

# Source of the Data and Accuracy of the Estimates for the 2015 Annual Social and Economic Supplement Microdata File

## Table of Contents

<b>SOURCE OF THE DATA.....</b>	<b>1</b>
Basic CPS.....	1
The 2015 Annual Social and Economic Supplement .....	4
Estimation Procedure .....	4
<b>ACCURACY OF THE ESTIMATES .....</b>	<b>5</b>
Sampling Error.....	5
Nonsampling Error.....	5
Nonresponse.....	6
Sufficient Partial Interview .....	6
Undercoverage .....	6
Comparability of Data.....	7
A Nonsampling Error Warning.....	8
Estimation of Median Incomes .....	8
Standard Errors and Their Use.....	9
Estimating Standard Errors .....	9
Generalized Variance Parameters .....	10
Standard Errors of Estimated Numbers .....	11
Standard Errors of Estimated Percentages.....	12
Standard Errors of Estimated Differences .....	13
Standard Errors of Estimated Ratios.....	15
Standard Errors of Estimated Medians .....	17
Standard Errors of Averages for Grouped Data.....	19
Standard Errors of Estimated Per Capita Deficits.....	21
Accuracy of State Estimates .....	22
Standard Errors of State Estimates .....	23
Standard Errors of Regional Estimates .....	24
Standard Errors of Groups of States .....	24
Standard Errors of Data for Combined Years.....	25
Standard Errors of Differences of 2-Year Averages.....	27
Standard Errors of Quarterly or Yearly Averages .....	28
Technical Assistance.....	28
Year-to-Year Factors .....	33
<b>REFERENCES.....</b>	<b>34</b>

## Tables

Table 1. Description of the March Basic CPS and CPS ASEC Sample Cases.....	3
Table 2. CPS Coverage Ratios: March 2015 .....	7
Table 3. Estimation Groups of Interest and Generalized Variance Parameters.....	11
Table 4. Parameters for Computation of Standard Errors for Labor Force Characteristics: March 2015 .....	29
Table 5. Parameters for Computation of Standard Errors for People and Families: 2015 CPS ASEC .....	30
Table 6. CPS Year-to-Year Correlation Coefficients for Income and Health Insurance Characteristics: 1961 to 2015.....	31
Table 7. CPS Year-To-Year Correlation Coefficients for Poverty Characteristics: 1971 to 2015.....	32
Table 8. Factors and Populations for State Standard Errors and Parameters: 2015 CPS ASEC ..	32
Table 9. Factors and Populations for Regional Standard Errors and Parameters: 2015 CPS ASEC .....	33

## Source of the Data and Accuracy of the Estimates for the 2015 Annual Social and Economic Supplement Microdata File

### SOURCE OF THE DATA

The data in this microdata file are from the 2015 Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS). The U.S. Census Bureau conducts the CPS ASEC over a 3-month period in February, March, and April, with most of the data collection occurring in the month of March. The CPS ASEC uses two sets of questions, the basic CPS and a set of supplemental questions. The CPS, sponsored jointly by the Census Bureau and the U.S. Bureau of Labor Statistics, is the country's primary source of labor force statistics for the entire population. The Census Bureau and the Bureau of Labor Statistics also jointly sponsor the CPS ASEC.

**Basic CPS.** The monthly CPS collects primarily labor force data about the civilian noninstitutionalized population living in the United States. The institutionalized population, which is excluded from the population universe, is composed primarily of the population in correctional institutions and nursing homes (98 percent of the 4 million institutionalized people in Census 2010). Interviewers ask questions concerning labor force participation about each member 15 years old and over in sample households. Typically, the week containing the nineteenth of the month is the interview week. The week containing the twelfth is the reference week (i.e., the week about which the labor force questions are asked).

The CPS uses a multistage probability sample based on the results of the decennial census, with coverage in all 50 states and the District of Columbia. The sample is continually updated to account for new residential construction. When files from the most recent decennial census become available, the Census Bureau gradually introduces a new sample design for the CPS.

Every ten years the CPS first stage sample is redesigned<sup>1</sup> reflecting changes based on the most recent decennial census. In the first stage of the sampling process, primary sampling units (PSUs)<sup>2</sup> were selected for sample. In the 2000 design, the United States was divided into 2,025 PSUs. These were then grouped into 824 strata and one PSU was selected for sample from each stratum. In the 2010 sample design, the United States was divided into 1,987 PSUs. These PSUs were then grouped into 852 strata. Within each stratum, a single PSU was chosen for the sample, with its probability of selection proportional to its population as of the most recent decennial census. In the case of strata consisting of only one PSU, the PSU was chosen with certainty.

In April 2014, the Census Bureau began phasing out the 2000 sample and replacing it with the 2010 sample, creating a mixed sampling frame. Two simultaneous changes occur during this phase-in period. First, within the PSUs selected for both the 2000 and 2010 designs, sample households from the 2010 design gradually replace sample households selected for the 2000 design. Second, new PSUs selected for only the 2010 design gradually replace outgoing PSUs

---

<sup>1</sup> For detailed information on the 2010 sample redesign, please see reference [1].

<sup>2</sup> The PSUs correspond to substate areas (i.e., counties or groups of counties) that are geographically contiguous.

selected for only the 2000 design. By July 2015, the new 2010 sample design will be completely implemented and the sample will come entirely from the 2010 redesigned sample.

Approximately 74,000 housing units were selected for sample from the sampling frame for the basic CPS. Based on eligibility criteria, 12 percent of these housing units were sent directly to computer-assisted telephone interviewing (CATI). The remaining units were assigned to interviewers for computer-assisted personal interviewing (CAPI).<sup>3</sup> Of all housing units in sample, about 61,100 were determined to be eligible for interview. Interviewers obtained interviews at about 52,900 of these units. Noninterviews occur when the occupants are not found at home after repeated calls or are unavailable for some other reason. Table 1 summarizes historical changes in the CPS design.

---

<sup>3</sup> For further information on CATI and CAPI and the eligibility criteria, please see reference [2].

Time period	Number of sample PSUs	Basic CPS housing units eligible		Total (CPS ASEC/ADS <sup>1</sup> + basic CPS) housing units eligible	
		Interviewed	Not interviewed	Interviewed	Not interviewed
<b>2015</b>	852	52,900	8,200	74,300	10,300
<b>2014 Redesign</b>	824	17,200	2,200	22,700	2,600
<b>2014 Traditional</b>	824	35,500	4,600	51,500	5,800
<b>2014</b>	824	52,700	6,800	--	--
<b>2013</b>	824	52,900	6,400	75,500	7,700
<b>2012</b>	824	53,300	5,800	75,100	7,200
<b>2011</b>	824	53,400	5,300	75,900	6,500
<b>2010</b>	824	54,100	4,600	77,000	5,700
<b>2009</b>	824	54,100	4,600	76,200	5,700
<b>2008</b>	824	53,800	5,100	75,900	6,400
<b>2007</b>	824	53,700	5,600	75,500	7,100
<b>2006</b>	824	54,000	5,400	76,000	7,100
<b>2005</b>	754/824 <sup>2</sup>	54,400	5,700	76,500	7,500
<b>2004</b>	754	55,000	5,200	77,700	7,000
<b>2003</b>	754	55,500	4,500	78,300	6,800
<b>2002</b>	754	55,500	4,500	78,300	6,600
<b>2001</b>	754	46,800	3,200	49,600	4,300
<b>2000</b>	754	46,800	3,200	51,000	3,700
<b>1999</b>	754	46,800	3,200	50,800	4,300
<b>1998</b>	754	46,800	3,200	50,400	5,200
<b>1997</b>	754	46,800	3,200	50,300	3,900
<b>1996</b>	754	46,800	3,200	49,700	4,100
<b>1995</b>	792	56,700	3,300	59,200	3,800
<b>1990 to 1994</b>	729	57,400	2,600	59,900	3,100
<b>1989</b>	729	53,600	2,500	56,100	3,000
<b>1986 to 1988</b>	729	57,000	2,500	59,500	3,000
<b>1985</b>	629/729 <sup>3</sup>	57,000	2,500	59,500	3,000
<b>1982 to 1984</b>	629	59,000	2,500	61,500	3,000
<b>1980 to 1981</b>	629	65,500	3,000	68,000	3,500
<b>1977 to 1979</b>	614	55,000	3,000	58,000	3,500
<b>1976</b>	624	46,500	2,500	49,000	3,000
<b>1973 to 1975</b>	461	46,500	2,500	49,000	3,000
<b>1972</b>	449/461 <sup>4</sup>	45,000	2,000	45,000	2,000
<b>1967 to 1971</b>	449	48,000	2,000	48,000	2,000
<b>1963 to 1966</b>	357	33,400	1,200	33,400	1,200
<b>1960 to 1962</b>	333	33,400	1,200	33,400	1,200
<b>1959</b>	330	33,400	1,200	33,400	1,200

1 The CPS ASEC was referred to as the Annual Demographic Survey (ADS) until 2002.

2 The Census Bureau redesigned the CPS following the Census 2000. During phase-in of the new design, housing units from the new and old designs were in the sample.

3 The Census Bureau redesigned the CPS following the 1980 Decennial Census of Population and Housing.

4 The Census Bureau redesigned the CPS following the 1970 Decennial Census of Population and Housing.

**The 2015 Annual Social and Economic Supplement.** In addition to the basic CPS questions, interviewers asked supplementary questions for the CPS ASEC. They asked these questions of the civilian noninstitutional population and also of military personnel who live in households with at least one other civilian adult. The additional questions covered the following topics:

- Household and family characteristics
- Marital status
- Geographic mobility
- Foreign-born population
- Income from the previous calendar year
- Poverty
- Work status/occupation
- Health insurance coverage
- Program participation
- Educational attainment

Including the basic CPS sample, approximately 99,500 housing units were in sample for the CPS ASEC. About 84,600 housing units were determined to be eligible for interview, and about 74,300 interviews were obtained (see Table 1).

The additional sample for the CPS ASEC provides more reliable data for Hispanic households, non-Hispanic minority households, and non-Hispanic White households with children 18 years or younger. These households were identified for sample from previous months and the following April. For more information about the households eligible for the CPS ASEC, please refer to reference [2].

**Estimation Procedure.** This survey's estimation procedure adjusts weighted sample results to agree with independently derived population estimates of the civilian noninstitutionalized population of the United States and each state (including the District of Columbia). These population estimates, used as controls for the CPS, are prepared monthly to agree with the most current set of population estimates that are released as part of the Census Bureau's population estimates and projections program.

The population controls for the nation are distributed by demographic characteristics in two ways:

- Age, sex, and race (White alone, Black alone, and all other groups combined).
- Age, sex, and Hispanic origin.

The population controls for the states are distributed by race (Black alone and all other race groups combined), age (0-15, 16-44, and 45 and over), and sex.

The independent estimates by age, sex, race, and Hispanic origin, and for states by selected age groups and broad race categories, are developed using the basic demographic accounting formula whereby the population from the 2010 Decennial Census data is updated using data on the

components of population change (births, deaths, and net international migration) with net internal migration as an additional component in the state population estimates.

The net international migration component in the population estimates includes a combination of the following:

- Legal migration to the United States.
- Emigration of foreign-born and native people from the United States.
- Net movement between the United States and Puerto Rico.
- Estimates of temporary migration.
- Estimates of net residual foreign-born population, which include unauthorized migration.

Because the latest available information on these components lags the survey date, it is necessary to make short-term projections of these components to develop the estimate for the survey date.

### **ACCURACY OF THE ESTIMATES**

A sample survey estimate has two types of error: sampling and nonsampling. The accuracy of an estimate depends on both types of error. The nature of the sampling error is known given the survey design; the full extent of the nonsampling error is unknown.

**Sampling Error.** Since the CPS estimates come from a sample, they may differ from figures from an enumeration of the entire population using the same questionnaires, instructions, and enumerators. For a given estimator, the difference between an estimate based on a sample and the estimate that would result if the sample were to include the entire population is known as sampling error. Standard errors, as calculated by methods described in “Standard Errors and Their Use,” are primarily measures of the magnitude of sampling error. However, they may include some nonsampling error.

**Nonsampling Error.** For a given estimator, the difference between the estimate that would result if the sample were to include the entire population and the true population value being estimated is known as nonsampling error. There are several sources of nonsampling error that may occur during the development or execution of the survey. It can occur because of circumstances created by the interviewer, the respondent, the survey instrument, or the way the data are collected and processed. For example, errors could occur because:

- The interviewer records the wrong answer, the respondent provides incorrect information, the respondent estimates the requested information, or an unclear survey question is misunderstood by the respondent (measurement error).
- Some individuals who should have been included in the survey frame were missed (coverage error).
- Responses are not collected from all those in the sample or the respondent is unwilling to provide information (nonresponse error).
- Values are estimated imprecisely for missing data (imputation error).
- Forms may be lost, data may be incorrectly keyed, coded, or recoded, etc. (processing error).

To minimize these errors, the Census Bureau applies quality control procedures during all stages of the production process including the design of the survey, the wording of questions, the review of the work of interviewers and coders, and the statistical review of reports.

Two types of nonsampling error that can be examined to a limited extent are nonresponse and undercoverage.

**Nonresponse.** The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. For the cases eligible for the 2015 ASEC, the basic CPS household-level nonresponse rate was 13.4 percent. The household-level nonresponse rate for the ASEC was an additional 14.2 percent. These two non-response rates lead to a combined supplement nonresponse rate of 25.7 percent.

In accordance with Census Bureau and Office of Management and Budget Quality Standards, the Census Bureau will conduct a nonresponse bias analysis to assess nonresponse bias in the 2015 ASEC.

**Sufficient Partial Interview.** A sufficient partial interview is an incomplete interview in which the household or person answered enough of the questionnaire for the supplement sponsor to consider the interview complete. The remaining supplement questions may have been edited or imputed to fill in missing values. Insufficient partial interviews are considered to be nonrespondents. Refer to the supplement overview attachment in the technical documentation for the specific questions deemed critical by the sponsor as necessary to be answered in order to be considered a sufficient partial interview.

**Undercoverage.** The concept of coverage in the survey sampling process is the extent to which the total population that could be selected for sample “covers” the survey’s target population. Missed housing units and missed people within sample households create undercoverage in the CPS. Overall CPS undercoverage for March 2015 is estimated to be about 13 percent. CPS coverage varies with age, sex, and race. Generally, coverage is larger for females than for males and larger for non-Blacks than for Blacks. This differential coverage is a general problem for most household-based surveys.

The CPS weighting procedure partially corrects for bias from undercoverage, but biases may still be present when people who are missed by the survey differ from those interviewed in ways other than age, race, sex, Hispanic origin, and state of residence. How this weighting procedure affects other variables in the survey is not precisely known. All of these considerations affect comparisons across different surveys or data sources.

A common measure of survey coverage is the coverage ratio, calculated as the estimated population before poststratification divided by the independent population control. Table 2 shows March 2015 CPS coverage ratios by age and sex for certain race and Hispanic groups. The CPS coverage ratios can exhibit some variability from month to month.



Table 2. CPS Coverage Ratios: March 2015

Age group	All people	Total		White only		Black only		Residual race		Hispanic	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-15	0.87	0.88	0.87	0.92	0.91	0.76	0.71	0.79	0.83	0.83	0.84
16-19	0.87	0.89	0.84	0.89	0.88	0.91	0.71	0.87	0.75	0.87	0.90
20-24	0.76	0.75	0.76	0.79	0.80	0.62	0.65	0.69	0.69	0.74	0.81
25-34	0.81	0.79	0.83	0.83	0.87	0.66	0.72	0.73	0.74	0.76	0.82
35-44	0.89	0.87	0.90	0.90	0.94	0.73	0.82	0.77	0.77	0.81	0.86
45-54	0.88	0.86	0.91	0.87	0.93	0.76	0.83	0.83	0.83	0.75	0.87
55-64	0.90	0.90	0.90	0.91	0.92	0.87	0.83	0.82	0.85	0.78	0.78
65+	0.93	0.93	0.94	0.94	0.95	0.82	0.88	0.92	0.81	0.84	0.81
15+	0.87	0.86	0.88	0.88	0.91	0.76	0.79	0.79	0.78	0.79	0.84
0+	0.87	0.86	0.88	0.89	0.91	0.75	0.77	0.79	0.79	0.80	0.84

- Notes: (1) The Residual race group includes cases indicating a single race other than White or Black, and cases indicating two or more races.
- (2) Hispanics may be any race. For a more detailed discussion on the use of parameters for race and ethnicity, please see the “Generalized Variance Parameters” section.

**Comparability of Data.** Data obtained from the CPS and other sources are not entirely comparable. This results from differences in interviewer training and experience and in differing survey processes. This is an example of nonsampling variability not reflected in the standard errors. Therefore, caution should be used when comparing results from different sources.

Data users should be careful when comparing estimates for 2014 in *Income and Poverty in the United States: 2014* and *Health Insurance Coverage in the United States: 2014* (which reflect Census 2010-based controls) with estimates for 1999 to 2010 (from March 2000 CPS to March 2011 CPS), which reflect 2000 census-based controls, and to 1992 to 1998 (from March 1993 CPS to March 1999 CPS), which reflect 1990 census-based controls. Ideally, the same population controls should be used when comparing any estimates. In reality, the use of the same population controls is not practical when comparing trend data over a period of 10 to 20 years. Thus, when it is necessary to combine or compare data based on different controls or different designs, data users should be aware that changes in weighting controls or weighting procedures could create small differences between estimates. See the following discussion for information on comparing estimates derived from different controls or different sample designs.

Data users should be careful when comparing the data from this microdata file, which reflects Census 2010-based controls, with microdata files from January 2003 through December 2011, which reflect 2000 census-based controls. Ideally, the same population controls should be used when comparing any estimates. In reality, the use of the same population controls is not practical when comparing trend data over a period of 10 to 20 years. Thus, when it is necessary to combine or compare data based on different controls or different designs, data users should be aware that changes in weighting controls or weighting procedures can create small differences between estimates. See the discussion following for information on comparing estimates derived from different controls or different sample designs.

Microdata files from previous years reflect the latest available census-based controls. Although the most recent change in population controls had relatively little impact on summary measures such as averages, medians, and percentage distributions, it did have a significant impact on levels. For example, use of Census 2010-based controls results in about a 0.2 percent increase from the 2000 census-based controls in the civilian noninstitutionalized population and in the number of families and households. Thus, estimates of levels for data collected in 2012 and later years will differ from those for earlier years by more than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain population subgroups than for the total population.

Users should also exercise caution because of changes caused by the phase-in of the Census 2010 files (see “Basic CPS”).<sup>4</sup> During this time period, CPS data were collected from sample designs based on different censuses. Three features of the new CPS design have the potential of affecting published estimates: (1) the temporary disruption of the rotation pattern from August 2014 through June 2015 for a comparatively small portion of the sample, (2) the change in sample areas, and (3) the introduction of the new Core-Based Statistical Areas. Most of the known effect on estimates during and after the sample redesign will be the result of changing from 2000 to 2010 geographic definitions. Research has shown that the national-level estimates of the metropolitan and nonmetropolitan populations should not change appreciably because of the new sample design. However, users should still exercise caution when comparing metropolitan and nonmetropolitan estimates across years with a design change, especially at the state level.

Caution should also be used when comparing Hispanic estimates over time. No independent population control totals for people of Hispanic origin were used before 1985.

**A Nonsampling Error Warning.** Since the full extent of the nonsampling error is unknown, one should be particularly careful when interpreting results based on small differences between estimates. The Census Bureau recommends that data users incorporate information about nonsampling errors into their analyses, as nonsampling error could impact the conclusions drawn from the results. Caution should also be used when interpreting results based on a relatively small number of cases. Summary measures (such as medians and percentage distributions) probably do not reveal useful information when computed on a subpopulation smaller than 75,000.

For additional information on nonsampling error including the possible impact on CPS data when known, refer to references [2] and [3].

**Estimation of Median Incomes.** The Census Bureau has changed the methodology for computing median income over time. The Census Bureau has computed medians using either Pareto interpolation or linear interpolation--depending on the size of the income interval--Pareto for intervals larger than \$2,500 in width, linear otherwise. Currently, we are using linear interpolation to estimate all medians. Pareto interpolation assumes a decreasing density of population within an income interval, whereas linear interpolation assumes a constant density of population within an income interval. The Census Bureau calculated estimates of median

---

<sup>4</sup> The phase-in process using the 2010 Census files began in April 2014.

income and associated standard errors for 1979 through 1987 using Pareto interpolation if the estimate was larger than \$20,000 for people or \$40,000 for families and households.

We calculated estimates of median income and associated standard errors for 1976, 1977, and 1978 using Pareto interpolation if the estimate was larger than \$12,000 for people or \$18,000 for families and households. This is because the width of the income interval containing the estimate is greater than \$1,000. All other estimates of median income and associated standard errors for 1976 through 2014 (2015 CPS ASEC) and almost all of the estimates of median income and associated standard errors for 1975 and earlier were calculated using linear interpolation.

Thus, use caution when comparing median incomes above \$12,000 for people or \$18,000 for families and households for different years. Median incomes below those levels are more comparable from year to year since they have always been calculated using linear interpolation. For an indication of the comparability of medians calculated using Pareto interpolation with medians calculated using linear interpolation, see reference [5].

**Standard Errors and Their Use.** The sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a range about a given estimate that has a specified probability of containing the average result of all possible samples. For example, if all possible samples were surveyed under essentially the same general conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

A particular confidence interval may or may not contain the average estimate derived from all possible samples, but one can say with specified confidence that the interval includes the average estimate calculated from all possible samples.

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common type of hypothesis is that the population parameters are different. An example of this would be comparing the percentage of men who were part-time workers to the percentage of women who were part-time workers.

Tests may be performed at various levels of significance. A significance level is the probability of concluding that the characteristics are different when, in fact, they are the same. For example, to conclude that two characteristics are different at the 0.10 level of significance, the absolute value of the estimated difference between characteristics must be greater than or equal to 1.645 times the standard error of the difference.

The Census Bureau uses 90-percent confidence intervals and 0.10 levels of significance to determine statistical validity. Consult standard statistical textbooks for alternative criteria.

**Estimating Standard Errors.** The Census Bureau uses replication methods to estimate the standard errors of CPS estimates. These methods primarily measure the magnitude of sampling error. However, they do measure some effects of nonsampling error as well. They do not measure systematic biases in the data associated with nonsampling error. Bias is the average over all possible samples of the differences between the sample estimates and the true value.

There are two ways to calculate standard errors for the 2015 CPS ASEC microdata file. They are:

- Direct estimates created from replicate weighting methods;
- Generalized variance estimates created from generalized variance function parameters a and b.

While replicate weighting methods provide the most accurate variance estimates, this approach requires more computing resources and more expertise on the part of the user. The Generalized Variance Function (GVF) parameters provide a method of balancing accuracy with resource usage as well as a smoothing effect on standard error estimates across time. For more information on calculating direct estimates, see reference [6]. For more information on generalized variance estimates refer to the “Generalized Variance Parameters” section.

**Generalized Variance Parameters.** While it is possible to compute and present an estimate of the standard error based on the survey data for each estimate in a report, there are a number of reasons why this is not done. A presentation of the individual standard errors would be of limited use, since one could not possibly predict all of the combinations of results that may be of interest to data users. Additionally, data users have access to CPS microdata files, and it is impossible to compute in advance the standard error for every estimate one might obtain from those data sets. Moreover, variance estimates are based on sample data and have variances of their own. Therefore, some methods of stabilizing these estimates of variance, for example, by generalizing or averaging over time, may be used to improve their reliability.

Experience has shown that certain groups of estimates have similar relationships between their variances and expected values. Modeling or generalizing may provide more stable variance estimates by taking advantage of these similarities. The generalized variance function is a simple model that expresses the variance as a function of the expected value of the survey estimate. The parameters of the generalized variance function are estimated using direct replicate variances. These generalized variance parameters provide a relatively easy method to obtain approximate standard errors for numerous characteristics.

The generalized variance parameters to use in computing standard errors are dependent upon the race/ethnicity group of interest. Table 3 summarizes the relationship between the race/ethnicity group of interest and the generalized variance parameters to use in standard error calculations.

In this source and accuracy statement, Table 4 provides the generalized variance parameters for labor force estimates, and Table 5 provides generalized variance parameters for characteristics from the 2015 CPS ASEC supplement. Also, tables are provided that allow the calculation of parameters for prior years and parameters for states and regions. Tables 6 and 7 contain

correlation coefficients for comparing estimates from consecutive years. Tables 8 and 9 provide factors and population controls to derive state and regional parameters.

The basic CPS questionnaire records the race and ethnicity of each respondent. With respect to race, a respondent can be White, Black, Asian, American Indian and Alaskan Native (AIAN), Native Hawaiian and Other Pacific Islander (NHOPI), or combinations of two or more of the preceding. A respondent's ethnicity can be Hispanic or non-Hispanic, regardless of race.

**Table 3. Estimation Groups of Interest and Generalized Variance Parameters**

<b>Race/ethnicity group of interest</b>	<b>Generalized variance parameters to use in standard error calculations</b>
Total population	Total or White
White alone, White AOIC, or White non-Hispanic population	Total or White
Black alone, Black AOIC, or Black non-Hispanic population	Black
Asian alone, Asian AOIC, or Asian non-Hispanic population	Asian, AIAN, NHOPI
AIAN alone, AIAN AOIC, or AIAN non-Hispanic population	
NHOPI alone, NHOPI AOIC, or NHOPI non-Hispanic population	
Populations from other race groups	Asian, AIAN, NHOPI
Hispanic population	Hispanic
Two or more races – employment/unemployment and educational attainment characteristics	Black
Two or more races – all other characteristics	Asian, AIAN, NHOPI

- Notes: (1) AIAN is American Indian and Alaska Native and NHOPI is Native Hawaiian and Other Pacific Islander.  
 (2) AOIC is an abbreviation for alone or in combination. The AOIC population for a race group of interest includes people reporting only the race group of interest (alone) and people reporting multiple race categories including the race group of interest (in combination).  
 (3) Hispanics may be any race.  
 (4) Two or more races refers to the group of cases self-classified as having two or more races.

**Standard Errors of Estimated Numbers.** The approximate standard error,  $s_x$ , of an estimated number from this microdata file can be obtained by using the formula:

$$s_x = \sqrt{ax^2 + bx} \quad (1)$$

Here  $x$  is the size of the estimate and  $a$  and  $b$  are the parameters in Table 4 or 5 associated with the particular type of characteristic. When calculating standard errors from cross-tabulations involving different characteristics, use the set of parameters for the characteristic that will give the largest standard error.

Illustration 1

Suppose there were 3,767,000 unemployed females in the civilian labor force. Use Formula (1) and the appropriate parameters from Table 4 to get

<b>Illustration 1</b>	
Number of unemployed females in the civilian labor force ( $x$ )	3,767,000
a parameter ( $a$ )	-0.000028
b parameter ( $b$ )	2,788
Standard error	101,000
90-percent confidence interval	3,601,000 to 3,933,000

The standard error is calculated as

$$s_x = \sqrt{-0.000028 \times 3,767,000^2 + 2,788 \times 3,767,000} = 101,000$$

and the 90-percent confidence interval is calculated as  $3,767,000 \pm 1.645 \times 101,000$ .

A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

Illustration 2

Suppose there were 60,010,000 married-couple family households. Use Formula (1) and the appropriate parameters from Table 5 to get

<b>Illustration 2</b>	
Number of married-couple family households ( $x$ )	60,010,000
a parameter ( $a$ )	-0.000005
b parameter ( $b$ )	1,285
Standard error	243,000
90-percent confidence interval	59,610,000 to 60,410,000

The standard error is calculated as

$$s_x = \sqrt{-0.000005 \times 60,010,000^2 + 1,285 \times 60,010,000} = 243,000$$

and the 90-percent confidence interval is calculated as  $60,010,000 \pm 1.645 \times 243,000$ .

A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

**Standard Errors of Estimated Percentages.** The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on both the size of

the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories, use the parameter from Table 4 or 5 as indicated by the numerator.

The approximate standard error,  $s_{y,p}$ , of an estimated percentage can be obtained by using the formula:

$$s_{y,p} = \sqrt{\frac{b}{y} p(100-p)} \quad (2)$$

Here  $y$  is the total number of people, families, households, or unrelated individuals in the base or denominator of the percentage,  $p$  is the percentage  $100 \cdot x/y$  ( $0 \leq p \leq 100$ ), and  $b$  is the parameter in Table 4 or 5 associated with the characteristic in the numerator of the percentage.

### Illustration 3

Suppose there were 212,589,000 out of 242,248,000 adults (aged 18 and older), or 87.8 percent, who graduated from high school. Use Formula (2) and the appropriate parameter from Table 5 to get

<b>Illustration 3</b>	
Percentage of adults who are high school graduates ( $p$ )	87.8
Base ( $y$ )	242,248,000
b parameter ( $b$ )	1,473
Standard error	0.08
90-percent confidence interval	87.7 to 87.9

The standard error is calculated as

$$s_{y,p} = \sqrt{\frac{1,473}{242,248,000} \times 87.8 \times (100 - 87.8)} = 0.08$$

The 90-percent confidence interval of the percentage of adults who graduated from high school is calculated as  $87.8 \pm 1.645 \times 0.08$ .

**Standard Errors of Estimated Differences.** The standard error of the difference between two sample estimates is approximately equal to

$$s_{x_1 - x_2} = \sqrt{s_{x_1}^2 + s_{x_2}^2 - 2rs_{x_1}s_{x_2}} \quad (3)$$

where  $s_{x_1}$  and  $s_{x_2}$  are the standard errors of the estimates,  $x_1$  and  $x_2$ . The estimates can be numbers, percentages, ratios, etc. Tables 7 and 8 contain the correlation coefficient,  $r$ , for CPS year-to-year comparisons. The correlations were derived for income, poverty, and health insurance estimates, but they can be used for other types of estimates where the year-to-year

correlation between identical households is high. For making other comparisons, assume that  $r$  equals zero. Making this assumption will result in accurate estimates of standard errors for the difference between two estimates of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. However, if there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

#### Illustration 4

Suppose there were 24,514,000 men over age 24 who were never married and 10,908,000 men over age 24 who were divorced. The apparent difference is 13,606,000. Use Formulas (1) and (3) with  $r = 0$  and the appropriate parameters from Table 5 to get

<b>Illustration 4</b>			
	Never married ( $x_1$ )	Divorced ( $x_2$ )	Difference
Number of males over age 24	24,514,000	10,908,000	13,606,000
a parameter ( $a$ )	-0.000010	-0.000010	-
b parameter ( $b$ )	3,240	3,240	-
Standard error	271,000	185,000	328,000
90-percent confidence interval	24,068,000 to 24,960,000	10,604,000 to 11,212,000	13,066,000 to 14,146,000

The standard error of the difference is calculated as

$$s_{x_1-x_2} = \sqrt{271,000^2 + 185,000^2} = 328,000$$

The 90-percent confidence interval around the difference is calculated as  $13,606,000 \pm 1.645 \times 328,000$ . Since this interval does not include zero, we can conclude with 90 percent confidence that the number of never married men over age 24 was higher than the number of divorced men over age 24.

#### Illustration 5

Suppose that the percentage of children in poverty in 2014 was 21.1 percent out of 73,556,000 children, and the percentage of children in poverty in 2013 was 21.5 percent out of 73,439,000 people. The apparent difference is 0.7 percent. Use Formulas (2) and (3) and the appropriate parameter and correlation coefficient from Tables 5 and 6 to get



Illustration 5			
	2013 ( $x_1$ )	2014 ( $x_2$ )	Difference
Percentage of children in poverty ( $p$ )	21.5	21.1	0.4
Base	73,439,000	73,556,000	-
b parameter ( $b$ )	10,859	4,974	-
Correlation coefficient ( $r$ )	-	-	0.45
Standard error	0.50	0.34	0.46
90-percent confidence interval	20.7 to 22.3	20.5 to 21.7	-0.4 to 1.2

The standard error of the difference is calculated as

$$s_{x_1-x_2} = \sqrt{0.50^2 + 0.34^2 - 2 \times 0.45 \times 0.50 \times 0.34} = 0.46$$

and the 90-percent confidence interval around the difference is calculated as  $0.4 \pm 1.645 \times 0.46$ . Since this interval includes zero, we can conclude with 90 percent confidence that the percentage of children in poverty in 2013 is not statistically different than the percentage of children in poverty in 2014.

**Standard Errors of Estimated Ratios.** Certain estimates may be calculated as the ratio of two numbers. Compute the standard error of a ratio,  $x/y$ , using

$$s_{x/y} = \frac{x}{y} \sqrt{\left(\frac{s_x}{x}\right)^2 + \left(\frac{s_y}{y}\right)^2 - 2r \frac{s_x s_y}{xy}} \quad (4)$$

The standard error of the numerator,  $s_x$ , and that of the denominator,  $s_y$ , may be calculated using formulas described earlier. In Formula (4),  $r$  represents the correlation between the numerator and the denominator of the estimate.

For one type of ratio, the denominator is a count of families or households and the numerator is a count of people in those families or households with a certain characteristic. If there is at least one person with the characteristic in every family or household, use 0.7 as an estimate of  $r$ . An example of this type is the average number of children per family with children.

For all other types of ratios,  $r$  is assumed to be zero. Examples are the average number of children per family and the family poverty rate. If  $r$  is actually positive (negative), then this procedure will provide an overestimate (underestimate) of the standard error of the ratio.

Note: For estimates expressed as the ratio of  $x$  per 100  $y$  or  $x$  per 1,000  $y$ , multiply Formula (4) by 100 or 1,000, respectively, to obtain the standard error.

**Illustration 6**

Suppose there were 9,907,000 males working part-time and 17,428,000 females working part-time. The ratio of males working part-time to females working part-time would be 0.568, or 56.8 percent. Use Formulas (1) and (4) with  $r = 0$  and the appropriate parameters from Table 4 to get

<b>Illustration 6</b>			
	Males (x)	Females (y)	Ratio
Number who work part-time	9,907,000	17,428,000	0.568
a parameter (a)	-0.000031	-0.000028	-
b parameter (b)	2,947	2,788	-
Standard error	162,000	200,000	0.011
90-percent confidence interval	9,641,000 to 10,173,000	17,099,000 to 17,757,000	0.549 to 0.587

The standard error is calculated as

$$s_{x/y} = \frac{9,907,000}{17,428,000} \sqrt{\left(\frac{162,000}{9,907,000}\right)^2 + \left(\frac{200,000}{17,428,000}\right)^2} = 0.011$$

and the 90-percent confidence interval is calculated as  $0.568 \pm 1.645 \times 0.011$ .

**Illustration 7**

Suppose that the number of families below the poverty level was 9,467,000 and the total number of families was 81,730,000. The ratio of families below the poverty level to the total number of families would be 0.116 or 11.6 percent. Use the appropriate parameters from Table 5 and Formulas (1) and (4) with  $r = 0$  to get

<b>Illustration 7</b>			
	In poverty (x)	Total (y)	Ratio (in percent)
Number of families	9,467,000	81,730,000	11.6
a parameter (a)	0.000052	-0.000005	-
b parameter (b)	1,518	1,285	-
Standard error	138,000	268,000	0.17
90-percent confidence interval	9,240,000 to 9,694,000	81,289,000 to 82,171,000	11.3 to 11.9

The standard error is calculated as

$$s_{x/y} = \frac{9,467,000}{81,730,000} \sqrt{\left(\frac{138,000}{9,467,000}\right)^2 + \left(\frac{268,000}{81,730,000}\right)^2} = 0.0017 = 0.17\%$$

and the 90-percent confidence interval of the percentage is calculated as  $11.6 \pm 1.645 \times 0.17$ .

**Standard Errors of Estimated Medians.** The sampling variability of an estimated median depends on the form of the distribution and the size of the base. One can approximate the reliability of an estimated median by determining a confidence interval about it. (See “Standard Errors and Their Use” for a general discussion of confidence intervals.)

Estimate the 68-percent confidence limits of a median based on sample data using the following procedure:

1. Using Formula (2) and the base of the distribution, calculate the standard error of 50 percent.
2. Add to and subtract from 50 percent the standard error determined in step 1. These two numbers are the percentage limits corresponding to the 68-percent confidence interval about the estimated median.
3. Using the distribution of the characteristic, determine upper and lower limits of the 68-percent confidence interval by calculating values corresponding to the two points established in step 2.

Note: The percentage limits found in step 2 may or may not fall in the same characteristic distribution interval.

Use the following formula to calculate the upper and lower limits:

$$X_p = \frac{pN - N_1}{N_2 - N_1}(A_2 - A_1) + A_1 \quad (5)$$

where

$X_p$  = estimated upper and lower bounds for the confidence interval ( $0 \leq p \leq 1$ ). For purposes of calculating the confidence interval,  $p$  takes on the values determined in step 2. Note that  $X_p$  estimates the median when  $p = 0.50$ .

$N$  = for distribution of numbers: the total number of units (people, households, etc.) for the characteristic in the distribution.

= for distribution of percentages: the value 100.

$p$  = the values obtained in Step 2.

$A_1, A_2$  = the lower and upper bounds, respectively, of the interval containing  $X_p$ .

$N_1, N_2$  = for distribution of numbers: the estimated number of units

(people, households, etc.) with values of the characteristic less than or equal to  $A_1$  and  $A_2$ , respectively.

= for distribution of percentages: the estimated percentage of units (people, households, etc.) having values of the characteristic less than or equal to  $A_1$  and  $A_2$ , respectively.

4. Divide the difference between the two points determined in step 3 by 2 to obtain the standard error of the median.

Note: Median incomes and their standard errors calculated as below may differ from those in published tables and reports showing income, since narrower income intervals were used in those calculations.

### Illustration 8

Suppose there were 124,587,000 households in 2015, and their income was distributed in the following way:

<b>Illustration 8</b>			
Income level	Number of households	Cumulative number of households	Cumulative percent of households
Under \$5,000	4,572,000	4,572,000	3.67%
\$5,000 to \$9,999	4,320,000	8,892,000	7.14%
\$10,000 to \$14,999	6,766,000	15,658,000	12.57%
\$15,000 to \$24,999	13,644,000	29,302,000	23.52%
\$25,000 to \$34,999	12,595,000	41,897,000	33.63%
\$35,000 to \$49,999	16,346,000	58,243,000	46.75%
\$50,000 to \$74,999	21,228,000	79,471,000	63.79%
\$75,000 to \$99,999	14,380,000	93,851,000	75.33%
\$100,000 and over	30,736,000	124,587,000	100.00%

1. Using Formula (2) with  $b = 1,393$ , the standard error of 50 percent on a base of 124,587,000 is about 0.17 percent.
2. To obtain a 68-percent confidence interval on an estimated median, add to and subtract from 50 percent the standard error found in step 1. This yields percentage limits of 49.83 and 50.17.
3. The lower and upper limits for the interval in which the percentage limits falls are \$50,000 and \$75,000, respectively.

Then the estimated numbers of households with an income less than or equal to \$50,000 and \$75,000 are 58,243,000 and 79,471,000, respectively.

Using Formula (5), the lower limit for the confidence interval of the median is found to be about

$$X_{0.4985} = \frac{0.4983 \times 124,587,000 - 58,243,000}{79,471,000 - 58,243,000} (75,000 - 50,000) + 50,000 = 54,521$$

Similarly, the upper limit is found to be about

$$X_{0.5015} = \frac{0.5017 \times 124,587,000 - 58,243,000}{79,471,000 - 58,243,000} (75,000 - 50,000) + 50,000 = 55,020$$

Thus, a 68-percent confidence interval for the median income for households is from \$54,521 to \$55,020.

4. The standard error of the median is, therefore,

$$\frac{55,020 - 54,521}{2} = 249.5$$

**Standard Errors of Averages for Grouped Data.** The formula used to estimate the standard error of an average for grouped data is

$$s_{\bar{x}} = \sqrt{\frac{b}{y}(S^2)} \quad (6)$$

In this formula,  $y$  is the size of the base of the distribution and  $b$  is the parameter from Table 4 or 5. The variance,  $S^2$ , is given by the following formula:

$$S^2 = \sum_{i=1}^c p_i \bar{x}_i^2 - \bar{x}^2 \quad (7)$$

where  $\bar{x}$ , the average of the distribution, is estimated by

$$\bar{x} = \sum_{i=1}^c p_i \bar{x}_i \quad (8)$$

where

- $c$  = the number of groups;  $i$  indicates a specific group, thus taking on values 1 through  $c$ .
- $p_i$  = estimated proportion of households, families, or people whose values for the characteristic being considered fall in group  $i$ .

$\bar{x}_i = (Z_{Li} + Z_{Ui})/2$  where  $Z_{Li}$  and  $Z_{Ui}$  are the lower and upper interval boundaries, respectively, for group  $i$ .  $\bar{x}_i$  is assumed to be the most representative value for the characteristic of households, families, or people in group  $i$ . If group  $c$  is open-ended, i.e., no upper interval boundary exists, use a group approximate average value of

$$\bar{x}_c = \frac{3}{2} Z_{Lc} \quad (9)$$

### Illustration 9

Suppose that there were 9,467,000 families in poverty and that the distribution of the income deficit (the difference between their family income and poverty threshold) for all families in poverty was

Income deficit	Number of families in poverty	Percentage of families in poverty ( $p_i$ )	Average income deficit ( $\bar{x}_i$ )
Under \$1000	593,000	6.3	500
\$1000 to \$2,499	907,000	9.6	1,750
\$2,500 to \$4,999	1,382,000	14.6	3,750
\$5,000 to \$7,499	1,333,000	14.1	6,250
\$7,500 to \$9,999	957,000	10.1	8,750
\$10,000 to \$12,499	902,000	9.5	11,250
\$12,500 to \$14,999	855,000	9.0	13,750
\$15,000 and over	2,537,000	26.8	22,500
Total	9,466,000*	100.0	

\*There may be a difference due to rounding

Using Formula (8),

$$\bar{x} = (0.063 \times 500) + (0.096 \times 1,750) + (0.146 \times 3,750) + (0.141 \times 6,250) + (0.101 \times 8,750) + (0.095 \times 11,250) + (0.090 \times 13,750) + (0.268 \times 22,500) = 10,848$$

and Formula (7),

$$S^2 = (0.063 \times 500^2) + (0.096 \times 1,750^2) + (0.146 \times 3,750^2) + (0.141 \times 6,250^2) + (0.101 \times 8,750^2) + (0.095 \times 11,250^2) + (0.090 \times 13,750^2) + (0.268 \times 22,500^2) - 10,848^2 = 62,638,000$$

Use the appropriate parameter from Table 5 and Formula (6) to get

<b>Illustration 9</b>	
Average income deficit for families in poverty ( $\bar{x}$ )	\$10,848
Variance ( $S^2$ )	62,638,000
Base ( $v$ )	9,467,000
b parameter ( $b$ )	1,518
Standard error	\$100
90-percent confidence interval	\$10,684 to \$11,013

The standard error is calculated as

$$s_{\bar{x}} = \sqrt{\frac{1,518}{9,467,000}(62,638,000)} = 100$$

and the 90-percent confidence interval is calculated as  $\$10,848 \pm 1.645 \times \$100$ .

**Standard Errors of Estimated Per Capita Deficits.** Certain average values in reports associated with the CPS ASEC data represent the per capita deficit for households of a certain class. The average per capita deficit is approximately equal to

$$x = \frac{hm}{p} \quad (10)$$

where

$h$  = number of households in the class.

$m$  = average deficit for households in the class.

$p$  = number of people in households in the class.

$x$  = average per capita deficit of people in households in the class.

To approximate standard errors for these averages, use the formula

$$s_x = \frac{hm}{p} \sqrt{\left(\frac{s_m}{m}\right)^2 + \left(\frac{s_p}{p}\right)^2 + \left(\frac{s_h}{h}\right)^2 - 2r\left(\frac{s_p}{p}\right)\left(\frac{s_h}{h}\right)} \quad (11)$$

In Formula (11),  $r$  represents the correlation between  $p$  and  $h$ .

For one type of average, the class represents households containing a fixed number of people. For example,  $h$  could be the number of 3-person households. In this case, there is an exact

correlation between the number of people in households and the number of households. Therefore,  $r = 1$  for such households. For other types of averages, the class represents households of other demographic types, for example, households in distinct regions, households in which the householder is of a certain age group, and owner-occupied and tenant-occupied households. In this and other cases in which the correlation between  $p$  and  $h$  is not perfect, use 0.7 as an estimate of  $r$ .

#### Illustration 10

Suppose there were 32,615,000 people living in families in poverty, and 9,467,000 families in poverty, with an average deficit income for families in poverty of \$10,848 with a standard error of \$100 (from Illustration 9). Use Formulas (1), (10), and (11) and the appropriate parameters from Table 5 and  $r = 0.7$  to get

Illustration 10				
	Number ( $h$ )	Number of people ( $p$ )	Average income deficit ( $m$ )	Average per capita deficit ( $x$ )
Value for families in poverty	9,467,000	32,615,000	\$10,848	\$3,149
a parameter ( $a$ )	+0.000052	-0.000020	-	-
b parameter ( $b$ )	1,518	6,452	-	-
Correlation ( $r$ )	-	-	-	0.7
Standard Error	138,000	435,000	\$100	\$45
90-percent confidence interval	9,240,000 to 9,694,000	31,899,000 to 33,331,000	\$10,684 to \$11,013	\$3,075 to \$3,223

The estimate of the average per capita deficit is calculated as

$$x = \frac{9,467,000 \times 10,848}{32,615,000} = 3,149$$

and the standard error is calculated as

$$s_x = \frac{9,467,000 \times 10,848}{32,615,000} \sqrt{\left(\frac{100}{10,848}\right)^2 + \left(\frac{435,000}{32,615,000}\right)^2 + \left(\frac{138,000}{9,467,000}\right)^2 - 2 \times 0.7 \times \left(\frac{435,000}{32,615,000}\right) \times \left(\frac{138,000}{9,467,000}\right)} = 45$$

The 90-percent confidence interval is calculated as  $\$3,149 \pm 1.645 \times \$45$ .

**Accuracy of State Estimates.** The redesign of the CPS following the 1980 census provided an opportunity to increase efficiency and accuracy of state data. All strata are now defined within state boundaries. The sample is allocated among the states to produce state and national estimates with the required accuracy while keeping total sample size to a minimum. Improved accuracy of state data was achieved with about the same sample size as in the 1970 design.

Since the CPS is designed to produce both state and national estimates, the proportion of the total population sampled and the sampling rates differ among the states. In general, the smaller the



population of the state the larger the sampling proportion. For example, in Vermont approximately 1 in every 250 households is sampled each month. In New York the sample is about 1 in every 2,000 households. Nevertheless, the size of the sample in New York is four times larger than in Vermont because New York has a larger population.

Note: The Census Bureau recommends the use of 3-year averages to compare estimates across states and 2-year averages to evaluate changes in state estimates over time. See “Standard Errors of Data for Combined Years” and “Standard Errors of Differences of 2-Year Averages.” The Census Bureau also recommends the American Community Survey microdata file as the preferred source for income and poverty state data in years 2006 (2005 estimates) to the present.

**Standard Errors for State Estimates.** The standard error for a state may be obtained by determining new state-level  $a$  and  $b$  parameters and then using these adjusted parameters in the standard error formulas mentioned previously. To determine a new state-level  $b$  parameter ( $b_{state}$ ), multiply the  $b$  parameter from Table 4 or 5 by the state factor from Table 9. To determine a new state-level  $a$  parameter ( $a_{state}$ ), use the following:

- (1) If the  $a$  parameter from Table 4 or 5 is positive, multiply it by the state factor from Table 9.
- (2) If the  $a$  parameter in Table 4 or 5 is negative, calculate the new state-level  $a$  parameter as follows:

$$a_{state} = \frac{-b_{state}}{POP_{state}} \quad (12)$$

where  $POP_{state}$  is the state population found in Table 9.

#### Illustration 11

Suppose there were 15,268,000 people living in New York state who were born in the United States. Use Formulas (1) and (12) and the appropriate parameter, factor, and population from Tables 5 and 8 to get

<b>Illustration 11</b>	
Number of people in NY who were born in the U.S. ( $x$ )	15,268,000
$b$ parameter ( $b$ )	3,240
New York state factor	1.19
State population	19,550,311
State $a$ parameter ( $a_{state}$ )	-0.000197
State $b$ parameter ( $b_{state}$ )	3,856
Standard error	114,000

Obtain the state-level  $b$  parameter by multiplying the  $b$  parameter, 3,240, by the state factor, 1.09. This gives  $b_{state} = 3,240 \times 1.19 = 3,856$ . Obtain the needed state-level  $a$  parameter by

$$a_{state} = \frac{-3,856}{19,550,311} = -0.000197$$

The standard error of the estimate of the number of people in New York state who were born in the United States can then be found by using Formula (1) and the new state-level  $a$  and  $b$  parameters, -0.000197 and 3,856, respectively. The standard error is given by

$$s_x = \sqrt{-0.000197 \times 15,268,000^2 + 3,856 \times 15,268,000} = 114,000$$

**Standard Errors of Regional Estimates.** To compute standard errors for regional estimates, follow the steps for computing standard errors for state estimates found in “Standard Errors for State Estimates” using the regional factors and populations found in Table 9.

Illustration 12

Suppose there were 19,531,000 of 118,253,000 people, or 16.5 percent, living in poverty in the South. Use Formulas (2) and (12) and the appropriate parameter, factor, and population from Tables 5 and 9 to get

<b>Illustration 12</b>	
Poverty rate in the South ( $p$ )	16.5
Base ( $y$ )	118,253,000
$b$ parameter ( $b$ )	6,452
South regional factor	1.13
Regional $b$ parameter ( $b_{region}$ )	7,291
Standard error	0.29
90-percent confidence interval	16.0 to 17.0

Obtain the region-level  $b$  parameter by multiplying the  $b$  parameter, 6,452, by the South regional factor, 1.06. This gives  $b_{region} = 6,452 \times 1.13 = 7,291$ .

The standard error of the estimate of the poverty rate for people living in the South can then be found by using Formula (2) and the new region-level  $b$  parameter, 7,291. The standard error is given by

$$s_{y,p} = \sqrt{\frac{7,291}{118,253,000} \times 16.5 \times (100 - 16.5)} = 0.29$$

and the 90-percent confidence interval of the poverty rate for people living in the South is calculated as  $16.5 \pm 1.645 \times 0.29$ .

**Standard Errors of Groups of States.** The standard error calculation for a group of states is similar to the standard error calculation for a single state. First, calculate a new state group factor for the group of states. Then, determine new state group  $a$  and  $b$  parameters. Finally, use these adjusted parameters in the standard error formulas mentioned previously.

Use the following formula to determine a new state group factor:

$$\text{state group factor} = \frac{\sum_{i=1}^n \text{POP}_i \times \text{state factor}_i}{\sum_{i=1}^n \text{POP}_i} \quad (13)$$

where  $\text{POP}_i$  and  $\text{state factor}_i$  are the population and factor for state  $i$  from Table 8. To obtain a new state group  $b$  parameter ( $b_{\text{state group}}$ ), multiply the  $b$  parameter from Table 4 or 5 by the state factor obtained by Formula (13). To determine a new state group  $a$  parameter ( $a_{\text{state group}}$ ), use the following:

- (1) If the  $a$  parameter from Table 4 or 5 is positive, multiply it by the state group factor determined by Formula (13).
- (2) If the  $a$  parameter in Table 4 or 5 is negative, calculate the new state group  $a$  parameter as follows:

$$a_{\text{state group}} = \frac{-b_{\text{state group}}}{\sum_{i=1}^n \text{POP}_i} \quad (14)$$

### Illustration 13

Suppose the state group factor for the state group Illinois-Indiana-Michigan was required. The appropriate factor would be

$$\text{state group factor} = \frac{12,695,960 \times 1.17 + 6,517,923 \times 1.11 + 9,810,823 \times 1.11}{12,695,960 + 6,517,923 + 9,810,823} = 1.14$$

**Standard Errors of Data for Combined Years.** Sometimes estimates for multiple years are combined to improve precision. For example, suppose  $\bar{x}$  is an average derived from  $n$

consecutive years' data, i.e.,  $\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$ , where the  $x_i$  are the estimates for the individual years.

Use the formulas described previously to estimate the standard error,  $s_{x_i}$ , of each year's estimate.

Then the standard error of  $\bar{x}$  is

$$s_{\bar{x}} = \frac{s_x}{n} \quad (15)$$

where

$$s_x = \sqrt{\sum_{i=1}^n s_{x_i}^2 + 2r \sum_{i=1}^{n-1} s_{x_i} s_{x_{i+1}}} \quad (16)$$

and  $s_{x_i}$  are the standard errors of the estimates  $x_i$ . Tables 6 and 7 contain the correlation coefficients,  $r$ , for the correlation between consecutive years  $i$  and  $i+1$ . Correlation between nonconsecutive years is zero. The correlations were derived for income and poverty estimates, but they can be used for other types of estimates where the year-to-year correlation between identical households is high.

The Census Bureau recommends the use of 3-year average estimates for certain small population subgroups<sup>5</sup> (see also “Accuracy of State Estimates.”) Two-year moving averages are recommended for these small population subgroups for comparisons across adjacent years (see “Standard Errors of Differences of 2-Year Averages.”)

#### Illustration 14

Suppose the 2012-2014 3-year average percentage of families with female householder, no husband present (FFH), in poverty was 30.7. Suppose the percentages and bases for 2012, 2013, and 2014 were 30.9, 32.2, and 30.6 percent and 15,489,000, 16,176,000 and 15,553,000 respectively. Use the appropriate parameters and correlation coefficients from Tables 5 and 6 and Formulas (2), (15), and (16) to get

Illustration 14				
	2012	2013	2014	2012-2014 avg
Percentage of families with female householder, no husband present (FFH), in poverty ( $p$ )	30.9	32.2	30.6	31.2
Base ( $y$ )	15,489,000	16,176,000	15,553,000	-
b parameter ( $b$ )	1,243	3,315	1,518	-
Correlation ( $r$ )	-	-	-	0.35, 0.35
Standard error	0.41	0.67	0.46	0.37
90-percent confidence interval	30.2 to 31.6	31.1 to 33.3	29.8 to 31.4	30.6 to 31.8

The standard error of the 3-year average is calculated as

$$s_{\bar{x}} = \frac{1.11}{3} = 0.37$$

where

$$s_x = \sqrt{0.41^2 + 0.67^2 + 0.46^2 + (2 \times 0.35 \times 0.41 \times 0.67) + (2 \times 0.35 \times 0.67 \times 0.46)} = 1.11$$

<sup>5</sup> Estimates of characteristics of the American Indian and Alaska Native (AIAN) and Native Hawaiian and Other Pacific Islander (NHOPI) populations based on a single-year sample would be unreliable due to the small size of the sample that can be drawn from either population. Accordingly, such estimates are based on multiyear averages.

The 90-percent confidence interval for the 3-year average percentage of the AIAN population without health insurance is  $31.2 \pm 1.645 \times 0.37$ .

**Standard Errors of Differences of 2-Year Averages.** Comparing two non-overlapping 2-year averages also improves precision for comparisons across years. Use the formulas described previously to estimate the standard error,  $s_{x_i}$ , of each year's estimate,  $x_i$ , and the standard error,  $s_{\bar{x}_{i,i+1}}$ , of each average,  $\bar{x}_{i,i+1}$ . Then the standard error of the difference of the two non-overlapping 2-year averages,  $\bar{x}_{1,2} - \bar{x}_{3,4}$ , is

$$s_{\bar{x}_{1,2} - \bar{x}_{3,4}} = \sqrt{s_{\bar{x}_{1,2}}^2 + s_{\bar{x}_{3,4}}^2 - \frac{1}{2} r s_{x_2} s_{x_3}} \quad (17)$$

### Illustration 15

Suppose that you want to calculate the standard error of the difference between the 2011, 2012 and 2013, 2014 averages of the percent of families with female householder, no husband present (FFH), in poverty. Use the following information along with Tables 5 and 8 and Formula (2) to get

	2011	2012	2013	2014
Percentage of families with female householder, no husband present (FFH), in poverty ( $p$ )	31.2	30.9	32.2	30.6
Base ( $y$ )	15,678,000	15,489,000	16,176,000	15,553,000
b parameter ( $b$ )	1,243	1,243	3,315	1,518
Standard error	0.41	0.41	0.67	0.46

Use this information, Formulas (15), (16), and (17), and the appropriate correlation coefficient from Table 6 to get

Illustration 15				
	2011, 2012	2012, 2013	2013, 2014	avg(2013, 2014) - avg(2011, 2012)
Average percent of families with female householder, no husband present (FFH), in poverty ( $\bar{x}$ )	31.1	-	31.4	-0.3
Correlation coefficient ( $r$ )	0.35	0.35	0.35	-
Standard error	0.34*	-	0.47*	0.53
90-percent confidence interval	30.5 to 31.7	-	30.6 to 32.2	-0.6 to 1.2

\*See "Standard Errors of Data for Combined Years" for instructions and illustrations on calculating these standard errors.

The standard error of the difference of the two 2-year averages is calculated as

$$s_{x_{1,2}-x_{3,4}} = \sqrt{0.34^2 + 0.47^2 - \frac{1}{2} \times 0.35 \times 0.41 \times 0.67} = 0.53$$

and the 90-percent confidence interval around the difference of the 2-year averages is calculated as  $-0.3 \pm 1.645 \times 0.53$ . Since this interval includes zero, we cannot conclude with 90 percent confidence that the 2013-2014 average percent of families with female householder, no husband present (FFH), in poverty was lower than the 2011-2012 average percent of families with female householder, no husband present (FFH), in poverty.

**Standard Errors of Quarterly or Yearly Averages.** For information on calculating standard errors for labor force data from the CPS which involve quarterly or yearly averages, please see the “Explanatory Notes and Estimates of Error: Household Data” section in *Employment and Earnings*, a monthly report published by the U.S. Bureau of Labor Statistics.

**Technical Assistance.** If you require assistance or additional information, please contact the Demographic Statistical Methods Division via e-mail at [dsmd.source.and.accuracy@census.gov](mailto:dsmd.source.and.accuracy@census.gov).

**Table 4. Parameters for Computation of Standard Errors for Labor Force Characteristics:  
March 2015**

Characteristic	A	b
<b>Total or White</b>		
<i>Civilian labor force, employed</i>	-0.000013	2,481
<i>Not in labor force</i>	-0.000013	2,432
<i>Unemployed</i>	-0.000017	3,244
<i>Civilian labor force, employed, not in labor force, and unemployed</i>		
Men	-0.000031	2,947
Women	-0.000028	2,788
Both sexes, 16 to 19 years	-0.000261	3,244
<b>Black</b>		
<i>Civilian labor force, employed, not in labor force, and unemployed</i>	-0.000117	3,601
Men	-0.000249	3,465
Women	-0.000190	3,191
Both sexes, 16 to 19 years	-0.001425	3,601
<b>Hispanic, may be of any race</b>		
<i>Civilian labor force, employed, not in labor force, and unemployed</i>	-0.000245	3,311
Men	-0.000537	3,397
Women	-0.000399	2,874
Both sexes, 16 to 19 years	-0.004078	3,311
<b>Asian, American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander</b>		
<i>Civilian labor force, employed, not in labor force, and unemployed</i>	-0.000087	3,316
Men	-0.000172	3,276
Women	-0.000158	3,001
Both sexes, 16 to 19 years	-0.000909	3,316

- NOTES: (1) These parameters are to be applied to basic CPS monthly labor force estimates.
- (2) The Total or White, Black, and Asian, AIAN, NHUPI parameters are to be used for both alone and in combination race group estimates.
- (3) For nonmetropolitan characteristics, multiply the a and b parameters by 1.5. If the characteristic of interest is total state population, not subtotaled by race or ethnicity, the a and b parameters are zero.
- (4) For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Black, Hispanic, and Asian, AIAN, NHUPI parameters.
- (5) For the groups self-classified as having two or more races, use the Asian, AIAN, NHUPI parameters for all employment characteristics.

Table 5. Parameters for Computation of Standard Errors for People and Families: 2015 CPS ASEC

Characteristics	Total or White		Black		Asian, AIAN, & NHOPI		Hispanic	
	a	b	a	b	a	b	a	b
<b>PEOPLE</b>								
Educational attainment	-0.000006	1,473	-0.000031	1,666	-0.000060	1,345	-0.000029	1,126
Employment	-0.000016	3,748	-0.000151	4,220	-0.000346	3,907	-0.000141	4,220
People by family income	-0.000012	3,047	-0.000065	3,488	-0.000156	3,488	-0.000089	3,488
Income characteristics								
Total	-0.000006	1,526	-0.000033	1,747	-0.000078	1,747	-0.000044	1,747
Male	-0.000013	1,526	-0.000070	1,747	-0.000165	1,747	-0.000089	1,747
Female	-0.000012	1,526	-0.000061	1,747	-0.000149	1,747	-0.000089	1,747
Age								
15 to 24	-0.000035	1,526	-0.000150	1,747	-0.000343	1,747	-0.000123	1,747
25 to 44	-0.000019	1,526	-0.000087	1,747	-0.000194	1,747	-0.000105	1,747
45 to 64	-0.000018	1,526	-0.000106	1,747	-0.000271	1,747	-0.000165	1,747
65 and over	-0.000033	1,526	-0.000254	1,747	-0.000641	1,747	-0.000479	1,747
Health insurance	-0.000010	3,240	-0.000114	4,653	-0.000274	4,653	-0.000086	4,653
Marital status, household and family								
Some household members	-0.000010	3,240	-0.000065	4,653	-0.000153	4,653	-0.000084	4,653
All household members	-0.000012	3,936	-0.000096	6,861	-0.000226	6,861	-0.000124	6,861
Mobility (movers)								
Educational attainment, labor force, marital status, HH, family, and income	-0.000006	1,783	-0.000025	1,783	-0.000059	1,783	-0.000032	1,783
US, county, state, region, or MSA	-0.000015	4,843	-0.000068	4,843	-0.000160	4,843	-0.000087	4,843
Below poverty								
Total	-0.000020	6,452	-0.000090	6,452	-0.000213	6,452	-0.000116	6,452
Male	-0.000042	6,452	-0.000191	6,452	-0.000440	6,452	-0.000232	6,452
Female	-0.000040	6,452	-0.000172	6,452	-0.000411	6,452	-0.000233	6,452
Age								
Under 15	-0.000081	4,974	-0.000289	4,974	-0.000620	4,974	-0.000307	4,974
Under 18	-0.000061	4,974	-0.000225	4,974	-0.000507	4,974	-0.000250	4,974
15 and over	-0.000025	6,452	-0.000117	6,452	-0.000277	6,452	-0.000143	6,452
15 to 24	-0.000056	2,441	-0.000209	2,441	-0.000479	2,441	-0.000172	2,441
25 to 44	-0.000030	2,441	-0.000121	2,441	-0.000271	2,441	-0.000146	2,441
45 to 64	-0.000029	2,441	-0.000148	2,441	-0.000379	2,441	-0.000230	2,441
65 and over	-0.000053	2,441	-0.000355	2,441	-0.000896	2,441	-0.000670	2,441
Unemployment	-0.000016	3,782	-0.000151	4,220	-0.000346	3,907	-0.000141	4,220
<b>FAMILIES, HOUSEHOLDS, OR UNRELATED INDIVIDUALS</b>								
Income	-0.000006	1,393	-0.000028	1,521	-0.000068	1,521	-0.000028	1,521
Marital status, HH and family, educational attainment, population by age/sex	-0.000005	1,285	-0.000022	1,163	-0.000052	1,163	-0.000022	1,163
Poverty	0.000052	1,518	0.000052	1,518	0.000052	1,518	0.000052	1,518

- NOTES: (1) These parameters are to be applied to the 2015 Annual Social and Economic Supplement data.  
(2) AIAN, NHOPI are American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander, respectively.  
(3) Hispanics may be any race. For a more detailed discussion on the use of parameters for race and ethnicity, please see the "Generalized Variance Parameters" section.



- (4) The Total or White, Black, and Asian, AIAN, NHOPI parameters are to be used for both alone and in-combination race group estimates.
- (5) For nonmetropolitan characteristics, multiply the  $a$  and  $b$  parameters by 1.5. If the characteristic of interest is total state population, not subtotaled by race or ancestry, the  $a$  and  $b$  parameters are zero.
- (6) For foreign-born and noncitizen characteristics for Total and White, the  $a$  and  $b$  parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Black, Asian, AIAN, NHOPI, and Hispanic.
- (7) For the group self-classified as having two or more races, use the Asian, AIAN, NHOPI parameters for all characteristics except employment, unemployment, and educational attainment, in which case use Black parameters.

**Table 6. CPS Year-to-Year Correlation Coefficients for Income and Health Insurance Characteristics: 1961 to 2015**

Characteristics	1961-2001 (basic) or 2001 (expanded)-2015		2000 (basic)- 2001 (expanded)	
	People	Families	People	Families
<b>Total</b>	<b>0.30</b>	<b>0.35</b>	<b>0.19</b>	<b>0.22</b>
White	0.30	0.35	0.20	0.23
Black	0.30	0.35	0.15	0.18
Other	0.30	0.35	0.15	0.17
Hispanic	0.45	0.55	0.36	0.28

- NOTES: (1) Correlation coefficients are not available for income data before 1961.
- (2) Hispanics may be any race. For a more detailed discussion on the use of parameters for race and ethnicity, please see the "Generalized Variance Parameters" section.
- (3) These correlation coefficients are for comparisons of consecutive years. For comparisons of nonconsecutive years, assume the correlation is zero.
- (4) For households and unrelated individuals, use the correlation coefficient for families.

**Table 7. CPS Year-to-Year Correlation Coefficients for Poverty Characteristics: 1971 to 2015**

Characteristics	1973-84, 1985-2001 (basic) or 2001 (expanded)-2015		2000 (basic)-2001 (expanded)		1984-1985		1972-1973		1971-1972	
	People	Families	People	Families	People	Families	People	Families	People	Families
<b>Total</b>	<b>0.45</b>	<b>0.35</b>	<b>0.29</b>	<b>0.22</b>	<b>0.39</b>	<b>0.30</b>	<b>0.15</b>	<b>0.14</b>	<b>0.31</b>	<b>0.28</b>
White	0.35	0.30	0.23	0.20	0.30	0.26	0.14	0.13	0.28	0.25
Black	0.45	0.35	0.23	0.18	0.39	0.30	0.17	0.16	0.35	0.32
Other	0.45	0.35	0.22	0.17	0.30	0.30	0.17	0.16	0.35	0.32
Hispanic	0.65	0.55	0.52	0.40	0.56	0.47	0.17	0.16	0.35	0.32

- NOTES: (1) Correlation coefficients are not available for income data before 1961.  
(2) Hispanics may be any race. For a more detailed discussion on the use of parameters for race and ethnicity, please see the "Generalized Variance Parameters" section.  
(3) These correlation coefficients are for comparisons of consecutive years. For comparisons of nonconsecutive years, assume the correlation is zero.  
(4) For households and unrelated individuals, use the correlation coefficient for families.

**Table 8. Factors and Populations for State Standard Errors and Parameters: 2015 CPS ASEC**

State	Factor	Population	State	Factor	Population
Alabama	1.11	4,777,204	Montana	0.21	1,012,902
Alaska	0.18	705,915	Nebraska	0.52	1,860,179
Arizona	1.25	6,673,829	Nevada	0.77	2,827,256
Arkansas	0.73	2,919,497	New Hampshire	0.33	1,314,170
California	1.28	38,497,170	New Jersey	1.15	8,857,996
Colorado	1.22	5,307,236	New Mexico	0.51	2,045,811
Connecticut	0.86	3,546,934	New York	1.19	19,550,311
Delaware	0.22	925,954	North Carolina	1.18	9,784,188
District of Columbia	0.17	655,492	North Dakota	0.17	734,769
Florida	1.14	19,758,009	Ohio	1.10	11,442,965
Georgia	1.15	9,947,436	Oklahoma	1.06	3,815,959
Hawaii	0.32	1,366,600	Oregon	1.07	3,955,151
Idaho	0.41	1,625,350	Pennsylvania	1.11	12,603,173
Illinois	1.17	12,695,960	Rhode Island	0.28	1,040,973
Indiana	1.11	6,517,923	South Carolina	1.07	4,761,878
Iowa	0.77	3,076,575	South Dakota	0.22	840,000
Kansas	0.82	2,847,866	Tennessee	1.10	6,478,545
Kentucky	1.13	4,333,831	Texas	1.32	26,738,307
Louisiana	1.01	4,556,832	Utah	0.53	2,941,673
Maine	0.39	1,316,942	Vermont	0.18	620,147
Maryland	1.15	5,906,593	Virginia	1.19	8,147,488
Massachusetts	1.10	6,696,193	Washington	1.18	7,001,607
Michigan	1.11	9,810,823	West Virginia	0.48	1,820,393
Minnesota	1.13	5,421,475	Wisconsin	1.13	5,694,823
Mississippi	0.69	2,925,298	Wyoming	0.16	574,436
Missouri	1.13	5,968,763			

- NOTES: (1) The state population counts in this table are for the 0+ population.  
(2) For foreign-born and noncitizen characteristics for Total and White, the *a* and *b* parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Black, Asian, AIAN, NHOPI, and Hispanic.

Table 9. Factors and Populations for Regional Standard Errors and Parameters: 2015 CPS ASEC		
Region	Factor	Population
Midwest	1.06	66,912,121
Northeast	1.07	55,546,839
South	1.13	118,252,904
West	1.12	74,534,936

- NOTES: (1) The state population counts in this table are for the 0+ population.  
 (2) For foreign-born and noncitizen characteristics for Total and White, the *a* and *b* parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Black, Asian, AIAN, NHOPI, and Hispanic.

### Year-to-Year Factors

In past years, the U.S. Bureau of the Census published a table of year factors for the Annual Social and Economic Supplement in the source and accuracies statement. User demand for these factors has diminished with the introduction of replicate weights. Data users producing estimates from prior years should consult the source and accuracy statements covering the years of their analysis to estimate standard errors. If further assistance is needed, please contact the Demographic Statistical Methods Division via e-mail at [dsmd.source.and.accuracy@census.gov](mailto:dsmd.source.and.accuracy@census.gov).

## References

- [1] Bureau of Labor Statistics, April 2014, “Redesign of the Sample for the Current Population Survey.” [http://www.bls.gov/cps/sample\\_redesign\\_2014.pdf](http://www.bls.gov/cps/sample_redesign_2014.pdf).
- [2] U.S. Census Bureau. 2006. *Current Population Survey: Design and Methodology*. Technical Paper 66. Washington, DC: Government Printing Office. (<http://www.census.gov/prod/2006pubs/tp-66.pdf>)
- [3] Brooks, C.A. and Bailar, B.A. 1978. *Statistical Policy Working Paper 3 - An Error Profile: Employment as Measured by the Current Population Survey*. Subcommittee on Nonsampling Errors, Federal Committee on Statistical Methodology, U.S. Department of Commerce, Washington, DC. (<http://www.fcsm.gov/working-papers/spp.html>)
- [4] U.S. Census Bureau. 1993. *Money Income of Households, Families, and Persons in the United States: 1992*. Current Population Reports, P60-184. Washington, DC: Government Printing Office. (<http://www2.census.gov/prod2/popscan/p60-184.pdf>)
- [5] U.S. Census Bureau. 1978. *Money Income in 1976 of Families and Persons in the United States*. Current Population Reports, P60-114. Washington, DC: Government Printing Office. (<http://www2.census.gov/prod2/popscan/p60-114.pdf>)
- [6] U.S. Census Bureau, July 15, 2009, “Estimating ASEC Variances with Replicate Weights Part I: Instructions for Using the ASEC Public Use Replicate Weight File to Create ASEC Variance Estimates.” ([http://www.bls.census.gov/pub/cps/march/Use\\_of\\_the\\_Public\\_Use\\_Replicate\\_Weight\\_File\\_final\\_PR.doc](http://www.bls.census.gov/pub/cps/march/Use_of_the_Public_Use_Replicate_Weight_File_final_PR.doc))

Accessed: February 3, 2012