

Public Use Microdata Sample (PUMS) Files:
2020 PUMS Top Coded and Bottom Coded Values
Experimental Weights

The following fourteen housing variables and ten person variables are either top coded only or are top and bottom coded in the Public Use Microdata Sample files:

Variable	Housing Variables (All Top Coded)	Variable	Person Variables (Top Coded)
BDSP	Number of bedrooms	AGEP	Age
CONP	Condominium fee (monthly amount)	JWMNP	Travel time to work
ELEP	Electricity cost (monthly cost)	OIP	All other income past 12 months
FULP	Fuel cost (yearly cost for fuels other than gas and electricity)	PAP	Public assistance income past 12 months
GASP	Gas cost (monthly cost)	RETP	Retirement income past 12 months
INSP	Fire/hazard/flood insurance (yearly amount)	SSIP	Supplementary Security income past 12 months
MHP	Mobile home costs (yearly amount)	SSP	Social Security income past 12 months
MRGP	First mortgage payment (monthly amount)	WAGP	Wages or salary income past 12 months
RMSP	Number of Rooms		
RNTP	Monthly rent	Variable	Person Variables (Top and Bottom Coded)
SMP	Total payment on all second and junior mortgages and home equity loans (monthly amount)	INTP	Interest, dividends, and net rental income past 12 months (signed)
TAXAMT	Property taxes (yearly real estate taxes)	SEMP	Self-employment income past 12 months (signed)
VALP	Property value		
WATP	Water cost (yearly cost)		

Note: Due to the unique circumstances during the pandemic, the PUMS files for data year 2020 use experimental weights. More information on the affect the pandemic had on the American Community Survey (ACS) as well as how and why experimental weights were used for the ACS and PUMS may be found on the ACS Experimental Data webpage:

<https://www.census.gov/programs-surveys/acs/data/experimental-data.html>.

The table above shows the variables categorized as either being top coded only or being top coded and bottom coded. These top codes and bottom codes were derived based on the following rules:

1. All top coded variables are top coded using the state mean of all cases greater than or equal to the top code state threshold value.
2. Income variables which can have negative values (Self-employment Income, or SEMP, and Interest, Dividends, and Net Rental Income, or INTP) are bottom coded using the state mean of all cases less than or equal to the bottom code state threshold value.

The state top coded and bottom coded thresholds and mean values may be found in the CSV file called “2020 ACS PUMS Top & Bottom Coded Values” on the ACS Experimental Weight page: <https://www.census.gov/programs-surveys/acs/data/experimental-data.html>.

Previous year’s CSV files may be found on the PUMS Documentation webpage: <https://www.census.gov/programs-surveys/acs/microdata/documentation.html>

In the CSV file, the top code threshold variable ends in “TPCT”. The mean top code variable name begins with “T_”. For example, for the “Electricity cost (monthly cost)” PUMS variable called “ELEP”, the top code threshold variable name is called “ELETPCT” and the top code mean variable name is called “T_ELE”.

On the PUMS data files, for each state there will be values for ELEP below the state top code threshold value. However, at or above the threshold value all records will only have the top code value. Note that the top code threshold variable and mean top code variable are not on the PUMS data files. In addition, if there are blank values for the top code threshold and mean value for a variable then no records were top coded for that particular state.

The naming convention for the top code mean value was first used beginning with the 2017 PUMS files. For the PUMS top code CSV files from 2009 through 2016, the top code mean variables has the same name as the variable on the PUMS data files that is top coded (for example, “ELEP”). For the PUMS top code CSV files from 2005 through

2008, the “P” at end of the variable name is not included (e.g. “ELE”). The top code threshold variable has used the same naming convention for this entire time period.

For bottom coded variables, the bottom code threshold variable name ends in “BPCT” and the bottom code mean variable name begins with “B_”. For example, “SEMBPCT” and “B_SEM”. Similar to the top code variables, this naming convention for the bottom code mean value variable was first used in 2017. For 2016 and prior years, the bottom code mean value variables are called BINT and BSEM, respectively. The bottom code threshold variable name has remained the same since 2005.