

How to Use the Data: Accuracy of the Data

1996

Introduction

The data contained in these Profiles, Summary Tape File tables and Public Use Microdata Samples are based on the American Community Survey (ACS) sample interviewed in 1996. The ACS is designed to provide accurate estimates for the housing units and household population of the four sites participating in the 1996 ACS. The ACS, like any other statistical activity, is subject to error. The purpose of this documentation is to provide data users with a basic understanding of the ACS sample design, estimation methodology, and accuracy of the ACS data.

Sample Design

- Sites -- Three urban sites: Multnomah County/Portland, OR, Rockland County, NY and Brevard County, FL -- and one rural site: Fulton County, PA, participated in the 1996 ACS. The primary sampling unit was the housing unit, including all occupants. Persons living in group quarters were not included in the sample.
- Master Address File -- In the urban sites the Census Bureau developed a Master Address File (MAF) that served as the housing unit sampling frame for the survey. The MAF was constructed by an automated match of the sites' 1990 Census Address Control File and the United States Postal Service 1995 Delivery Sequence File. For the Fulton County site the Census Bureau compiled a sampling frame by canvassing and listing each housing unit in the county, then computerizing the list.
- Sampling Rates -- An automated procedure designated the sample units. Two sampling rates were employed. Most of the housing units selected into the survey were sampled at a rate of 15 percent. For functioning governmental units in which there were fewer than 1,000 housing units on the sampling frame, the sampling rate was 30 percent. All of Fulton County and a few governmental units in the three urban sites were sampled at 30 percent. A variable sampling rate was used for the purpose of providing relatively more reliable estimates for small areas.
- Public Use Microdata Sample -- A stratified systematic sample with probability equal to the housing unit weight was used to select units for the public use microdata sample (PUMS) file. Three sampling rates were used to account for subsampling for the field personal visit cases and for the furlough adjustment. Housing units were stratified to improve the reliability of estimates derived from the PUMS. The PUMS was selected to represent roughly 5% of the population universe for each site.

Data Collection

Three data collection modes were used to conduct the 1996 ACS: Mail, Computer Assisted Telephone Interviewing (CATI), and Computer Assisted Personal Interviewing (CAPI).

These three modes are described below.

- Mail Phase -- The Mail phase began with a prenotice letter mailed to each housing unit on the next to last Wednesday of the month preceding the sample month. The ACS questionnaire was mailed one week later, followed by a reminder card one week after that. A replacement questionnaire was mailed three weeks later if the original questionnaire had not yet been checked in at the processing site. Check-in of mail return questionnaires for a sample panel was cut off at the start of the third month following the sample month.
- CATI Phase -- Approximately five weeks after the mailout of the original ACS questionnaire, the CATI staff began contacting non-responding sample households by telephone. Late mail returns were removed from the CATI workload on a daily basis. This phase of nonresponse follow-up lasted for approximately four weeks.
- CAPI Phase -- The CAPI universe consisted of all outstanding non-response cases remaining after the completion of the CATI phase. A 1 in 3 subsample was selected from the outstanding cases and forwarded to the field interviewers. Field interviewers visited each assigned housing unit and attempted to conduct an interview. Late mail returns were removed from the CAPI workload on a daily basis. The CAPI phase of nonresponse follow-up lasted approximately one month.

Confidentiality of the Data

- Confidentiality Edit -- To maintain the confidentiality required by law (Title 13, United States Code), the Census Bureau applies a confidentiality edit to the ACS data to assure that published data do not disclose information about specific individuals, households, or housing units. As a result, a small amount of uncertainty is introduced into the estimates of ACS characteristics. The sample itself provides adequate protection for most areas for which sample data are published since the resulting data are estimates of the actual characteristics. However, small areas require more protection. The confidentiality edit is implemented by identifying a subset of individual housing units from the sample data files as having a unique combination of specified person and household characteristics within a block group. The confidentiality edit is controlled so that the basic structure of the data is preserved.

Errors in the Data

- Sampling Error -- The data in the ACS products are estimates of the actual figures that would have been obtained by interviewing the entire population using the same methodology. The estimates from the chosen sample also differ from other samples of housing units and persons within those housing units. Sampling error in data arises due to the use of probability sampling, which is necessary to insure the integrity and representativeness of sample survey results. The implementation of statistical sampling procedures provides the basis for the statistical analysis of sample data.
- Nonsampling Error -- In addition to sampling error, data users should realize that other types of errors may be introduced during any of the various complex operations used to collect and process survey data. For example, operations such as editing, reviewing, or keying data from questionnaires may introduce error into the estimates. These and other sources of error contribute to the nonsampling error component of the total error of survey estimates. Nonsampling errors may affect the data in two ways. Errors that are introduced randomly increase the variability of the data. Systematic errors which are consistent in one direction introduce bias into the results of a sample survey. The Census Bureau protects against the effect of systematic errors on survey estimates by conducting extensive research and evaluation programs on sampling techniques, questionnaire design, and data collection and processing procedures. In addition, an important goal of the ACS is to minimize the amount of nonsampling error introduced through nonresponse for sample housing units. One way of accomplishing this is by following up on mail nonrespondents during the CATI and CAPI phases.
- Standard Errors -- The standard error is a measure of the deviation of a sample estimate from the average of all possible samples. Sampling errors

and some types of nonsampling errors are estimated by the standard error. The sample estimate and its estimated standard error permit the construction of interval estimates with a prescribed confidence that the interval includes the average result of all possible samples. The next section describes the method of calculating standard errors and confidence intervals for the estimates in this ACS product.

Calculation of Standard Errors

Direct Standard Errors

- **Methodology Used** -- Direct estimates of the standard errors were calculated for all estimates reported in this product. They are provided in the Profiles and in tables of Summary Tape File estimates. The standard errors, in most cases, are calculated using standard variance estimation software using a methodology that takes into account the sample design and estimation procedures.
- **Exceptions** -- There are four cases for which the direct standard error estimates are not appropriate.
 1. The estimate of the *number* or *proportion* of people, households, housing units or families in a geographic area with a specific characteristic is zero. A special procedure was used to estimate the standard error.
 2. There are no sample observations available to compute an estimate of a proportion or per capita amount or an estimate of its standard error. The estimate is represented in the tables by "--" and the standard error estimate by ""*"".
 3. Only a small number of *identical values* are reported and used to calculate an aggregate, median, mean or per capita amount. In this case, there are too few sample observations to compute a stable estimate of the standard error. The standard error estimate is represented in the tables by ""*"".
 4. The estimate of the *number* of people having a specified characteristic is *controlled* to be equal to an independently derived population estimate at the county level. These county estimates are produced by the Census Bureau's Population Estimates Program using standard demographic analysis techniques. For these cases the standard error is zero. (See "Estimation Procedure" for a further explanation.)

- **Sums and Differences** -- The standard errors estimated from these tables are for individual estimates. Additional calculations are required to estimate the standard errors for sums of and differences between two sample estimates. The estimate of the standard error of a sum or difference is approximately the square root of the sum of the two individual standard errors squared; that is, for standard errors and of estimates and :

This method, however, will underestimate (overestimate) the standard error if the two items in a sum are highly positively (negatively) correlated or if the two items in a difference are highly negatively (positively) correlated.

- **Ratios** -- Frequently, the statistic of interest is the ratio of two variables, where the numerator may or may not be a subset of the denominator. The standard error of the ratio between two sample estimates is approximated as follows:
- **Confidence Intervals** -- A sample estimate and its estimated standard error may be used to construct confidence intervals about the estimate. These intervals are ranges that will contain the average value of the estimated characteristic that results over all possible samples, with a known probability.

For example, if all possible samples that could result under the 1996 ACS sample design were independently selected and surveyed under the same conditions, and if the estimate and its estimated standard error were calculated for each of these samples, then:

1. Approximately 68 percent of the intervals from one estimated standard error below the estimate to one estimated standard error above the estimate would contain the average result from all possible samples;
2. Approximately 90 percent of the intervals from 1.65 times the estimated standard error below the estimate to 1.65 times the estimated standard error above the estimate would contain the average result from all possible samples.
3. Approximately 95 percent of the intervals from two estimated standard errors below the estimate to two estimated standard errors above the estimate would contain the average result from all possible samples.

The intervals are referred to as 68 percent, 90 percent, and 95 percent confidence intervals, respectively.

- **Confidence Intervals of Ratios, Sums, and Differences** -- Confidence intervals also may be constructed for the ratio, sum of, or difference between two sample figures. This is done by first computing the ratio, sum, or difference, then obtaining the standard error of the ratio, sum, or difference (using the formulas given earlier), and finally forming a confidence interval for this estimated ratio, sum, or difference as above. One can then say with specified confidence that this interval includes the ratio, sum, or difference that would have been obtained by averaging the results from all possible samples.

Limitations -- The user should be careful when computing and interpreting confidence intervals.

1. The estimated standard errors included in this data product do not include all portions of the variability due to nonsampling error that may be present in the data. In particular, the standard errors do not reflect the effect of correlated errors introduced by interviewers, coders, or other field or processing personnel. Thus, the standard errors calculated represent a lower bound of the total error. As a result, confidence intervals formed using these estimated standard errors may not meet the stated levels of confidence (i.e., 68, 90, or 95 percent). Thus, some care must be exercised in the interpretation of the data in this data product based on the estimated standard errors.
2. **Zero or small estimates; very large estimates** -- The value of almost all ACS characteristics is greater than or equal to zero by definition. For zero or small estimates use of the method given previously for calculating confidence intervals relies on large sample theory, and may result in negative values which for most characteristics are not admissible. In this case the lower limit of the confidence interval is set to zero by default. A similar caution holds for estimates of totals close to a control total or estimated ratios near one, where the upper limit of the confidence interval is set to its largest admissible value. In these situations the level of confidence of the adjusted range of values is less than the prescribed confidence level.
3. **Small tabulation areas** -- Estimates for small tabulation areas (particularly small areas with high nonresponse rate) are based on small samples. Again, using large sample theory methodology to construct confidence intervals may result in values that are not admissible for the characteristic of interest or in overstating the confidence for a given range. The user should exercise caution in the analysis of data for these areas.

Generalized Standard Errors

The information provided in Tables A through C can be used to approximate the standard errors of sample estimates of *totals* and *proportions* in the Profiles and Summary Tape File tables, and from the PUMS. Tables A and B give the basic standard error for an estimate of a characteristic that would result under a simple random sampling design. The estimates are for person, family, and housing unit characteristics. Design factors by subject are provided in Table C. The term "subject" refers to a characteristic, such as age for persons, tenure for housing units, and poverty for families. The design factors reflect the effects of the actual sample design and estimation procedures used for the 1996 American Community Survey. Details of the sample design and estimation procedures are provided elsewhere in this chapter.

To approximate the standard error of an estimate of a total or a proportion using Tables A through C follow the steps described in the next section. A proportion is defined as a ratio of two estimates where the numerator is a subset of the denominator. For example, the proportion of Black lawyers is the ratio of Black lawyers to all lawyers.

An inspection of the formulas used to calculate the simple random sampling standard errors suggests that when dealing with zero estimates or very small estimates of totals and percentages the standard error estimates approach zero. This is also the case for very large estimates of totals and percentages. Zero or small estimates, like any other sample estimates, are still subject to sampling variability and therefore an estimated standard error of zero or close to zero is not adequate. For an estimated total that is less than 75 or within 75 of the total size of the tabulation area, use a basic standard error of 21. For estimated percentages that are less than 2 or greater than 98, use the basic standard errors in Table B that are shown in the "2 or 98" row.

Confidence intervals can be constructed from generalized standard errors just as they are from direct standard errors. However, for estimates other than totals and proportions, generalized standard errors cannot be calculated from Tables A through C.

1. To get standard errors for estimates of *means, medians, aggregates* and *per capita amounts* the user should use direct standard errors since generalized standard errors are not provided for these estimates. If two or more tabulation areas are combined to obtain the median value of a characteristic, the user is referred to "Medians" below for a description of how to approximate the standard error. The method relies on linear interpolation. The standard error approximations that result from this method may not be as adequate for characteristics that deviate from the linear assumption. Users are recommended to exercise caution when analyzing these characteristics.
2. Table C includes a design factor for "Children Ever Born." However, the standard errors shown in Table A and the formula provided below the table are not appropriate to calculate standard errors for fertility estimates. The user should use direct standard errors for these estimates (Table P28.) Direct standard errors are not available for estimates of areas formed by combining two or more standard tabulation areas or for estimates derived from the PUMS.

Approximate the basic standard error using the formula , where e is the estimate of "Children Ever Born" and w is the unweighted number of women in a specific age group. e is obtained from P28 and w is approximated by the product f , where f is the site level sampling fraction and t is the total estimate of women in the specific age group. Note that this method provides adequate standard error estimates for PUMS estimates; however, the approximation deteriorates as the size of the tabulation area (or the specific age group) decreases. To obtain the adjusted standard error follow the procedure explained in the section "Use of Tables to Approximate Standard Errors." Standard error estimates approximated using this methodology are conservative and therefore the specified confidence of statistical intervals calculated as described in "Confidence Intervals" may be overstated.

Use of Tables to Approximate Standard Errors

Tables A through C are used in the following manner to approximate standard errors:

1. Obtain the basic standard error from either Table A (for an estimate of a total) or Table B (for an estimate of a percentage) or use the formula given below the appropriate table. Obtain the number of persons, number of housing units, or number of families for the county in the appropriate matrix for estimates of total. Use these numbers to determine which column to look under in Table A. When working with the PUMS, multiply this basic standard error by 1.83.
2. Use Table C to obtain the appropriate design factor for the characteristic; for example, educational attainment or ancestry. Multiply the basic standard error by this factor to get the ACS standard error estimate.

Medians -- For the standard error of the median of a characteristic, it is necessary to examine the distribution from which the median is derived, as the estimated number of persons, households, families or housing units with the characteristic and the distribution of the characteristic affect the standard error. An approximate method is given here. As the first step, compute one-half of the estimated number having the characteristic on which the median is based (refer to this result as $B/2$). Treat $B/2$ as if it were an ordinary estimate and obtain its standard error as instructed above. Compute the desired confidence interval about $B/2$. Starting with the lowest value of the characteristic, cumulate the frequencies in each category of the characteristic until the sum equals or first exceeds the lower limit of the confidence interval about $B/2$. By linear interpolation, obtain a value of the characteristic corresponding to this sum. This is the lower limit of the confidence interval of the median. In a similar manner, continue cumulating frequencies until the sum equals or exceeds the count in excess of the upper limit of the interval about $B/2$. Interpolate as before to obtain the upper limit of the confidence interval for the estimated median.

When interpolation is required in the upper open-ended interval of a distribution to obtain a confidence bound, use 1.5 times the lower limit of the open-ended confidence interval as the upper limit of the open-ended interval.

The following examples of standard errors and confidence intervals are based on real data from the 1996 ACS Test.

Example 1 - Proportion or Percentage Estimate

The estimated poverty rate for census tract 601 in Brevard County is 19.3 percent. The base of the estimated percentage is 6,063. From Table B, use the row corresponding to "20 or 80" percent and interpolating between the two columns 5,000 and 7,500 one can approximate the basic standard error of this estimate as follows:

$$\text{BasicSE}(19.3) = 1.4 - (1.4 - 1.1) \cdot [(6,063 - 5,000) / (7,500 - 5,000)] = 1.3.$$

The design factor for "Poverty Status in the Past 12 Months (persons)" for Brevard County is 1.8. The approximate standard error estimate for the estimated poverty rate of 19.3 percent is determined by multiplying the basic standard error 1.3 by the design factor 1.8 from Table C. This yields an estimated standard error of 2.3. (The level of precision on each calculation is the same as for the estimates.)

To avoid interpolation use the formula given below Table B. The use of the formula is illustrated here.

$$\text{BasicSE}(19.3) =$$

(Note that the two basic standard error estimates are not identical.) Again, to get the final standard error of 2.2 multiply 1.2 by the design factor 1.8.

To calculate the lower and upper bounds of the 90 percent confidence interval around 19.3 percent using the second final standard error, simply multiply 2.2 by 1.65, then add and subtract the product from 19.3. Thus the 90 percent confidence interval for this estimated percentage is found to be

$$[19.3 - 1.65(2.2)] \text{ to } [19.3 + 1.65(2.2)] \text{ or } 15.7 \text{ to } 22.9.$$

Example 2 - Total Estimate

Consider the data in example 1. The estimate of persons in poverty in census tract 601 in Brevard County is 1,171. From Table A, use the column labeled "400,000" and interpolating between the rows 1,000 and 2,500, approximate the basic standard error as follows:

$$\text{BasicSE}(1,171) = 119 - (119 - 75) \cdot [(2,500 - 1,171) / (2,500 - 1,000)] = 80.$$

Multiply 80 by the design factor 1.8 to approximate the ACS standard error estimate. The standard error estimate is found to be 144 persons.

Avoid linear interpolation by using the formula given below Table A. The population of Brevard county is 447,597. Thus,

$$\text{BasicSE}(1,171) = 82.$$

Note that using the formula yields a slightly different result. Multiply 82 by the design factor 1.8 to get a final standard error estimate equal to 147. (Keep in mind that the two results are approximations of the standard error estimate.)

Proceed as before to construct a 90 percent confidence interval around the estimate of 1,171 persons in poverty. The upper and lower bounds of the 90 percent confidence interval are

$[1,171 - 1.65(147)]$ to $[1,171 + 1.65(147)]$ or 928 to 1,414.

Example 3 - Difference Between Two Sample Estimates

The following is an illustration of the calculations required to construct a confidence interval for the estimated difference between two sample estimates. The 1996 ACS poverty rate estimates for census tracts 601 and 611 in Brevard County are 19.3 and 5.2 percent, respectively. The obvious question in comparing these two areas is whether the two areas are really different with respect to the characteristic of interest or is the apparent difference just the result of the use of sampling. The difference in the poverty rate for the two tracts is 14.1 percent. To compute the final standard error estimate of the difference use the formula given in "Sums and Differences."

First, calculate the standard error estimate of 5.2 and combine it with the results obtained in example 1, as follows (the design factors are incorporated into the calculation):

$SE(14.1) = .$

The 90 percent confidence interval for the difference is computed as before:

$[14.1 - 1.65(2.6)]$ to $[14.1 + 1.65(2.6)]$ or 9.8 to 18.4.

Since this confidence interval doesn't include zero, we can conclude with 90 percent confidence that the poverty rates in the two tracts are really different.

Example 4 - Ratio of Two Estimates

For reasonably large areas or large samples ratio estimates are normally distributed. The methods described in the previous examples can be used to calculate a confidence interval around a ratio estimate. Suppose that one wished to express the poverty rate of census tract 601 relative to the poverty rate of census tract 611. The ratio of the two estimates of interest is

$19.3/5.2 = 3.7$. Thus, the poverty rate of census tract 601 is 3.7 times higher than the corresponding rate of census tract 601. Using the formula to calculate the standard error of a ratio estimate we have:

$= 1.0$

Using the result above, the 90 percent confidence interval for this ratio is

$[3.7 - 1.65(1.0)]$ to $[3.7 + 1.65(1.0)]$ or 2.0 to 5.4.

Example 5 - Median Estimate

Compute a 90 percent confidence interval for median adjusted household income in the Brevard County Profile. This median is computed using all of the 183,251 households in Brevard County. Half of that number is $91,626 = B/2$. To avoid interpolation in finding the basic standard error, use the formula given below Table A.

Multiply 511 by the design factor 1.3 for "Household Income in the Past 12 Months" from Table C to get the final standard error of 664.

Calculate the 90 percent confidence interval bounds around $B/2 = 91,626$.

$[91,626 - 1.65(664)]$ to $[91,626 + 1.65(664)]$ or 90,530 to 92,722.

Use the Profile "1996 Adjusted Income" table to determine the confidence interval for the median. The table below shows the number of households in each income category. A column of cumulative totals has been added. From this table, it is clear that the 90,530th household has an income that falls in the range \$25,000 to \$34,999.

1996 ADJUSTED INCOME Households	Number in Range	Cumulative Number
Less than \$5,000	7,490	7,490
\$5,000 to \$9,999	12,191	19,681
\$10,000 to \$14,999	15,105	34,786
\$15,000 to \$24,999	32,972	67,758
\$25,000 to \$34,999	29,648	97,406
\$35,000 to \$49,999	34,457	131,863
\$50,000 to \$74,999	31,556	163,419
\$75,000 to \$99,999	11,377	174,796
\$100,000 to \$149,999	6,242	181,038
\$150,000 or more	2,213	183,251

Interpolating for the lower bound gives:

The upper bound also falls in the \$25,000 to \$34,999 range and is computed in a similar fashion.

ESTIMATION PROCEDURE

The estimates that appear in this product are obtained from a raking ratio estimation procedure that results in the assignment of two sets of weights: a weight to each sample person record and a weight to each sample housing unit record. For any given tabulation area, a characteristic total is estimated by summing the weights assigned to the persons, households, families or housing units possessing the characteristic in the tabulation area. Estimates of person characteristics are based on the person weight and estimates of family, household or housing unit characteristics are based on the housing unit weight.

Each sample person or housing unit record is assigned exactly one weight to be used to produce estimates of all characteristics. For example, if the weight given to a sample person or housing unit had the value 6, all characteristics of that person or housing unit would be tabulated with the weight of

6. The estimation procedure, however, does assign weights varying from person to person or housing unit to housing unit.

The estimation procedure used to assign the weights was performed independently within each of the 1996 ACS sites.

- Initial Housing Unit Weighting Factors - This process produced the following factors:

- Base Weight (BW) - This factor was assigned to every housing unit based on its sampling stratum and is the inverse of the housing unit's sampling rate. Base weights were either 3.3333 or 6.6667.
- Unduplication Adjustment Factor (UAF) - Addresses already included in other Census Bureau surveys were not subjected to sampling. This factor adjusted the base weight to account for the probability that an address had already been selected into some other survey's sample. Factors were computed and assigned based on the following groups.

Site x County Block Type x 1990 Census Address Control File Status

- CAPI Subsampling Factor (SSF) - The weights of the CAPI cases were adjusted to reflect the results of CAPI subsampling. This factor was assigned to each record as follows:

Selected in CAPI subsampling: SSF = 3.0

Not selected in CAPI subsampling: SSF = 0.0

Not a CAPI case: SSF = 1.0

- Variation in Monthly Response by Mode (VMS) - This factor made the total weight of the Mail, Delivery, CATI, and CAPI records to be tabulated in a month equal to the total base weight of all cases originally mailed for that month. The value of VMS for Mail and Delivery cases was 1.0. For CATI and CAPI cases, VMS was computed and assigned based on the following groups.

Site x Month

- Noninterview Factor (NIF) - This factor adjusted the weight of *all* responding occupied housing units to account for both responding and non-responding housing units. This factor was computed in two stages: NIF1 and NIF2. NIF1 was computed and assigned to occupied housing units based on the following groups.

Site x County x Building Type¹ x Tract

After having adjusted each occupied housing unit for NIF1, NIF2 was computed and assigned to occupied housing units based on the following groups.

Site x Building Type x Month

NIF was then computed for each occupied housing unit as the product of NIF1 and NIF2. Vacancies were assigned a value of NIF = 1.0.

Non-responding housing units were now assigned a weight of 0.0.

- Noninterview Factor - Mode (NIFM) - This factor adjusted the weight of just the responding CAPI occupied housing units to account for both CAPI respondents and all non-respondents. This factor was computed as if NIF had not already been assigned to every occupied housing unit record. This factor was not used directly but rather as part of computing the next factor: MBF. NIFM was computed and assigned to occupied CAPI housing units based on the following groups.

Site x Building Type x Month

Mail, Delivery and CATI cases received a value of NIFM = 1.0. Vacancies received a value of NIFM = 1.0.

- Mode Bias Factor (MBF) - This factor made the total weight of the groups below the same as if NIFM had been used instead of NIF. MBF was computed and assigned to occupied housing units based on the following groups.

Site x Tenure² x Month x Marital Status³

Vacancies received a value of MBF = 1.0.

- Furlough Adjustment Factor (FAF) - This factor adjusted the weights of the February 1996 CAPI records to account for the "missing" January 1996 CAPI cases caused by the government-wide furlough and bad weather in Dec 95/Jan 96. This factor was assigned to each record as follows:

Feb 1996 CAPI records: FAF 2.0⁴

All other records: FAF = 1.0

- First Housing Unit Post-Stratification Factor (HPSF1) - This factor made the number of housing units in a tract equal to the tract counts from the 1997 Master Address File (MAF), after all factors through FAF had been applied. HPSF1 was computed and assigned to all housing units based on the following groups.

Site x County x Tract

- Person Weighting Factors - Initially each person in an occupied housing unit received the housing unit weight as their person weight. At this point everyone in the household had the same weight. These person weights were then individually adjusted based on each person's age, race, sex, and Hispanic origin as described below.
- Person Post-Stratification Factor (PPSF) - This factor was applied to individual persons based on their age, race, sex and Hispanic origin. It adjusts the person weights so that the weighted sample counts will match county population control counts by age, race, sex, and Hispanic origin. These population control counts are independently derived by the Census Bureau's Population Division.

This is an iterative procedure that first computes PPSF using the following groups:

County x Race⁵ x Sex⁶ x Age Groups used for Race adjustment

After applying the value of PPSF computed above, a second stage value of PPSF is computed using the following groups:

County x Hispanic Origin⁷ x Sex x Age Groups used for Hispanic adjustment

The above two steps were repeated up to six times, or until the change to PPSF from one iteration to the next became small.

- Rounding - The final product of all person weights (BW x . . . x HPSF1 x PPSF) was rounded to an integer. Rounding was performed so that the sum of the rounded weights was within one person of the sum of the unrounded weights for any of the groups listed below:

County

County x Race
 County x Race x Hisp
 County x Race x Hisp x Sex
 County x Race x Hisp x Sex x Age
 County x Race x Hisp x Sex x Age x Tract
 County x Race x Hisp x Sex x Age x Tract x Block

For example, the estimate of the number of White, Hispanic, Males, Age 30 using the rounded weights is within one of the number produced using the unrounded weights.

- Final Housing Unit Weighting Factors - This process produced the following factors:
- Principal Person Factor (PPF) - This factor adjusted for differential response depending on the race, Hispanic origin, sex, and age of the principal person in the household. The principal person was defined as the female spouse of the responding householder. If there was no such person, then the responding householder was the principal person.

The value of PPF was the PPSF of the principal person.

- Second Housing Unit Post-stratification Factor (HPSF2) - This factor made the number of housing units in a tract again equal to the 1997 MAF control count totals after all factors (BW x . . . x HPSF1 x PPF) had been applied. HPSF2 was computed and assigned to all housing units based on the following groups.

Site x County x Tract

- Rounding - The final product of all housing unit weights (BW x . . . x HPSF1 x PPF x HPSF2) was rounded to an integer. Rounding was performed so that total rounded weight was within one housing unit of the total unrounded weight for any of the groups listed below:

Site
 Site x County
 Site x County x Tract
 Site x County x Tract x Block

CONTROL OF NONSAMPLING ERROR

As mentioned earlier, sample data are subject to nonsampling error. This component of error could introduce serious bias into the data, and the total error could increase dramatically over that which would result purely from sampling. While it is impossible to completely eliminate nonsampling error from a survey operation, the Census Bureau attempts to control the sources of such error during the collection and processing operations. Described below are the primary sources of nonsampling error and the programs instituted for control of this error. The success of these programs, however, is contingent upon how well the instructions actually were carried out during the survey.

- Undercoverage C It is possible for some sample housing units or persons to be missed entirely by the survey. The undercoverage of persons and housing units can introduce biases into the data. A major way to avoid undercoverage in a survey is to ensure that its sampling frame, for ACS an address list in each site, is as complete and accurate as possible.

The source of addresses in the three urban sites was a new product, the Master Address File (MAF), currently being developed by the Census Bureau. The MAF is created by combining the 1990 Census Address Control File and the Delivery Sequence File of the United States Postal Service. An attempt is made to assign all appropriate geographic codes to each MAF address via an automated procedure using the Census Bureau TIGER files. A manual coding operation based in the appropriate regional offices is attempted for addresses which could not be automatically coded. The MAF was used as the source of addresses for selecting sample housing units and mailing questionnaires. TIGER produced the location maps for personal visit CAPI assignments.

In Fulton County, PA the Census Bureau conducted a manual listing and map-spotting operation in the summer of 1995. Interviewers hand delivered ACS questionnaires to the addresses from this listing.

In the CATI and CAPI nonresponse follow-up phases, efforts were made to minimize the chances that housing units that were not part of the sample were interviewed in place of units in sample by mistake. If a CATI interviewer called a mail nonresponse case and was not able to reach the exact address, no interview was conducted and the case was eligible for CAPI. During CAPI follow-up, the interviewer had to locate the exact address for each sample housing unit. In some multi-unit structures the interviewer could not locate the exact sample unit or found a different number of units than expected. In these cases the interviewers were instructed to list the units in the building and follow a specific procedure to select a replacement sample unit.

- Respondent and Interviewer Error -- The person answering the questionnaire or responding to the questions posed by an interviewer could serve as a source of error, although the questions were phrased as clearly as possible based on testing, and detailed instructions for completing the questionnaire were provided to each household. In addition, respondents' answers were edited for completeness, and problems were followed up as necessary.
- Interviewer monitoring -- The interviewer may misinterpret or otherwise incorrectly enter information given by a respondent; may fail to collect some of the information for a person or household; or may collect data for households that were not designated as part of the sample. To control these problems, the work of interviewers was monitored carefully. Field staff were prepared for their tasks by using specially developed training packages that included hands-on experience in using survey materials. A sample of the households interviewed by CAPI interviewers was reinterviewed to control for the possibility that interviewers may have fabricated data.
- Item Nonresponse -- Nonresponse to particular questions on the survey questionnaire and instrument allows for the introduction of bias into the data, since the characteristics of the nonrespondents have not been observed and may differ from those reported by respondents. As a result, any imputation procedure using respondent data may not completely reflect this difference either at the elemental level (individual person or housing unit) or on the average.

Some protection against the introduction of large biases is afforded by minimizing nonresponse. In the ACS, nonresponse for the CATI and CAPI operations was reduced substantially by the requirement that the automated instrument receive a response to each question before the next one could be asked. For mail responses, the clerical edit and follow-up operations were aimed at obtaining a response for every question on selected questionnaires. Values for any items that remain unanswered were imputed by computer using reported data for a person or housing unit with similar characteristics.

- Clerical Review -- Questionnaires returned by mail were edited for completeness and acceptability. They were reviewed by clerks for content omissions and population coverage. If necessary, a telephone follow-up was made to obtain missing information. Potential coverage errors were included in this follow-up, as well as questionnaires with too many omissions to be accepted as returned.
- Processing Error -- The many phases involved in processing the survey data represent potential sources for the introduction of nonsampling error.

The processing of the survey questionnaires includes the clerical editing, follow-up by telephone, and keying of data from completed questionnaires; the manual coding of write-in responses; and the electronic data processing. The various field, coding and computer operations undergo a number of quality control checks to insure their accurate application.

- Automated Editing -- After data collection was completed, any remaining incomplete or inconsistent information was imputed during the final automated edit of the collected data. Imputations, or computer assignments of acceptable codes in place of unacceptable entries or blanks, were needed most often when an entry for a given item was lacking or when the information reported for a person or housing unit on that item was inconsistent with other information for that same person or housing unit. As in other surveys and previous censuses, the general procedure for changing unacceptable entries was to assign an entry for a person or housing unit that was consistent with entries for persons or housing units with similar characteristics. Assigning acceptable values in place of blanks or unacceptable entries enhances the usefulness of the data.

Table A. Unadjusted Standard Error for Estimated Totals

[Based on a 15 percent simple random sample.

For estimates from the PUMS multiply standard errors from the table or the formula by 1.83.]

Estimate ¹	Size of publication Area ²											
	500	750	1,000	2,500	5,000	7,500	10,000	25,000	50,000	100,000	250,000	400,000
75	21	21	21	21	21	21	21	21	21	21	21	21
125	23	24	25	26	26	26	27	27	27	27	27	27
250	27	31	33	36	37	37	37	38	38	38	38	38
500	.	31	38	48	51	52	52	53	53	53	53	53
750	.	.	33	55	60	62	63	64	65	65	65	65
1,000	.	.	.	58	68	70	72	74	75	75	75	75
2,500	84	97	103	113	116	118	119	119
5,000	97	119	151	160	165	167	168
7,500	103	173	191	199	204	205
10,000	185	214	226	234	236
15,000	185	245	270	283	287
25,000	267	327	358	366
50,000	377	477	499
75,000	327	547	589
100,000	585	654
250,000	731

¹ To get a better approximation of the standard error of an estimated count use the formula below.

N = Size of area
= Estimate of characteristic total

² The population estimate of the county the tabulation area is in if the estimate is a person characteristic, the estimate of housing units in the county if the estimate is a housing characteristic, or the estimate of families if the estimate is a family characteristic.

Table B. Unadjusted Standard Error in Percentage Points for Estimated Percentages

[Based on a 15 percent simple random sample.

For estimates from the PUMS multiply standard errors from the table or the formula by 1.83.]

Estimated Percentage	Base of Estimated Percentage ¹											
	500	750	1,000	2,500	5,000	7,500	10,000	25,000	50,000	100,000	250,000	400,000
2 or 98	1.4	1.1	1.0	0.6	0.4	0.4	0.3	0.2	0.1	0.1	0.1	0.1
5 or 95	2.3	1.9	1.6	1.0	0.7	0.6	0.5	0.3	0.2	0.2	0.1	0.1
10 or 90	3.2	2.6	2.3	1.4	1.0	0.8	0.7	0.5	0.3	0.2	0.1	0.1
15 or 85	3.9	3.1	2.7	1.7	1.2	1.0	0.9	0.5	0.4	0.3	0.2	0.1
20 or 80	4.3	3.5	3.0	1.9	1.4	1.1	1.0	0.6	0.4	0.3	0.2	0.2
25 or 75	4.6	3.8	3.3	2.1	1.5	1.2	1.0	0.7	0.5	0.3	0.2	0.2
30 or 70	5.0	4.0	3.5	2.2	1.5	1.3	1.1	0.7	0.5	0.3	0.2	0.2
35 or 65	5.1	4.2	3.6	2.3	1.6	1.3	1.1	0.7	0.5	0.4	0.2	0.2
40 or 60	5.2	4.3	3.7	2.3	1.6	1.4	1.2	0.7	0.5	0.4	0.2	0.2
50	5.3	4.4	3.8	2.4	1.7	1.4	1.2	0.8	0.5	0.4	0.2	0.2

¹ For an estimated percentage not shown in the table, use the formula below to get standard error approximations. Use this table only for proportions, that is, where the numerator is a subset of the denominator.

B = Base of estimated percentage
= Estimated percentage

Table C. Standard Error Design Factors - 1996 ACS Test

[Design factors are site specific, use the appropriate column]

Characteristics	Brevard	Rockland	Multnomah	Fulton
Population				
Persons	1.6	1.7	1.6	1.8
Families	1.0	0.9	1.0	1.4
Households	0.8	0.6	0.7	0.9
Age	1.5	1.5	1.5	1.1
Sex	1.4	1.5	1.3	1.1
Race	1.7	2.0	1.6	1.5
Hispanic Origin	1.7	1.8	1.6	1.5
Marital Status	1.2	1.2	1.3	0.9
Ancestry	1.8	2.0	1.8	1.6

Household Size	1.2	1.3	1.2	1.0
Household Type and Relationship	1.3	1.3	1.3	1.1
Children Ever Born	1.3	1.3	1.4	1.3
Work Disability and Functional Limitation	1.4	1.4	1.3	1.0
Place of Birth	1.7	1.8	1.7	1.6
Residence 5 years ago	1.9	2.0	1.8	1.7
Year of Entry	1.7	2.0	2.4	1.0
Language Spoken at Home and English Ability	1.5	1.8	1.4	1.4
Educational Attainment	1.4	1.5	1.4	1.2
School Enrollment	1.4	1.3	1.3	1.1
Family Type	1.2	1.3	1.2	1.1
Employment Status	1.3	1.3	1.2	0.9
Industry	1.5	1.5	1.4	1.1
Occupation	1.5	1.5	1.4	1.1
Class of Worker	1.5	1.5	1.4	1.1
Hours Per Week and Weeks Worked in past 12 months	1.3	1.3	1.3	1.0
Number of Workers in Family	1.2	1.2	1.2	1.0
Place of Work	1.5	1.4	1.3	1.1
Means of Transportation to Work	1.4	1.4	1.4	1.1
Travel Time to Work	1.5	1.5	1.4	1.2
Private Vehicle Occupancy	1.4	1.4	1.4	1.1
Time Leaving Home to Go to Work	1.5	1.5	1.4	1.1
Type of Income in the Past 12 Months	0.9	0.8	0.8	0.8
Household Income in the Past 12 Months	1.3	1.4	1.3	1.1
Family Income in the Past 12 Months	1.3	1.4	1.3	1.1
Poverty Status in the Past 12 Months (persons)	1.8	1.9	1.8	1.5
Poverty Status in the Past 12 Months (families)	1.3	1.3	1.2	1.1
Armed Forces and Veteran Status	1.3	1.2	1.2	0.9
Housing				
Housing units	0.5	0.5	0.7	1.1
Age of Householder	1.3	1.3	1.3	1.0
Race of Householder	1.1	1.1	1.1	0.9
Hispanic Origin of Householder	1.5	1.6	1.7	1.0
Condominium Status	0.9	0.9	0.9	0.8
Units in Structure	1.1	1.1	1.0	0.9
Occupied by Tenure	1.0	1.1	1.0	0.9
Vacant	1.9	1.9	1.9	1.3
Gross Rent	1.5	1.5	1.4	1.1
Year Structure Built	1.1	1.2	1.2	1.0
Rooms, Bedrooms	1.2	1.3	1.2	1.0
Kitchen Facilities, Plumbing Facilities, and Source of Water	0.6	0.5	0.5	0.7
Heating System	0.6	0.4	0.7	0.8
Sewage Disposal	0.7	0.6	0.6	0.9
House Heating Fuel	0.9	0.7	1.1	0.9
Telephone	1.0	1.0	1.0	0.9
Vehicles Available	1.2	1.3	1.2	1.0
Length at Residence	1.3	1.3	1.3	1.1
Mortgage Status and SMOG	1.2	1.2	1.2	1.1
Gross Rent as a Percent of Household Income	1.6	1.6	1.4	1.2
SMOG as a Percent of Household Income	1.2	1.3	1.2	1.0
Air Conditioning	0.7	1.1	1.2	1.0
Value	1.2	1.2	1.2	1.0
Aggregate Persons by Tenure by Units in Structure	2.0	2.2	2.0	2.0

Footnotes:

- (1) Single-Unit: Only one housing unit at the basic street address, or Multi-Unit: More than one housing unit at the basic street address. Back
- (2) Owner, Renter, or Not applicable. "Not applicable" applies to temporarily occupied housing units. Back
- (3) Married/Widowed, or Other Back
- (4) A temporary duplicate of each February 1996 CAPI record was created prior to estimation and assigned a tabulation month of January 1996. This temporary record was weighted separately through the factor MBF. In the FAF stage, the total weight of this temporary January 1996 CAPI record was added to the total weight of its corresponding original February CAPI record. The temporary record was then deleted. Back
- (5) White, Black, or Other Back
- (6) Male or Female Back
- (7) Hispanic or Non-Hispanic Back

Top of page

Source: U.S. Census Bureau, Demographic Surveys Division,
American Community Survey Office
Created: Wednesday May 29, 2002
Last revised: Wednesday August 25, 2010