitled “Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties” by Robert Fay and George Train (<https://www.census.gov/content/dam/Census/library/working-papers/1995/demo/faytrain95.pdf>).

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.

## IX. CHANGES TO PUMS VARIABLES FOR THE 2018 PUMS 1-YEAR FILES

The 2018 1-year PUMS includes most of the variables that were included in the 2017 1-year PUMS. There were some new variables and one variable which was removed. There were also some variables with new codes, modified codes, or cosmetic changes to variable labels or value labels. See the 2018 1-year PUMS Data Dictionary for a complete listing of the variables and values contained in the 2018 1-year PUMS data file.

The PUMS Data Dictionaries may be found on the PUMS Technical Documentation page located at: <https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html>.

Data users may wish to compare PUMS estimates to previous year estimates, or reuse a program written using previous year data. The PUMS Data Dictionary for the current year and previous year(s) may be used to find changes to the values for a specific variable.

**Variables added since previous 1-year PUMS:** ELEFP, FULFP, GASFP, TAXAMT, WATFP.

**Variables deleted since previous 1-year PUMS:** TAXP.

**Variables with new or modified codes since the previous 1-year PUMS:**

ADJINC, ANC, ANC1P, ANC2P, CITWP, CONP, ELEP, FINCP, FULP, GASP, HHLANP, HINCP, INDP, INTP, LANP, MARHD, MARHYP, NAICSP, OCCP, OIP, PAP, PERNP, PINCP, RETP, SEMP, SERIALNO, SOCP, SPORDER, SSIP, SSP, WAGP, WAOB, WATP, YBL, YOEP.

**Variables that changed from character to numeric:** SPORDER

**Variables with cosmetic changes to variable labels or value labels:**

Variables with cosmetic changes, such as rewording a variables description, are not noted this year. This is due to the large number of variables with cosmetic changes. The main change is that numeric variables will no longer be zero-padded. For example, a numeric variable of length 3 and a value of 27 is formatted to display as 27 this year instead of 027.

Zero-padding numeric variables is a holdover from when the PUMS data was in a fixed-width format. The values for all variables were displayed in a single string. The position within the string determined the value for a variable. Zero-padding numeric variables ensured that they took up the appropriate width. As each variable is individually defined, this format is no longer required and was removed from the majority of numeric variables.

### Detailed Changes by Variable (as compared to 2017 ACS 1-year PUMS Data Dictionary)

Variable	Description of Change
	<i>New Variables</i>
ELEFP	New Flag Variable for ELEP
FULFP	New Flag Variable for FULP
GASFP	New Flag Variable for GASP



<b>Variable</b>	<b>Description of Change</b>
TAXAMT	Replaces TAXP for Property Taxes
WATFP	New Flag Variable for WATP
	<i>Deleted Variable</i>
TAXP	Removed and replaced with TAXAMT
	<i>Variables with Metadata Changes</i>
SPORDER	Formerly character variable, now numeric
	<i>Variables with Annual Code Updates</i>
ADJINC	Annual Update of Inflation Factors
CITWP	Current year (2018) added
MARHYP	Current year (2018) added
YBL	Current year (2018) added
	<i>Variables with Annual Code Updates and Code Changes</i>
YOEP	New collapsed categories for years prior to 1945 and Current year (2018) added
	<i>Variables with Code Changes</i>
ANC	Category added to represent suppressed values
ANC1P	Category added to represent suppressed values
ANC2P	Category added to represent suppressed values
CONP	Categories updated to reflect income rounding rules
ELEP	Removed non-cost responses
FINCP	Categories updated to reflect income rounding rules
FULP	Removed non-cost responses
GASP	Removed non-cost responses
HHLANP	Removed non-valid category
HINCP	Categories updated to reflect income rounding rules
INDP	INDP categories updated to reflect new collapsing categories
INTP	Categories updated to reflect income rounding rules
LANP	Removed non-valid category
MARHD	Category added to represent suppressed values
NAICSP	NAICSP categories updated to reflect new collapsing categories
OCCP	OCCP categories updated to reflect new collapsing categories
OIP	Categories updated to reflect income rounding rules
PAP	Categories updated to reflect income rounding rules
PERNP	Categories updated to reflect income rounding rules and Universe for blank values updated
PINCP	Categories updated to reflect income rounding rules
RETP	Categories updated to reflect income rounding rules
SEMP	Categories updated to reflect income rounding rules
SERIALNO	Updated to represent new format for variable
SOCP	SOCP categories updated to reflect new collapsing categories
SSIP	Categories updated to reflect income rounding rules

Variable	Description of Change
SSP	Categories updated to reflect income rounding rules
WAGP	Categories updated to reflect income rounding rules
WAOB	American Samoa changed categories for World Area of Birth
WATP	Removed non-cost responses

In 2017, there was a need to collapse the categories of certain variables due to data disclosure requirements. In 2018, the variables for Industry and Occupation were also updated. The list of collapsed variables are:

- ANC1P
- ANC2P
- CITWP
- HHLANP
- INDP
- LANP
- MARHYP
- MIGSP
- NAICSP
- OCCP
- POBP
- RAC2P
- RAC3P
- SOCP
- YOEP

**X. EXPLANATION OF VARIABLE CHANGES**

This section is for additional information on changes that occurred to the PUMS Variables.

**A. Note on New Utility Cost Flag Variables (ELEFP, FULFP, GASFP, and WATFP)**

Previously, the utility cost variables ELEP, FULP, GASP, and WATP contained values for the variable that were unrelated to utility costs. For example, for 2017 and earlier data, the value of 1 for ELEP represented “Included in rent or condo fee” and 2 represented “No charge or electricity not used”. Only values of 3 or higher represented true monthly electricity costs. Starting in 2018, these non-cost values have been removed from the variable and placed in a flag variables. These new flag variables end in “FP”. For example, the flag variable for electricity costs is called ELEFP.

**B. Note on Changes to Utility Cost variables (ELEP, FULP, GASP, and WATP)**

As noted above, values that did not represent an actual cost were removed from these variables and placed in a flag variable.

**C. Note on Change to Property Tax (TAXP and TAXAMT)**

Beginning in 2018, the categorical property tax variable (TAXP) has been replaced with a numeric property tax variable, called TAXAMT. Beginning in 2018, property tax is top-coded using state-level thresholds and values. The largest TAXP category (TAXP = “68”) had a national-level threshold of \$10,000 or higher. To allow data users to use the state-level top-code values, a numeric version of the property tax variable was created.

**D. Note on Collapsing Year of Entry (YOEP) Categories**

Most values for Year of Entry represent a given year. However, for 1944 and earlier, the years are placed in categories. The collapsed categories are different from the 2017 and early collapsed categories for these year.

### **New Collapsed Categories for Year of Entry for Years Prior to 1945**

<b>YOEP Category</b>	<b>Description</b>
1929	1929 or earlier (Bottom-coded)
1930	1930 - 1934
1935	1935 - 1938
1939	1939 - 1940
1941	1941 - 1942
1943	1943 - 1944

#### **E. 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. 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”.

#### **F. Note on Condo Fee (CONP)**

Due to changes in the universe, a value of 0 now indicate a condo fee of \$0. Note that CONP does not include vacant condo units. Data users are therefore cautioned from using it to obtain an estimate of condos in a geographic area.

#### **G. Note on Language variables (LANP and HHLANP)**

For PUMS, the language code for Zuni is collapsed into a broader language category. However, it was incorrectly listed in previous years’ PUMS Data Dictionaries as a separate uncollapsed language code for LANP and HHLANP. This category was therefore removed as it does not appear in the PUMS microdata.

The ACS language categories are collapsed in PUMS to protect confidentiality. Information about the collapsing may be found in the PUMS Code Lists, located at: <https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html>.

#### **H. Information on new Industry and Occupation Codes (INDP, NAICSP, SOCP, and OCCP)**

The Industry and Occupation branch has updated the Industry and Occupation codes. Data users may be interested in crosswalking older codes to the new codes. ACS Code Lists and Crosswalks may be found here: <https://www.census.gov/topics/employment/industry-occupation/guidance.html>

Information on how the ACS codes are collapsed into PUMS codes may be found in the PUMS Code Lists documentation located on the PUMS Technical Documentation site (<https://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html>).

**I. Note on New Format for Serial Numbers (SERIALNO)**

Beginning in 2018, the format for the PUMS serial number (SERIALNO) variable has changed. The format begins with the 4-digit year, followed by a 2-digit code (either “GQ” or “HU”) followed by a 7-digit serial number. For example, “2018GQ0000001”. The “GQ” and “HU” portion of SERIALNO replaces “00”, which was used for 2017 and earlier data.

The “HU” and “GQ” denote whether a record is a group quarter record or household record. For individuals on the Person files, all records for a given household will have the same serial number. Individuals on the Person files may be uniquely identified by using serial number variable (SERIALNO) and the person number variable (SPORDER).

**J. Note on Removing “\$0 or Break Even” Category for Income Variables**

PUMS income variables are subject to rounding rules to protect data confidentiality. In previous years, for income variables, the PUMS Data Dictionary displayed values of 1 with a description of “\$1 or break even”. However, due to the rounding rules, this value was rounded and is not on the microdata files. The Data Dictionary was updated to reflect this reality.

**K. 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

**L. Note on Universe for Total Person’s Earnings (PERNP)**

The definition for blank values has been updated in the Data Dictionary. Previously, it had a definition of “N/A (less than 15 years old)”. This is updated to “N/A (less than 16 years old)” to reflect the universe. Note that there is no change to the responses in the microdata.

## **XI. ADDITIONAL NOTES AND USEFUL INFORMATION**

### **A. 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. In 2017 they were published as a csv file. Prior to that they were located at the end of the PUMS Accuracy document in a set of tables. The design factors for the 2018 PUMS 1-year data are the same as the ones published for 2017. There are no changes.

### **B. 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 2018 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.

### **C. 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 2018 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 1000000. 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.

### **D. Note on Public Use Microdata Area (PUMA) Variable**

Public use microdata areas (PUMAs) are the smallest geography available on the PUMS files. They are designed to have a population of roughly 100,000 or more people. In order to uniquely identify a PUMA code it must be paired with the state FIPS code (ST). PUMA codes within a state are unique. However, multiple PUMAs in different states may share the state PUMA code.

### **E. 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>.

**F. Note on Selected Values for Industry and Occupation (INDP, NAICSP, and OCCP)**

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

**Pseudo-Codes Values for Select Variables**

Variable	Value	Description
INDP	999920	Unemployed And Last Worked 5 Years Ago Or Earlier Or Never Worked
NAICSP	999920	Unemployed And Last Worked 5 Years Ago Or Earlier Or Never Worked
OCCP	9830	MIL-Military, Rank Not Specified
OCCP	9920	Unemployed And Last Worked 5 Years Ago Or Earlier Or Never Worked
INDP	9920	Unemployed And Last Worked 5 Years Ago Or Earlier Or Never Worked

**G. 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 stands for North American Industry Classification System and is pronounced as “nakes”.

## **H. Additional Information on PUMS Industry and Occupation Codes**

Data users may wish to consult the Code Lists on the PUMS Technical Documentation site (<https://www.census.gov/programs-surveys/acs/technical-documentation/pums/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>.

## **I. Note on Change to World Area of Birth (WAOB)**

The categories for World Area of Birth (WAOB) have not changed. However, beginning in 2018, the code for POBP = “060” (American Samoa) was moved from WAOB = “2” (PR and US Island Areas) to WAOB = “8” (Oceania and at Sea). This was due to a change in how citizenship status was determined for respondents who listed their place of birth as American Samoa. See the User Note located at: <https://www.census.gov/programs-surveys/acs/technical-documentation/user-notes/2019-01.html> for more information.

World Area of Birth (WAOB) is determined using the PUMS Place of Birth (POBP) variable. Data users may create their own version of WAOB should they wish to retain the older category definitions.

## **J. 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 for 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).

## **K. Note on Top-coded and Bottom-coded Variables**

To protect confidentiality, selected PUMS variables are top-coded or bottom-coded. That is, a threshold value is calculated. For top-coded variables, all values at or above the threshold value are replaced with the top-code value. For bottom-coded variables, all values at or below the bottom-code threshold are replaced with the bottom-code value.

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, TAXP, VALP, WAGP, WATP

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

**L. 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”.

**M. Location of PUMS Design Factors**

PUMS provides two methods for data users to calculate variances. One uses a generalized variance function (GVF) which involves design factors (DF). 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 are now published in a comma-separated value (CSV) file. A description of the variables is provided in the table below.

Note that for some years the PUMS Accuracy document organized the DF in an appendix called Attachment A. There was a table of contents that allowed data users to quickly go to a particular state’s design factor table. The ATTACHMENT variable is included in the CSV file to provide continuity for this system.

**PUMS Design Factor Variables in CSV File**

<b>Variable</b>	<b>Description</b>
DF_ID	ID variable to identify the Characteristic with a numeric code.
YEAR	4-digit year
PERIOD	Time period (1-year or 5-year)
ATTACHMENT	A-1 through A-53. The attachments were referenced in the 2016 PUMS Accuracy document
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

**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](https://service.govdelivery.com/accounts/USCENSUS/subscriber/new?category_id=USCENSUS_C12).

In addition, PUMS errata notes may be found here: <https://www.census.gov/programs-surveys/acs/technical-documentation/errata.html>.

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

Data users may also email [acso.users.support@census.gov](mailto:acso.users.support@census.gov) with any PUMS-related questions.