

**Documentation
Describing the Topcoding
and Other Confidentiality
Measures that Occur on
the**

**American Housing Surveys
(AHS)
Public Use Files**



U.S. Department of Housing and Urban Development
Office of Policy Development and Research



Documentation describing the topcoding and other confidentiality measures that occur on the American Housing Surveys (AHS) public use files

The purpose of this paper is to document the variables on the AHS public use microdata files that are affected by topcoding, bottomcoding, and/or other measures to preserve the confidentiality of the survey's respondents. Hopefully, it will explain why some data user's tabulations using certain AHS variables from the public use microdata files do not match the numbers in the AHS publications.

Topcoding and bottomcoding

A number of the variables on the AHS public use microdata files are topcoded and/or bottomcoded to preserve the survey respondent's confidentiality. Currently, variables are topcoded in a number of different ways. Some variables are topcoded at the 97th percentile. Some variables are topcoded at the 99.5th percentile. Some variables are topcoded and/or bottomcoded at a predetermined value. Some variables are topcoded using the topcode of a different variable. The lists below indicate which variables are topcoded and/or bottomcoded using which method.

The topcodes are calculated differently for the AHS-National and the AHS-Metropolitan surveys. The topcodes are calculated using the entire data file for the AHS-National. In the AHS-Metropolitan survey, the topcodes are calculated uniquely for each MSA that is enumerated in the survey year, except when one or more Metropolitan areas don't have enough cases with valid values for an individual variable, then that variable's topcode is calculated using all cases surveyed. The categories below apply to both National and Metropolitan surveys.

Variables topcoded at the 97th percentile

AMMORT, AMMRT2, AMMRT3, AMMRT4, AMRTZ, AMRTZ2, AMTE, AMTF, AMTG, AMTI, AMTM, AMTO, AMTT, AMTW, AMTX, CASH, CASH2, CONFEE, CPRICE, HEBAM1, HEBAM2, HEBAM3, HECR1, HECR2, HECR3, HEPMT1, HEPMT2, HEPMT3, LODRNT, LOT, LPRICE, LRENT, LVALUE, MHOTFE, PMAMT2, PMIAMT, PMT, PMT2, PMT3, PMT4, PRENT, PVALUE, RENT, RESMOR, RESMR2, TERM, TERM2, UNITSF, VALUE

Variables topcoded at the 99.5th percentile

CSTMNT, RAD, SAL, PVOTHER, ZINCN

Note: The topcode for RAD is computed individually for each type of job reported.
The topcodes for AMTX and CONFEE are computed on the individual variables needed to create them.

Variables topcoded and/or bottomcoded at a specific value

AGE	- topcoded at 90
*INUSYR	- bottomcoded at (survey year - 90)
*ROOMS	- topcoded at 21
*TPARK	- topcoded at 2
*ZADULT	- topcoded at 11
PVOTHER	- bottomcoded at -1
ZINCN	- bottomcoded at -1

Note: The listed topcode actually appears on the microdata file for the variables marked with an ‘*’. The mean of all cases at or above the topcode appears on the microdata file for the other variables.

Variables topcoded at the 97th percentile within geographic combinations

CLIMB, FLOORS, NUNITS

Note: For the AHS-National survey, each of these variables has a topcode computed for each SMSA-METRO3 combination shown on the public use microdata file.

For the AHS-Metropolitan survey, each of these variables has a topcode computed for each SMSA-ZONE combination shown on the public use microdata file.

Variables created using the variables which have been topcoded

VOTHER & VOTHER2 – Each are the sum of the appropriate PVOTHERs associated with the household.

ZINC & ZINC2 – Each are the sum of the appropriate PVOTHERs and SALs associated with the household.

Other things to note about the topcoding/bottomcoding operation

- 1. The actual topcode/bottomcode value that appears on the public use microdata file**
Only where the variable is marked with a ‘*’ in the list above does the specified topcode or bottomcode value actually appear on the public use microdata file. Otherwise, the mean value for all cases equal to or greater than the topcode is calculated and displayed on the public use microdata file.
- 2. Number of cases used to compute the mean for the topcodes**
The Census Bureau’s Disclosure Review Board’s policy is that there must be at least three cases included in the calculation of this mean. It is not unusual in the AHS surveys, particularly in the AHS-Metropolitan survey, for a variable’s universe of cases to be so small that there is not a minimum of three cases greater than or equal to the topcode predetermined or calculated for that variable. In these instances, the value of the topcode is lowered until there are at least three cases that can be included in the calculation of the mean. In the rare instances where there are not three eligible cases in the entire universe for a variable, no topcode is used at all. In the AHS-Metropolitan survey, there are occasions where a variable may have a large enough universe to obtain at least three cases for the mean in some MSAs, but not in others. When this happens, the topcode and its related mean are calculated across all MSAs enumerated in that survey year.
- 3. Longitudinal topcodes**
Topcodes for variables that describe the physical characteristics of housing units that are not likely to change over time are compared to topcodes used in previous enumerations of the AHS survey. The larger of the topcode cutoff between the current survey year and the largest topcode ever used in a previous survey year is the topcode that will be used for the current survey year. The variables that fall into this category are CLIMB, FLOORS, LOT, NUNITS, and UNITSF.
- 4. Special topcoding of mortgage data**
The Disclosure Review Board told us that we must ensure that the following variables must not sum to more than the topcode for the variable VALUE: AMMORT, AMMRT2, AMMRT3, AMMRT4, HEBAM1, HEBAM2, HEBAM3, HECR1, HECR2, HECR3. If the sum of these variables does exceed the topcode for VALUE, the variables are changed via an algorithm to ensure the sum is less than or equal to the topcode for VALUE.

Other Confidentiality Measures Used

In addition to topcoding and bottomcoding variables, a number of other variables on the public use microdata file are reviewed and possibly changed to protect the respondent's confidentiality. The Census Bureau requires that no geographic areas in a sample survey containing less than 100,000 in population be identified on a public use microdata file. The AHS public use microdata files contain a number of geographic variables. We take steps to ensure that no combination of these geographic variables will identify an area with a population less than 100,000 by either suppressing or changing the values of one or more of these variables. Likewise, steps are taken to eliminate rare events that could allow a respondent to be identified. Details documenting the variables affected are described next.

Geographic variables changed and/or suppressed to comply with the 100,000 population rule

These variables are changed and/or suppressed in the AHS-National survey:
CMSA, DEGREE, METRO3, SMSA

These variables are changed and/or suppressed in the AHS-Metropolitan survey:
COUNTY, LONGIT, METRO, PSUDOTCT, STATE

Other variables changed to preserve the respondent's confidentiality

AGE	this variable is changed in a number of "rare event" situations
AMTX	the values are rounded to the nearest integer in the sequence 5, 15, 25, 35, 45, . . .
CONDO	cooperative units are combined with condominium units
CONFEE	the values are rounded to the nearest integer in the sequence 50, 150, 250, 350, . . .
NATVTY	ethnic and/or political affiliations are collapsed into broad categories