Quick Guide to Estimating Errors for Summary Table Users: 1973 to Current

LAST UPDATED: MARCH 2017

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
U.S. CENSUS BUREAU
U.S. DEPARTMENT OF COMMERCE

Contents

Exhibits

1. Overview

The American Housing Survey (AHS) data can be used to produce estimates of many housing statistics in the United States. The U.S. Department of Housing and Urban Development (HUD) and the Census Bureau provide AHS data to the public in two ways, summary tables and public use file (PUF) microdata. The AHS includes both a national longitudinal sample and metropolitan area longitudinal oversamples; summary tables and PUF microdata are available for both.

The AHS is a longitudinal survey and survey estimates generally come with some error. Any given survey has potentially six sources of error.[1] Exhibit 1.1 describes the sources of error and what AHS users should know about each error type.

Exhibit 1.1. Sources of Error in the AHS

Type of Error	Description	What AHS Users Need to Know
Validity	The failure to design a survey question that accurately measures the construct of interest.	HUD and the Census Bureau carefully test each new survey question to ensure it is measuring the construct of interest. There are no additional steps AHS users must take to account for validity error.
Sampling error	Refers to how the estimates vary with respect to selecting different samples from the same universe and with the same sample design versus observing the entire universe.	Errors from sampling are estimated and made available to AHS users. See the section Errors From Sampling in this guide.
Coverage error	The failure to give some units in the target population any chance of selection into the sample (undercoverage), or giving units more than one chance of selection (overcoverage).	HUD and the Census Bureau correct for potential overcoverage or undercoverage by controlling AHS estimates to official estimates or specific demographic characteristics at specific levels of geography. Additional details are available in the Sample Design document associated with each survey. There are no additional steps AHS users must take to account for coverage error.
Survey nonresponse error	The failure to collect complete data from all units in the sample.	HUD and the Census Bureau correct for nonresponse by implementing noninterview adjustment factors into the weighting process. Additional details are available in the Sample Design document associated with each survey. There are no additional steps AHS users must take to account for nonresponse error.
Measurement error: missing question responses	Errors introduced by missing responses due to various factors, including respondents' inability to answer questions or interviewer error.	The AHS has several hundred questions across numerous topics. Some AHS respondents are unable to answer some questions, and others refuse to answer questions. For some survey questions, HUD and the Census Bureau impute missing values. For others, the response is coded as a refusal. Errors from missing responses are made available to AHS users but with plenty of caveats. See the section Errors From Missing Data in this guide.
Measurement error: inaccurate question responses	Errors introduced from inaccurate responses due to various factors, including effects of interviewers on the respondents' answers to survey questions, respondents' inability to	The AHS has several hundred questions across numerous topics. It is likely that AHS respondents will give incorrect answers to some survey questions. From time to time, HUD and the Census Bureau commission studies to determine if

[1] Groves, Robert. 1989. *Survey Errors and Survey Costs*. New York: Wiley.

Type of Error	Description	What AHS Users Need to Know
	answer questions, faulty wording of survey questions, or mode effects.	certain responses are accurate or if responses are consistent across time. These studies are listed in Errors From Inaccurate Data in this guide.
Processing error	Errors introduced after the data are collected, including data capture errors, data coding and classification errors, and data editing and imputation errors.	HUD and the Census Bureau carefully test all aspects of the data capture, coding, classification, editing, and imputation procedures. Although mistakes are possible, HUD and the Census Bureau believe they are minimal. If a processing error is discovered, HUD and the Census Bureau will let AHS users know and, in some cases, will publish revised estimates. There are no additional steps AHS users must take to account for processing error.

2. Errors From Sampling

The AHS is a longitudinal survey with a complex sample design. As such, each estimate is subject to some error because of the design of the sample. Many AHS users are interested in determining the precision (amount of sampling error) in their AHS estimate of interest, which is often expressed through a confidence interval. The calculation of confidence intervals requires an estimate of variance or the square root of the variance, which is known as the standard error.

For AHS summary table users, HUD and the Census Bureau published errors from sampling estimates in three ways depending on the year of the survey. The three methods are abbreviated standard error tables, generalized variance functions, and full standard error tables. For the first two methods, numerous examples are available in the survey year-specific documentation referenced in the previous section; we encourage users to see those examples for further information.

2.1. Abbreviated Errors From Sampling Standard Error Tables

The abbreviated standard error tables provide standard errors for small subset of general estimate values. Users must locate their summary table estimate of interest, then use the abbreviated standard error table to estimate a standard error by interpolation.

For instance, to estimate the standard error of an estimate of two-bedroom single-family homes, the user finds the estimate of two-bedroom single-family homes (15,000,000) in the summary tables. Using the following abbreviated standard error table, the user can estimate the standard error by interpolating between the standard error values for estimate sizes of 10,000,000 and 25,000,000. A linear interpolation method is suggested.[2]

---

[2] A linear interpolation formula is available at http://formulas.tutorvista.com/math/interpolation-formula.html.

Exhibit 2.1. Abbreviated Errors

Size of Estimate	Standard Error	Size of Estimate	Standard Error
25,000	5,000	2,500,000	50,000
50,000	7,000	5,000,000	70,000
100,000	10,000	10,000,000	96,000
250,000	16,000	25,000,000	134,000
500,000	23,000	50,000,000	138,000
1,00,000	32,000		

3. Generalized Variance Functions (aka Formulas for Computing Errors)

Generalized Variance Functions (GVFs) are formulas that can be used to estimate standard error from sampling. The parameters of the formula are provided. Users must locate the correct GVF for their estimate of interest and parameters to use in the GVF.

For instance, to estimate the standard error of an estimate of two-bedroom single-family homes, the user finds the estimate of two-bedroom single-family homes (15,000,000) in the summary tables. The user then finds the GVF that is appropriate for an estimate of standard error for two-bedroom single-family homes:

$$standard\ error = \sqrt{6.68\ x\ A - .000050\ x\ A^2}.$$

The user then plugs in the value of 15,000,000 into the A value in the GVF to produce the standard error estimate.

4. Full Errors From Sampling Standard Error Tables

The full standard error tables include standard error estimates for every cell presented in the summary tables. The standard error estimates were made using replicate weights. Users simply need to find the summary table estimate of interest, then locate the corresponding standard error estimate in the standard error table.

4.2. Where to Find Errors From Sampling Standard Error Estimates

Exhibit 4.2 describes where to find standard error estimates for each survey, as well as what method of presenting standard errors is available. Within each source are numerous examples.

Exhibit 4.2. Locations of Errors From Sampling Standard Error Estimates

Survey Year	Method	Where To Find
National, 1973–1989	Abbreviated SE tables	19XX AHS National Source and Accuracy
Metropolitan area, 1974–1989	Abbreviated SE tables	19XX AHS Metropolitan Source and Accuracy
National, 1991	Abbreviated SE tables and GVF	1991 AHS National Source and Accuracy
Metropolitan area, 1990–1992	Abbreviated SE tables and GVF	19XX AHS Metropolitan Source and Accuracy
National, 1993–2007	Abbreviated SE tables and GVF	19XX AHS National Errors 19XX AHS National Supplement Errors*
Metropolitan area, 1993–2007	Abbreviated SE tables and GVF	19XX AHS Metropolitan Errors

Survey Year	Method	Where To Find
		19XX AHS Metropolitan Supplement Errors*
National, 2009–2013	Abbreviated SE tables and GVF	20XX AHS National Errors
National, 2009–2013	Full SE tables	AHS 20XX National Summary Report and Tables (in Excel format)
Metropolitan area, 2009–2013	Abbreviated SE tables and GVF	20XX AHS Metropolitan Errors
National and metropolitan area, 2015 and later	GVF	20XX AHS Sample Design, Weighting, and Errors

*Supplements are not present in every year.

5. Errors From Missing Data

Missing data occurs because people refuse to answer certain questions, do not know answers, or provide an answer that is clearly erroneous and is treated as though it were missing. For some questions with missing responses, HUD and the Census Bureau impute missing values according to rules that subject matter specialists developed. The rules generally assume that the true value for a missing response is similar to a value reported by a similar respondent. However, HUD and the Census Bureau do not know how close the imputed values are to the actual values.

Incompleteness can cause large errors. For example, when a particular question misses only 10 percent of homes, that represents about 13.5 million homes that have to be estimated on little or no basis (about 135 million homes are in the United States). To give users a sense of the bias caused by missing data, most of the errors in the documents referenced in Table 2 also include an Errors for Incomplete Data Bias table. This table can be used when a particular survey estimate is based on one or more variables with completeness rates of 50 to 90 percent. The values in the table are based on a 1990 analysis conduct by the Census Bureau. The analysis showed the estimated the standard error from missing data to be:

> $.0012 \times U + .0363 \times$ (lesser of A or $U-A$)), where A is any count from the AHS and U is the total number of housing units in the U.S. or metropolitan area (both in thousands, result also in thousands).

Because of the large number of variables in the AHS, HUD and the Census Bureau typically do not publish completeness rates for individual survey estimates. AHS users interested in completeness rates should consider using the AHS PUF microdata to estimate completeness rates. When using the PUF to estimate completeness rates, users should be aware of the following.

- PUF variables with a value of "not applicable" are not considered missing. Not applicable means the question corresponding to the variable was not asked of the AHS respondent, because the respondent was not eligible to be asked the question. For instance, if a respondent reported living in an apartment building, the respondent is not asked questions about mobile home features.

- Most AHS variables have a corresponding edit/allocation flag variable that indicates whether the variable's value was imputed for the respondent. The edit/allocation flag variables are the same as the variable name but are preceded by the letter J. For instance, if a respondent did not report a value for the *HFUEL* (heating fuel) variable but the respondent's value was imputed, the variable *JHFUEL* will equal 2, indicating an imputation.

- The edit/allocation flag will take a value of 1 if the respondent's reported value was edited. Edited values should not be considered missing.

- The correct way to calculate a completeness rate in the AHS is—

(Sum of all respondents with a reported value – sum of respondents with an imputed value) / (sum of respondents with a reported value + sum of respondents with a "missing," "don't know," and "refused" value).

6. Errors From Inaccurate Data: Studies of Response Accuracy in the AHS and Other Household Surveys

The AHS typically includes a few hundred questions. It is likely that some respondents will provide inaccurate responses to some questions. Generally, as the inaccurate responses are random among the sample, estimates generated from the responses will not be biased. However, the confidence interval around any estimate could be larger because of additional error.

A handful of studies have attempted to measure inaccurate responses to certain survey questions. These studies are listed in the following table and are available at huduser.gov/portal/datasets/ahs.html.

Exhibit 6.1. Data Errors – A Study

Title	Focus	Findings
Between-Survey Changes in the Number of Bedrooms in a Unit	Number of bedrooms	Evidence of variation in the number of bedrooms reported by respondents for the same housing units. Findings suggest some inaccurate reporting by respondents.
American Housing Survey: Housing Adequacy and Quality as Measured by the AHS	Various measures of housing quality	Numerous findings.
Streamlining the American Housing Survey	Various survey measures	Numerous findings.

U.S. Department of Housing and Urban Development

Office of Policy Development and Research

Washington, DC 20410-6000