Data Dictionary for HUD Public Use Microdata Sample (PUMS)

Version 18, Updated 02/10/2025

I. Programs

The PUMS contains separate datasets for households assisted by programs administered by HUD's Office of Public and Indian Housing (PIH) and Office of Housing. Data for two of the largest PIH programs are provided: Public Housing, and the Housing Choice Voucher program (HCVP), as are data on three Office of Housing multifamily programs: Project Based Section 8, Section 202, and Section 811.

HCVP tenants find rental units in the private market, paying at least 30 percent of their income toward rent and utilities, with the remainder subsidized by a HUD voucher. Public Housing tenants live in housing owned and operated by public housing agencies. Project Based Section 8 tenants live in privately owned and operated developments, where some or all of the housing units are HUD subsidized. Section 202 subsidizes supportive housing for the elderly, and Section 811 subsidizes supportive housing for persons with disabilities.

II. Data Sources

Household characteristics for households in PIH programs were extracted from HUD's Inventory Management System/PIH Information Center (IMS/PIC) data system.[1] Characteristics for households in Office of Housing programs were extracted from HUD's Tenant Rental Assistance Certification System (TRACS) data system.[2]

Both IMS/PIC and TRACS are transaction based. The PUMS are based on the most recent certification transaction for each household at the end of the calendar year. For most households, the record is at most 18 months old. For public housing and HCVP households assisted by housing agencies in the Moving to Work demonstration program, the record is, at most, three years old.[3]

III. Sample Design

The PUMS are 5 percent samples, without replacement, of tenant records for the 50 states, DC, and Puerto Rico. The samples are stratified by state and program. The samples contain approximately 5 percent of records for each stratum.

The precise sample size for each stratum was determined by Neyman Allocation, a statistical method giving a greater sample size for strata with more diverse populations.[4] The standard deviation of adjusted household income was used as a proxy for diversity. The sample for each

[1] http://portal.hud.gov/hudportal/HUD?src=/program_offices/public_indian_housing/systems/pic/about
[2] http://portal.hud.gov/hudportal/HUD?src=/program_offices/housing/mfh/trx/trxsum
[3] http://portal.hud.gov/hudportal/HUD?src=/program_offices/public_indian_housing/programs/ph/mtw
[4] See Lohr, *Sampling Design and Analysis* , Duxbury Press 1999, p. 108.

stratum is proportional to tenants multiplied by the standard deviation of adjusted household income.

The data sets contain a weight (variable "weight") which is the inverse of the sampling probability. The weight is the same for all households in a given strata. Weighting makes the samples nationally representative.

IV. Confidentiality

To ensure confidentiality, some variables were masked for some households. State is suppressed for tenants with unique combinations of variables in the population. If a tenant is uniquely identified after suppressing state, one or more additional variables are suppressed. For sampling purposes, households with suppressed state were treated as a separate state.

Four variables in the sample were top-coded to ensure confidentiality. Household size is top-coded at 7, and number of bedrooms is top-coded at 4. Household income and adjusted household income are both top-coded at $90,000.

V. Variables

This section describes each of the 15 variables in the PUMS.

1. Program

Program is identified by character variable "prog". For Public Housing tenants, this variable equals "P". For HCVP tenants, this variable equals "V". For Project Based Section 8 tenants, this variable equals "Sec 8". For Section 202 tenants, this variable equals "Sec 202". For Section 811 tenants, this variable equals "Sec 811".

2. State

States are identified by two character variables. Variable "state" contains the mixed-case state name, and variable "fipst" contains the state FIPS code. Missing or suppressed state names are set to "Z Missing"; missing or suppressed fips codes are set to "99".

3. Strata

Sampling strata are identified by variable "strata", which is a concatenation of the state FIPS code (variable "fipst") and program (variable "prog").

4. Census Tract Poverty Rate

Census tract poverty rates are measured by character variable "poverty". Poverty rates for 2009 and 2010 PUMS are based on American Community Survey (ACS) data averaged over 2005-09, using 2000 tract boundaries. Poverty rates for 2012 PUMS are based on ACS data averaged over 2007-11, using 2010 tract boundaries. Poverty rates for 2013 PUMS are based on ACS data averaged over 2008-12, using 2010 tract boundaries. Poverty rates for 2014 PUMS are based on ACS data averaged over 2009-13, using 2010 tract boundaries. Poverty rates for 2015 PUMS

are based on ACS data averaged over 2010-14, using 2010 tract boundaries. Poverty rates for 2016 are based on ACS data averaged over 2011-15, using 2010 tract boundaries. Poverty rates for 2017 are based on ACS data averaged over 2012-16, using 2010 tract boundaries. Poverty rates for 2018 are based on ACS data averaged over 2013-17, using 2010 tract boundaries. Poverty rates for 2019 are based on ACS data averaged over 2014-18, using 2010 tract boundaries. Poverty rates for 2020 are based on ACS data averaged over 2014-18, using 2010 tract boundaries. Poverty rates for 2021 are based on ACS data averaged over 2015-19, using 2010 tract boundaries. Poverty rates for 2023 are based on ACS data averaged over 2016-20, using 2020 tract boundaries. Poverty rates for 2024 are based on ACS data averaged over 2016-20, using 2020 tract boundaries. Poverty rates are reported in four five categories: 0%-9%, 10%-19%, 20%-29%, 30%-39%, and 40% and above. Missing or suppressed values are blank.

5. Urban/Rural Status

Urban/rural status is measured by character variable “UR”. For urban areas, UR equals “U”. For rural areas, UR equals “R”. Missing or suppressed values are blank.

6. Metropolitan Status

Metropolitan status is measured by character variable “metro”. For metropolitan areas, metro equals “Metropolitan”. For micropolitan areas, metro equals “Micropolitan”. Non-metropolitan areas are defined as areas outside of Core Based Statistical Areas (CBSAs). For non-metropolitan areas, metro equals “Non-CBSA”. Missing or suppressed values are blank.

7. Household Type

Household type is measured by character variable “H6”. H6 equals “1” for households with an elderly head or spouse, and no children. H6 equals “2” for non-elderly households where the head or spouse has disabilities, and there are no children. H6 equals “3” for other households with no children. H6 equals “4” for households with an elderly head or spouse, with children. H6 equals “5” for non-elderly households where the head or spouse has disabilities, with children present. H6 equals “6” for other households with children. Elderly is defined as age 62 and above.

8. Household Members

Household member counts are measured by numeric variable “mbrs”, taking values of 1-7. Household members is top-coded at 7, thus 7 represents 7 or more householders. Missing or suppressed values are blank.

9. Bedrooms

Bedroom counts are measured by numeric variable “bdrms”, taking values of 0-4. Bedroom counts are top-coded at 4, thus 4 represents 4 or more bedrooms. Missing or suppressed values are blank.

10. Race/Ethnicity of Household Head

Race and ethnicity of the household head is measured by character variable "race_eth". Race and ethnicity are reported in four categories. For Hispanics of any race, race_eth equals "Hispanic". For non-Hispanic whites, race_eth equals "White". For non-Hispanic blacks, race_eth equals "Black". For other non-Hispanics, race_eth equals "Other". Missing or suppressed values are blank.

11. Sex of Household Head

Sex of household head is measured by character variable "sex". Males are coded as "M", and females are coded as "F". Missing or suppressed values are blank.

12. Household Income

Annual household income is measured by character variable "inc5000".

The 2009 PUMS reports income in 13 categories. The first 12 categories report income in $5000 increments: 0-$5000, $5001-$10,000,…, $55,001-$60,000. The thirteenth category is for households with income between $60,001 and $90,000. Income is top-coded at $90,000.

Starting in 2010, the PUMS reports income in 12 categories. The first eight categories report income in $2,500 increments: 0-$2,500, $2,501-$5,000, … ,$17,501-$20,000. The next two categories report income in $5,000 increments: $20,001-$25,000, and $25,001-$30,000. The 11th category is for households with income between $30,001 and $40,000. The 12th category is for households with income between $40,001 and $90,000. Income is top-coded at $90,000.

13. Adjusted Household Income

Eligibility for HUD rental assistance programs is based on adjusted household income. Adjusted income is calculated by subtracting off certain expenses from household income. Details of the calculation for PIH programs are reported on HUD form 50058[5], and on HUD form 50059[6] for Office of Housing programs. Annual adjusted household income is reported in the same categories as annual household income, and is also top-coded at $90,000.

Adjusted household annual income is measured by character variable "ainc5000". Adjusted household annual income is reported in the same categories as annual household income, and is also top-coded at $90,000.

14. Rent Burden

Rent burden is defined as gross rent (including utility costs) divided by monthly adjusted household income. Rent burden is undefined for households with $0 adjusted income.

Rent burden is measured by character variable "brdn". Rent burden is reported in four categories: 0%-31%, 32%-39%, 40%-49%, and 50% and above.

[5] http://portal.hud.gov/hudportal/HUD?src=/program_offices/public_indian_housing/systems/pic/50058
[6] http://www.hud.gov/offices/adm/hudclips/forms/files/50059.pdf

15. Survey Weight

The survey weight is measured by variable “weight”, which is the inverse of the sampling probability. The weight is the same for all households in a given stratum.