Source and Accuracy of Estimates for *Income and Poverty in the United States: 2018*, *Health Insurance Coverage in the United States: 2018*, and *The Supplemental Poverty Measure: 2018*

SOURCE OF DATA

The estimates in the reports *Income and Poverty in the United States: 2018*, *Health Insurance Coverage in the United States: 2018*, and *The Supplemental Poverty Measure: 2018* come from the 2019 Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC).[1] The U.S. Census Bureau conducts the CPS ASEC over a 3-month period, in February, March, and April, with most 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 U.S. 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, consists primarily of the population in correctional institutions and nursing homes (98 percent of the 4.0 million institutionalized people in the 2010 Census). Starting in August 2017, college and university dormitories were also excluded from the population universe because most residents had usual residences elsewhere. Interviewers ask questions concerning labor force participation of 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.[2]

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 occurred during this phase-in period. First, primary sampling units (PSUs)[3] selected for only the 2010 design gradually replaced those selected for the 2000 design. This involved 10 percent of the sample. Second, within PSUs selected for both the 2000 and 2010 designs, sample households from the 2010 design gradually replaced those from the 2000 design. This involved about 90 percent of the sample. The new sample design was completely implemented by July 2015.

In the first stage of the sampling process, PSUs are selected for sample. The United States is divided into 1,987 PSUs. The PSUs were redefined for this design to correspond to the Office of Management and Budget definitions of Core Based Statistical Area definitions and to improve efficiency in field operations. These PSUs are grouped into 852 strata. Within each stratum, a single PSU is chosen for the sample, with its probability of selection proportional to its population as of the most recent decennial census. This PSU represents the entire stratum from which it was selected. In the case of strata consisting of only one PSU, the PSU is chosen with certainty.

Approximately 71,000 housing units were selected for sample from the sampling frame in March for the basic CPS. Based on eligibility criteria, 11 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).[4] Of all housing units in sample, about 60,000 were determined to be eligible for interview. Interviewers conducted interviews at about 48,900 of these units. Noninterviews occur when the occupants are not reached at home (after repeated attempts) or are unavailable for some other reason.

---

[1] Portions of the health insurance data in the report are based on the American Community Survey (ACS). Please refer to the ACS Source and Accuracy Statement <https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/ACS_Accuracy_of_Data_2017.pdf>.

[2] For detailed information on the 1990 sample redesign, see the U.S. Department of Labor, Bureau of Labor Statistics report, *Employment and Earnings*. "Redesign of the Sample for the Current Population Survey," Volume 51, Number 11, May 2004.

[3] The PSUs correspond to substate areas (i.e., counties or groups of counties) that are geographically contiguous.

[4] For further information on CATI and CAPI and the eligibility criteria, please see Technical Paper 66, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2002, <www.census.gov/prod/2006pubs/tp-66.pdf>.

Table 1 summarizes changes in the CPS design for the years in which data appear in this report.

Table 1.
Description of the March Basic CPS and ASEC Sample Cases

Time period	Number of sample PSUs	Basic CPS housing units eligible		Total (ASEC/ADS[A] + basic CPS) housing units eligible	
		Interviewed	Not interviewed	Interviewed	Not interviewed
2019	852	48,900	11,100	68,300	13,600
2018	852	50,800	9,900	67,900	11,500
2017	852	52,400	9,300	70,000	10,900
2016	852	52,000	9,100	69,500	10,600
2015	852	52,900	8,200	74,300	10,300
2014 Traditional	824	17,200	2,200	22,700	2,600
2014 Redesign	824	35,500	4,600	51,500	5,800
2013	824	52,900	6,400	75,500	7,700
2012	824	53,300	5,807	75,100	7,200
2011	824	53,400	5,300	75,900	6,500
2010	824	54,100	4,600	77,000	5,700
2009	824	54,100	4,600	76,900	5,700
2008	824	53,800	5,100	76,600	6,400
2007	824	53,700	5,600	76,100	7,100
2006	824	54,000	5,400	76,700	7,100
2005	[B]754/824	54,400	5,700	77,200	7,500
2004	754	55,000	5,200	77,700	7,000
2003	754	55,500	4,500	78,300	6,800
2002	754	55,500	4,500	78,300	6,600
2001	754	46,800	3,200	49,600	4,300
2000	754	46,800	3,200	51,000	3,700
1999	754	46,800	3,200	50,800	4,300
1998	754	46,800	3,200	50,400	5,200
1997	754	46,800	3,200	50,300	3,900
1996	754	46,800	3,200	49,700	4,100
1995	792	56,700	3,300	59,200	3,800
1990 to 1994	729	57,400	2,600	59,900	3,100
1989	729	53,600	2,500	56,100	3,000
1986 to 1988	729	57,000	2,500	59,500	3,000
1985	[C]629/729	57,000	2,500	59,500	3,000
1982 to 1984	629	59,000	2,500	61,500	3,000
1980 to 1981	629	65,500	3,000	68,000	3,500
1977 to 1979	614	55,000	3,000	58,000	3,500
1976	624	46,500	2,500	49,000	3,000
1973 to 1975	461	46,500	2,500	49,000	3,000
1972	[D]449/461	45,000	2,000	45,000	2,000

Table continues on next page.

Table 1.
Description of the March Basic CPS and ASEC Sample Cases—Con.

Time period	Number of sample PSUs	Basic CPS housing units eligible		Total (ASEC/ADS[A] + basic CPS) housing units eligible	
		Interviewed	Not interviewed	Interviewed	Not interviewed
1967 to 1971	449	48,000	2,000	48,000	2,000
1963 to 1966	357	33,400	1,200	33,400	1,200
1960 to 1962	333	33,400	1,200	33,400	1,200
1959	330	33,400	1,200	33,400	1,200

[A] The ASEC was referred to as the Annual Demographic Survey (ADS) until 2002.
[B] The Census Bureau redesigned the CPS following the 2000 Census. During phase-in of the new design, housing units from the new and old designs were in the sample.
[C] The Census Bureau redesigned the CPS following the 1980 Decennial Census of Population and Housing.
[D] The Census Bureau redesigned the CPS following the 1970 Decennial Census of Population and Housing.
Source: U.S. Census Bureau, Current Population Survey, 1959–2019 Annual Social and Economic Supplements.

The 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 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.
- Work status/occupation.
- Health insurance coverage.
- Program participation.
- Educational attainment.

Including the basic CPS sample, approximately 94,600 housing units were in the sample for the ASEC. About 81,900 housing units were determined to be eligible for interview and about 68,300 interviews were conducted (see Table 1).

The additional sample for the CPS ASEC provides more reliable data than the Basic CPS 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 Technical Paper 66, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2006, <www.census.gov/prod/2006pubs/tp-66.pdf>.

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, each state and 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 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 the demographic characteristics listed above 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:

- Net international migration of the foreign-born.
- Net migration between the United States and Puerto Rico.
- Net migration of natives to and from the United States.
- Net movement of the armed forces population to and from the United States.

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

The CPS ASEC estimation procedure includes a further adjustment to give married and unmarried partners the same weight.

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.

Answers to questions about money income often depend on the memory or knowledge of one person in a household. Recall problems can cause underestimates of income in survey data because it is easy to forget minor or irregular sources of income. Respondents may also misunderstand what the Census Bureau considers money income or may simply be unwilling to answer these questions correctly because the questions are considered too personal. See Appendix C, Current Population Reports, Series P60-184, *Money Income of Households, Families, and Persons in the United States: 1992,* <https://www2.census.gov/prod2/popscan/p60-184.pdf> for more details.

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. The basic CPS household-level unweighted nonresponse rate was 16.6 percent. The household-level CPS

ASEC unweighted nonresponse rate was an additional 18.9 percent. These two nonresponse rates lead to a combined supplement unweighted nonresponse rate of 32.4 percent.[5] 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 2019 CPS 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 nonrespondents. Refer to the supplement overview attachment in the technical documentation for the specific questions deemed critical by the sponsor as necessary to answer to be considered a sufficient partial interview.

As part of the nonsampling error analysis, the item response rates, item refusal rates, and edits are reviewed. For the CPS ASEC, the unweighted item refusal rates range from 0.0 percent to 23.0 percent. The unweighted item allocation rates range from 2.4 percent to 74.1 percent.

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 2019 is estimated to be about 11.0 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 2019 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 2019

Age	All people			White only		Black only		Residual race[A]		Hispanic[B]	
	Total	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0 to 15 years	0.87	0.86	0.88	0.90	0.92	0.72	0.71	0.80	0.82	0.79	0.80
16 to 19 years	0.83	0.85	0.80	0.89	0.82	0.69	0.75	0.79	0.71	0.88	0.83
20 to 24 years	0.76	0.77	0.75	0.80	0.79	0.58	0.62	0.83	0.69	0.80	0.77
25 to 34 years	0.82	0.79	0.85	0.83	0.88	0.55	0.73	0.78	0.78	0.75	0.83
35 to 44 years	0.90	0.88	0.91	0.92	0.95	0.68	0.81	0.80	0.79	0.78	0.85
45 to 54 years	0.90	0.89	0.91	0.90	0.94	0.80	0.78	0.87	0.81	0.81	0.90
55 to 64 years	0.94	0.93	0.95	0.96	0.97	0.79	0.89	0.85	0.84	0.95	0.95
65 years and older	0.97	0.98	0.97	1.00	0.98	0.91	0.92	0.86	0.83	0.86	0.88
15 years and older	0.89	0.88	0.90	0.91	0.93	0.71	0.80	0.82	0.79	0.82	0.86
0 years and older	0.89	0.88	0.89	0.91	0.93	0.71	0.78	0.82	0.80	0.81	0.84

[A] The Residual race group includes cases indicating a single race other than White or Black, and cases indicating two or more races.
[B] Hispanics may be any race.
Note: For a more detailed discussion on the use of parameters for race and ethnicity, please see the "Generalized Variance Parameters" section.
Source: U.S. Census Bureau, Current Population Survey, 2019 Annual Social and Economic Supplement.

[5] Because the ASEC is at the household level, the overall/combined ASEC response rate is a product of the basic CPS response rate and the ASEC response rate.

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 exercise caution when comparing estimates from the CPS ASEC for data year 2018 to estimates from earlier years. The data for 2018 estimates use the new data processing system. This new system introduces demographic edit changes to account for same-sex couples, revised procedures for editing income and health insurance variables, and several new income and health insurance variables. Changes to the editing procedures encompass both changes to the resolution of logically inconsistent data and changes to the imputation methods. The 2018 CPS ASEC estimates can be compared to the 2018 CPS ASEC Bridge Files,[6] which contain data year 2017 estimates, and to the 2017 CPS ASEC Research Files,[7] which contain estimates for data year 2016. The 2017 Research Files and the 2018 Bridge Files both use the new processing system and serve as a bridge between the old production files and the new processing system. Data users should be aware that the estimates from the 2017 and 2018 CPS ASEC files for data years 2016 and 2017 using the old production system are not directly comparable to the data year 2018 estimates from the 2019 CPS ASEC file.

Data users should be careful when comparing estimates for 2018 in *Income and Poverty in the United States: 2018* and *Health Insurance Coverage in the United States: 2018* (which reflect 2010 Census-based controls) with estimates for 2001 to 2010 (from March 2002 CPS to March 2011 CPS), which reflect 2000 Census-based controls, and to 1993 to 2000 (from March 1994 CPS to March 2001 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.

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 2010 Census-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. However, 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 2010 Census files (see "Basic CPS").[8] During this time period, CPS data were collected from sample designs based on different censuses. Two 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 and (2) the change in sample 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.

---

[6] For additional information on the 2018 CPS ASEC Bridge Files, see the Documentation and User Notes at <www.census.gov/data/datasets/2018/demo/income-poverty/cps-asec-bridge.html>.

[7] For additional information on the 2017 CPS ASEC Research Files, see the Documentation and User Notes at <www.census.gov/data/datasets/2017/demo/income-poverty/2017-cps-asec-research-file.html>.

[8] The phase-in process using the 2010 Census files began in April 2014.

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:

- Statistical Policy Working Paper 3, *An Error Profile: Employment as Measured by the Current Population Survey*, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1978. <https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/spwp3.pdf>.
- Technical Paper 66, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2006. <www.census.gov/prod/2006pubs/tp-66.pdf>.

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 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 2018 (2019 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 Series P-60, Number 114, *Money Income in 1976 of Families and Persons in the United States* <https://www2.census.gov/prod2/popscan/p60-114.pdf>.

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 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 Whites in poverty to the percentage of Blacks in poverty.

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.

The tables in *Income and Poverty in the United States: 2018* and *Health Insurance Coverage in the United States: 2018* list estimates followed by a number labeled "Margin of Error (±)." This number can be added to and subtracted from the estimates to calculate upper and lower bounds of the 90 percent confidence interval. For example, Table 1 in *Health Insurance Coverage in the United States: 2018* shows the numbers for health insurance. For the statement, "8.5 percent of people were uninsured for the entire calendar year," the 90 percent confidence interval for the estimate, 8.5 percent, is 8.5 (± 0.2) percent, or 8.3 percent to 8.7 percent.[9]

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 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 CPS ASEC 2019 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 generalized variance estimates refer to the "Generalized Variance Parameters" section. For more information on calculating direct estimates, see 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* at <https://usa.ipums.org/usa/resources/repwt/Use_of_the_Public_Use_Replicate_Weight_File_final_PR.doc>.

The *Income and Poverty in the United States: 2018*, *Health Insurance Coverage in the United States: 2018*, and *The Supplemental Poverty Measure: 2018* reports use replicate weights to calculate the standard errors of the estimates seen in tables and throughout the reports. The examples in this source and accuracy statement are for guidance calculating standard errors using the generalized variance parameters. The use of generalized variance parameters is the recommended method of calculating standard errors for data users who do not have the ability to calculate the standard errors using replicate weights.

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

The GVF 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 GVF parameters to use in standard error calculations.

[9] Note that the confidence interval here does not match the confidence interval given in Illustration 2 because the standard errors/margin of errors were calculated in two different ways. The margin of errors within the tables in the reports are calculated using direct estimates, whereas the standard errors within the illustrations later in this document are calculated using generalized variance estimates.

In this source and accuracy statement, Table 4 provides GVF parameters for characteristics from the 2019 CPS ASEC. Also, tables are provided that allow the calculation of parameters and standard errors for comparisons to adjacent years and the calculation of parameters for U.S. states and regions. Tables 5 and 6 contain correlation coefficients for comparing estimates from consecutive years. Table 7 contains the correlation coefficients for comparing race categories that are subsets of one another. Tables 8 and 9 provide factors and populations to derive U.S. 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

Race/ethnicity group of interest	Generalized variance parameters to use in standard error calculations
Total population	Total or White
Total White, White alone or in combination (AOIC), or White non-Hispanic population	Total or White
Total Black, Black AOIC, or Black non-Hispanic population	Black
Total Asian, Asian AOIC, or Asian non-Hispanic population	Asian, American Indian and Alaska Native (AIAN), Native Hawaiian and Other Pacific Islander (NHOPI)
Total AIAN, AIAN AOIC, or AIAN non-Hispanic population	Asian, AIAN, NHOPI
Total NHOPI, NHOPI AOIC, or NHOPI non-Hispanic population	Asian, AIAN, NHOPI
Populations from other race groups	Asian, AIAN, NHOPI
Hispanic[A] population	Hispanic[A]
Two or more races[B] —employment/unemployment and educational attainment characteristics	Black
Two or more races[B] —all other characteristics	Asian, AIAN, NHOPI

[A] Hispanics may be any race.
[B] Two or more races refers to the group of cases self-classified as having two or more races.
Notes: 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).
Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Standard Errors of Estimated Numbers. The approximate standard error, s_x, of an estimated number can be obtained 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 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.

Table 4.

Parameters for Computation of *Income and Poverty in the United States: 2018* and *Health Insurance Coverage in the United States: 2018* Standard Errors

Characteristic	Total or White		Black		Asian, AIAN, NHOPI[A]		Hispanic[B]	
	a	b	a	b	a	b	a	b
BELOW POVERTY LEVEL								
People								
Total	–0.000009	3,051	–0.000040	3,027	–0.000087	2,931	–0.000047	2,790
Male	–0.000020	3,111	–0.000085	3,088	–0.000174	2,819	–0.000098	2,935
Female	–0.000018	2,912	–0.000077	3,094	–0.000162	2,797	–0.000093	2,786
Age								
Under 15	–0.000062	3,812	–0.000276	4,853	–0.000559	4,699	–0.000260	4,362
Under 18	–0.000033	2,718	–0.000165	3,740	–0.000295	3,065	–0.000147	3,052
15 and older	–0.000013	3,332	–0.000054	3,246	–0.000117	3,050	–0.000065	3,215
15 to 24	–0.000078	3,307	–0.000244	2,850	–0.000532	2,921	–0.000179	2,686
25 to 44	–0.000032	2,751	–0.000136	3,005	–0.000266	2,675	–0.000150	2,676
45 to 64	–0.000036	3,008	–0.000141	2,469	–0.000360	2,566	–0.000211	2,548
65 and older	–0.000063	3,289	–0.000275	2,311	–0.000746	2,560	–0.000522	2,370
Households, Families, and Unrelated Individuals								
Total	0.000082	3,631	0.000314	3,646	0.000857	3,379	0.000206	3,565
ALL INCOME LEVELS								
People								
Total	–0.000008	2,641	–0.000030	2,310	–0.000070	2,338	–0.000037	2,209
Male	–0.000017	2,663	–0.000065	2,366	–0.000144	2,342	–0.000078	2,352
Female	–0.000014	2,238	–0.000052	2,077	–0.000128	2,218	–0.000070	2,106
Age								
15 to 24	–0.000073	3,099	–0.000250	2,917	–0.000467	2,568	–0.000162	2,442
25 to 44	–0.000031	2,645	–0.000120	2,664	–0.000260	2,620	–0.000149	2,657
45 to 64	–0.000034	2,846	–0.000132	2,296	–0.000355	2,528	–0.000202	2,434
65 and older	–0.000058	3,033	–0.000240	2,011	–0.000714	2,453	–0.000466	2,115
People by family income	–0.000013	4,333	–0.000062	4,722	–0.000119	3,984	–0.000069	4,122
Households, Families, and Unrelated Individuals								
Total	–0.000010	2,980	–0.000151	3,235	–0.000077	2,957	–0.000026	2,898
NONINCOME CHARACTERISTICS								
People								
Employment status	–0.000013	2,481	–0.000117	3,601	–0.000245	3,311	–0.000087	3,316
Educational attainment	–0.000009	3,021	–0.000035	2,688	–0.000078	2,628	–0.000406	2,434
Health insurance	–0.000008	2,688	–0.000031	2,334	–0.000072	2,400	–0.000041	2,460
Total, Marital Status, Other								
Some household members	–0.000007	2,197	–0.000035	2,673	–0.000055	1,841	–0.000033	1,990
All household members	–0.000008	2,450	–0.000029	2,236	–0.000060	2,023	–0.000032	1,937
Households, Families, and Unrelated Individuals								
Total	–0.000003	2,712	–0.000054	2,224	–0.000044	2,975	–0.000003	2,811

[A] AIAN, NHOPI are American Indian and Alaska Native and Native Hawaiian and Other Pacific Islander, respectively.

[B] Hispanics may be any race.

Notes: 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. For same-sex households, multiply the a- and b-parameters by 1.3. For foreign-born and noncitizen characteristics for Total and White, a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for other race/ethnicity groups. The Total or White, Black, and Asian, AIAN, NHOPI parameters are to be used for both alone and in-combination race group estimates. For the group self-classified as having two or more races, use the Asian, AIAN, NHOPI parameters for all characteristics except employment status and educational attainment—in which case, use Black parameters. For a more detailed discussion on the use of parameters for race and ethnicity, please see the "Generalized Variance Parameters" section.

Source: U.S. Census Bureau, Current Population Survey, 2019 Annual Social and Economic Supplement.

Illustration 1
In *Income and Poverty in the United States: 2018*, Table A-1 shows that there were 128,579,000 households in the United States in 2018. Use the appropriate parameters from Table 4 and Formula (1) to get

Number of households (x)	128,579,000
a parameter (a)	–0.000003
b parameter (b)	2,712
Standard error	547,000
90 percent confidence interval	127,679,000 to 129,479,000

The standard error is calculated as

$$s_x = \sqrt{-0.000003 \times 128{,}579{,}000^{\,2} + 2{,}712 \times 128{,}579{,}000}$$

which rounds to 547,000. The 90 percent confidence interval is calculated as 128,579,000 ± 1.645 × 547,000.

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 as indicated by the numerator. However, for calculating standard errors for different characteristics of families in poverty, use the standard error of a ratio equation (see Formula (4) in "Standard Errors of Estimated Ratios").

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

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

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

Illustration 2
In *Health Insurance Coverage in the United States: 2018*, Table 2 shows that there were 27,462,000 out of 323,688,000 people, or 8.5 percent, who did not have health insurance. Use the appropriate parameter from Table 4 and Formula (2) to get

Percentage of people without health insurance (p)	8.5
Base (y)	323,688,000
b parameter (b)	2,688
Standard error	0.08
90 percent confidence interval	8.4 to 8.6

The standard error is calculated as

$$s_{y,p} = \sqrt{\frac{2{,}688}{323{,}688{,}000} \times 8.5 \times (100 - 8.5\,)} = 0.08$$

The 90 percent confidence interval of the percentage of people without health insurance is calculated as 8.5 ± 1.645 × 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 5 and 6 contain the correlation coefficient r for year-to-year comparisons for CPS poverty, income, and health insurance estimates of numbers and proportions. Table 7 contains the correlation coefficient r for making comparisons between race categories that are subsets of one another. For example, to compare the number of people in poverty who listed White as their only race to the number of people in poverty who are White alone or in combination with another race, a correlation coefficient is needed to account for the large overlap between the two groups. For making other comparisons (including race overlapping where one group is not a complete subset of the other), 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.

Table 5.

CPS Year-to-Year Correlation Coefficients for Poverty Estimates: 1970 to 2018[A]

Characteristics	1972–1983, 1984–2000 (basic) or 2000 (expanded)–2018		1999 (basic)–2000 (expanded)		1983–1984		1971–1972		1970–1971	
	People	Families[B]	People	Families[B]	People	Families[B]	People	Families[B]	People	Families[B]
Total	**0.45**	**0.35**	**0.29**	**0.22**	**0.39**	**0.30**	**0.15**	**0.14**	**0.31**	**0.28**
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[C]	0.65	0.55	0.52	0.40	0.56	0.47	0.17	0.16	0.35	0.32

[A] Correlation coefficients are not available for poverty estimates before 1970.
[B] For households and unrelated individuals, use the correlation coefficient for families.
[C] 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.
Note: These correlations are for comparisons of consecutive years. For comparisons of nonconsecutive years, assume the correlation is zero.
Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Table 6.

CPS Year-to-Year Correlation Coefficients for Income and Health Insurance Estimates: 1960 to 2018[A]

Characteristics	1960–2000 (basic) or 2000 (expanded)–2018		1999 (basic)–2000 (expanded)	
	People	Families[B]	People	Families[B]
Total	**0.30**	**0.35**	**0.19**	**0.22**
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[C]	0.45	0.55	0.36	0.28

[A] Correlation coefficients are not available for income and health insurance estimates before 1960.
[B] For households and unrelated individuals, use the correlation coefficient for families.
[C] 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.
Note: These correlations are for comparisons of consecutive years. For comparisons of nonconsecutive years, assume the correlation is zero.
Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Table 7.

CPS Correlation Coefficients Between Race Groups and Subgroups: 2018

Race 1 (subgroup)	Race 2	r
White alone, not Hispanic	White alone	0.82
White alone, not Hispanic	White alone or in combination, not Hispanic	0.98
Black alone	Black alone or in combination	0.95
Asian alone	Asian alone or in combination	0.92

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Illustration 3

In *Income and Poverty in the United States: 2018*, Table B-1 shows that 11,869,000 out of 73,284,000 children, or 16.2 percent, were reported as in poverty in 2018, and that 12,759,000 out of 73,470,000, or 17.4 percent, were in poverty in 2017. The apparent difference is 1.2 percent. Use the appropriate parameters and correlation coefficient from Tables 4 and 5 and Formulas (2) and (3) to get

	2017 (x_1)	2018 (x_2)	Difference
Percentage of children in poverty (p)	17.4	16.2	1.2
Base (y)	73,470,000	73,284,000	–
b parameter (b)	4,974[A]	2,718	–
Correlation (r)	–	–	0.45
Standard error	0.31	0.22	0.29
90 percent confidence interval	16.9 to 17.9	15.8 to 16.6	0.7 to 1.7

[A] This value comes from the Source and Accuracy Statement for the 2017 Annual Social and Economic Supplement, Appendix G, Table 5 in <https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar17.pdf>.

The standard error of the difference is calculated as

$$s_{x_1-x_2} = \sqrt{0.31^2 + 0.22^2 - 2 \times 0.45 \times 0.31 \times 0.22} = 0.29$$

and the 90 percent confidence interval around the difference is calculated as 1.2 ± 1.645 × 0.29. Because this interval does not include zero, we can conclude that the percentage of children in poverty in 2018 was significantly different than the percentage of children in poverty in 2017.

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 year-to-year and subsetted race correlation coefficients see “Standard Errors of Estimated Differences.” 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 4

In *Income and Poverty in the United States: 2018,* Table B-2 shows that the number of families below the poverty level, x, was 7,504,000 and the total number of families, y, was 83,508,000. The ratio of families below the poverty level to the total number of families would be 0.090 or 9.0 percent. Use the appropriate parameters from Table 4 and Formulas (1) and (4) with r = 0 to get

	In poverty (x)	Total (y)	Ratio (in percent)
Number of families	7,504,000	83,508,000	9.0
a parameter (a)	0.000082	–0.000003	–
b parameter (b)	3,631	2,712	–
Standard error	179,000	453,000	0.22
90 percent confidence interval	7,210,000 to 7,798,000	82,763,000 to 84,253,000	8.6 to 9.4

The standard error is calculated as

$$s_{x/y} = \frac{7{,}504{,}000}{83{,}508{,}000}\sqrt{\left(\frac{179{,}000}{7{,}504{,}000}\right)^2 + \left(\frac{453{,}000}{83{,}508{,}000}\right)^2} = 0.0022 = 0.22\%$$

and the 90 percent confidence interval of the percentage is calculated as 9.0 ± 1.645 × 0.22.

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 (one standard error) 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:

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

   where

   X_{pN} = estimated upper and lower bounds for the confidence interval ($0 \le p \le 1$). For purposes of calculating the confidence interval, p takes on the values determined in step 2. Note that X_{pN} 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_{pN}.

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 two to obtain the standard error of the median.

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

Illustration 5

Suppose you want to calculate the standard error of the median of total money income for households with the following distribution:

Income Level	Number of households	Cumulative number of households	Cumulative percentage of households
Under $5,000	4,283,000	4,283,000	3.33%
$5,000 to $9,999	3,337,000	7,620,000	5.93%
$10,000 to $14,999	5,510,000	13,130,000	10.21%
$15,000 to $24,999	11,444,000	24,574,000	19.11%
$25,000 to $34,999	11,290,000	35,864,000	27.89%
$35,000 to $49,999	15,438,000	51,302,000	39.90%
$50,000 to $74,999	22,115,000	73,417,000	57.10%
$75,000 to $99,999	16,046,000	89,463,000	69.58%
$100,000 and over	39,117,000	[A]128,579,000	100.00%
Total number of households	128,579,000		

[A] This value may not equal the sum of the number of households due to rounding.

1. Using Formula (2) with b = 2,980 from Table 4, the standard error of 50 percent on a base of 128,579,000 is about 0.24 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 49.76 and 50.24.

3. The lower and upper limits for the interval in which the percentage limits fall are $50,000 and $75,000, respectively.

   Therefore, the estimated numbers of households with an income less than or equal to $50,000 and $75,000 are 51,302,000 and 73,417,000, respectively.

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

$$X_{pN} = \frac{0.4976 \times 128{,}579{,}000 - 51{,}302{,}000}{73{,}417{,}000 - 51{,}302{,}000}(75{,}000 - 50{,}000) + 50{,}000 = 64{,}333$$

   Similarly, the upper limit is found to be about

$$X_{pN} = \frac{0.5024 \times 128{,}579{,}000 \ - 51{,}302{,}000}{73{,}417{,}000 - 51{,}302{,}000}(75{,}000 - 50{,}000) + 50{,}000 = 65{,}031$$

Thus, a 68 percent confidence interval for the median income for households is from $64,333 to $65,031.

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

$$\frac{65{,}031 - 64{,}333}{2} = 349.0$$

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

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

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

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

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

$$\bar{x} = \sum_{i=1}^{c} p_i \bar{x}_i \tag{8}$$

and

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 (*x*-values) being considered, fall in group i.

$\bar{x}_i$ = $(Z_{L_i} + Z_{U_i})/2$ where Z_{L_i} and Z_{U_i} 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

$$\bar{x}_c = \frac{3}{2} Z_{L_c} \tag{9}$$

Illustration 6

In *Income and Poverty in the United States: 2018*, Table B-4 shows that there were 7,504,000 families in poverty and that the distribution of the income deficit for all families in poverty was

Income deficit	Number of families	Percentage of families (p_i)	Average ($\bar{x}_i$)
Under $1,000	536,000	7.1	500
$1,000 to $2,499	649,000	8.6	1,750
$2,500 to $4,999	935,000	12.5	3,750
$5,000 to $7,499	1,031,000	13.7	6,250
$7,500 to $9,999	861,000	11.5	8,750
$10,000 to $12,499	669,000	8.9	11,250
$12,500 to $14,999	621,000	8.3	13,750
$15,000 +	2,203,000	29.4	22,500
Total number of families	[A]7,504,000		

[A] This value may not equal the sum of the number of families due to rounding.

Using Formula (8),

$$\bar{x} = (0.071 \times 500) + (0.086 \times 1{,}750) + (0.125 \times 3{,}750) + (0.137 \times 6{,}250) + (0.115 \times 8{,}750) + (0.089 \times 11{,}250) + (0.083 \times 13{,}750) + (0.294 \times 22{,}500) = 11{,}275$$

and Formula (7),

$$S^2 = (0.071 \times 500^2) + (0.086 \times 1{,}750^2) + (0.125 \times 3{,}750^2) + (0.137 \times 6{,}250^2) + (0.115 \times 8{,}750^2) + (0.089 \times 11{,}250^2) + (0.083 \times 13{,}750^2) + (0.294 \times 22{,}500^2) - 11{,}275^2 = 64{,}863{,}000$$

Use the appropriate parameter from Table 4 and Formula (6) to get

Average income deficit for families in poverty $(\bar{x})$	$11,275
Variance (S^2)	64,863,000
Base (y)	7,504,000
b parameter (b)	3,631
Standard error	$177
90 percent confidence interval	$10,984 to $11,566

Note: This result is different from the average deficit for families in poverty and its 90 percent confidence interval in Table B-4 of the report because the report value is calculated using all (ungrouped) weighted data points.

The standard error is calculated as

$$s_{\bar{x}} = \sqrt{\frac{3{,}631}{7{,}504{,}000}(64{,}863{,}000)} = 177$$

and the 90 percent confidence interval is calculated as $11,275 ± 1.645 × $177.

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

$$x = \frac{hm}{p} \tag{10}$$

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

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

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 three-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 7

According to *Income and Poverty in the United States: 2018* Table B-2, there are 25,489,000 people living in families in poverty and 7,504,000 families in poverty. Continuing with Illustration 6, the average deficit income for families in poverty was $11,275 with a standard error of $177. Use the appropriate parameters from Table 4 and Formulas (1), (6), and (11) and r = 0.7 to get

	Number (h)	Number of people (p)	Average income deficit (m)	Average per capita deficit (x)
Value for families in poverty	7,504,000	25,489,000	$11,275	$3,319
a parameter (a)	0.000082	–0.000009	–	–
b parameter (b)	3,631	3,051	–	–
Correlation (r)	–	–	–	0.70
Standard error	179,000	268,000	$177	$80
90 percent confidence interval	7,210,000 to 7,798,000	25,048,000 to 25,930,000	$10,984 to $11,566	$3,187 to $3,451

Note: This result is different than the average per capita deficit for families in poverty and its standard error in Table B-4 of the report because of the different average income deficit calculated in Illustration 6.

The estimate of the average per capita deficit is calculated as

$$x = \frac{7{,}504{,}000 \times 11{,}275}{25{,}489{,}000} = 3{,}319$$

and the estimate of the standard error is calculated as

$$s_x = \frac{7{,}504{,}000 \times 11{,}275}{25{,}489{,}000} \sqrt{\left(\frac{177}{11{,}275}\right)^2 + \left(\frac{268{,}000}{25{,}489{,}000}\right)^2 + \left(\frac{179{,}000}{7{,}504{,}000}\right)^2 - 2 \times 0.7 \times \left(\frac{268{,}000}{25{,}489{,}000}\right) \times \left(\frac{179{,}000}{7{,}504{,}000}\right)} = 80$$

The 90 percent confidence interval is calculated as $3,319 $\pm$ 1.645 $\times$ $80.

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 *Income and Poverty in the United States* report no longer presents state estimates. The American Community Survey now provides those estimates. For CPS ASEC health insurance estimates for years prior to the 2014 CPS ASEC, the Census Bureau recommends using 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." The health insurance redesign of 2014 constitutes a "break-in-series" to the historical health insurance estimates, so data users should not use the 2015, 2016 production, or 2017 production CPS ASEC data with data prior to 2014 to create multiyear averages. Due to the change in

processing system, data users should not use the 2016 data from the 2017 CPS ASEC Research Files, 2017 data from the 2018 CPS ASEC Bridge Files, or 2018 CPS ASEC data with data from the 2016 production files, 2017 production files, or data prior to 2016 to create multiyear averages.

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 by the state factor from Table 8. To determine a new state-level a parameter (a_{state}), use the following.

(1) If the a parameter from Table 4 is positive, multiply the a parameter by the state factor from Table 8.

(2) If the a parameter in Table 4 is negative, calculate the new state-level a parameter as follows:

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

The state population is found in Table 8.

Table 8.
Factors and Populations for State Standard Errors and Parameters: 2018

State	Factor	Population	State	Factor	Population
Alabama	1.11	4,817,448	Montana	0.21	1,052,885
Alaska	0.18	710,173	Nebraska	0.52	1,907,145
Arizona	1.25	7,144,486	Nevada	0.77	3,038,584
Arkansas	0.73	2,966,518	New Hampshire	0.33	1,344,098
California	1.28	39,142,085	New Jersey	1.15	8,817,965
Colorado	1.22	5,654,833	New Mexico	0.51	2,060,106
Connecticut	0.86	3,522,174	New York	1.19	19,269,650
Delaware	0.22	958,985	North Carolina	1.18	10,253,278
District of Columbia	0.17	695,996	North Dakota	0.17	746,433
Florida	1.14	21,190,666	Ohio	1.10	11,534,632
Georgia	1.15	10,383,880	Oklahoma	1.06	3,868,308
Hawaii	0.32	1,362,507	Oregon	1.07	4,180,705
Idaho	0.41	1,757,250	Pennsylvania	1.11	12,612,112
Illinois	1.17	12,524,599	Rhode Island	0.28	1,041,657
Indiana	1.11	6,613,762	South Carolina	1.07	5,027,879
Iowa	0.77	3,120,802	South Dakota	0.22	870,059
Kansas	0.82	2,848,728	Tennessee	1.10	6,704,245
Kentucky	1.13	4,393,800	Texas	1.32	28,470,921
Louisiana	1.01	4,548,950	Utah	0.53	3,173,249
Maine	0.39	1,325,658	Vermont	0.18	621,051
Maryland	1.15	5,954,827	Virginia	1.19	8,337,856
Massachusetts	1.10	6,853,403	Washington	1.18	7,499,082
Michigan	1.11	9,903,633	West Virginia	0.48	1,769,612
Minnesota	1.13	5,582,357	Wisconsin	1.13	5,756,166
Mississippi	0.69	2,914,863	Wyoming	0.16	566,982
Missouri	1.13	6,026,740			

Source: U.S. Census Bureau, Current Population Survey, 2019 Annual Social and Economic Supplement, and internal data files for the 2010 Design.

Standard Errors for 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.

Table 9.
Factors and Populations for Regional Standard Errors and Parameters: 2018

Region	Factor	Population
Northeast	1.07	67,435,056
Midwest	1.06	55,407,768
South	1.13	123,258,032
West	1.12	77,342,927

Source: U.S. Census Bureau, Current Population Survey, 2019 Annual Social and Economic Supplement, and internal data files for the 2010 Design.

Illustration 8

In *Income and Poverty in the United States: 2018*, Table B-1 shows that there were 16,757,000 people living in poverty in the South. Use the appropriate parameter, factor, and population from Tables 4 and 9 and Formulas (1) and (12) to get:

Number of people living in poverty in the South	16,757,000
b parameter (b)	3,051
South factor	1.13
South population	123,258,032
South b parameter (b_{region})	3,448
South a parameter (a_{region})	–0.000028
Standard error	223,000
90 percent confidence interval	16,390,000 to 17,124,000

Obtain the region-level *b* parameter by multiplying the *b* parameter, 3,051, by the South regional factor, 1.13. This gives b_{region} = 3,051 × 1.13 =3,448. Obtain the needed region-level *a* parameter by

$$a_{region} = \frac{-3{,}448}{123{,}258{,}032} = -0.000028$$

The standard error of the estimate of the number of people living in the South in poverty can then be found by using Formula (1) and the new region-level a and b parameters, –0.000028 and 3,448, respectively. The standard error is given by

$$s_x = \sqrt{-0.000028 \times 16{,}757{,}000^{\,2} + 3{,}448 \times 16{,}757{,}000}$$

Which, rounded to the nearest thousand, is 223,000. The 90 percent confidence interval of the number of people living in poverty in the South is calculated as $16{,}757{,}000 \pm 1.645 \times 223{,}000$.

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:

$$state\ group\ factor = \frac{\sum_{i=1}^{n} POP_i \times state\ factor_i}{\sum_{i=1}^{n} POP_i} \qquad (13)$$

where POP_i and *state factor*$_i$ are the population and factor for state i from Table 8.

To obtain a new state group b parameter ($b_{state\ group}$), multiply the b parameter from Table 4 by the state factor obtained by Formula (13). To determine a new state group a parameter ($a_{state\ group}$), use the following:

(1) If the a parameter from Table 4 is positive, multiply the a parameter by the state group factor determined by Formula (13).

(2) If the a parameter in Table 4 is negative, calculate the new state group a parameter as follows:

$$a_{state\ group} = \frac{-b_{state\ group}}{\sum_{i=1}^{n} POP_i} \tag{14}$$

Illustration 9

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

$$state\ group\ factor = \frac{12{,}524{,}599 \times 1.17 + 6{,}613{,}762 \times 1.11 + 9{,}903{,}633 \times 1.11}{12{,}524{,}599 + 6{,}613{,}762 + 9{,}903{,}633} = 1.14$$

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

$$s_{\bar{x}} = \frac{s_x}{n} \tag{15}$$

where

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

and s_{x_i} are the standard errors of the estimates x_i. Tables 5 and 6 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, 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. Tables showing state data on income and poverty are available on the Internet and use 3-year average estimates for state-to-state comparisons and for certain race/ethnicity groups for health insurance estimates.[10] These tables also use 2-year averages to compare state estimates across years.

[10] Estimates of characteristics of the AIAN and 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.

Illustration 10

In *Income and Poverty in the United States: 2018*, Table B-7 provides the percentages of families in poverty. Suppose the percentage and bases of families with a female householder, no husband present in poverty for 2016,[11] 2017,[12] and 2018 are 26.7, 26.2, and 24.9 percent and 15,411,000, 15,305,000, and 15,052,000, respectively. Then the 2016–2018 3-year average percentage of families with a female householder, no husband present in poverty is 25.9. Use the appropriate parameters and correlation coefficients from Tables 4 and 5 and Formulas (15) and (16) to get:

	2016	2017	2018	2016–2018 avg
Percentage of families with a female householder in poverty (x)	26.7	26.2	24.9	25.9
Base (y)	15,411,000	15,305,000	15,052,000	–
b parameter (b)	[A]1,518	[B]1,518	3,631	–
Correlation (r)	–	–	–	0.35
Standard error	0.44	0.44	0.67	0.36
90 percent confidence interval	26.0 to 27.4	25.5 to 26.9	23.8 to 26.0	25.3 to 26.5

[A] This value comes from the Source and Accuracy Statement for the 2016 Annual Social and Economic Supplement, Appendix G, Table 5 in <https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar16.pdf>.

[B] This value comes from the Source and Accuracy Statement for the 2017 Annual Social and Economic Supplement, Appendix G, Table 5 in <https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar17.pdf>.

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

$$s_{\bar{x}} = \frac{1.09}{3} = 0.36$$

where

$$s_x = \sqrt{0.44^2 + 0.44^2 + 0.67^2 + (2 \times 0.35 \times 0.44 \times 0.44) + (2 \times 0.35 \times 0.44 \times 0.67)} = 1.09$$

The 90 percent confidence interval for the 3-year average percentage of families with a female householder, no husband present in poverty is $25.9 \pm 1.645 \times 0.36$.

Other Standard Errors. In the reports *Income and Poverty in the United States: 2018*, *Health Insurance Coverage in the United States: 2018*, and *The Supplemental Poverty Measure: 2018*, margins of error are provided for all key estimates. For other estimates, the standard errors can be calculated using the formulas in this Source and Accuracy Statement. For more information or questions about calculating standard errors, please contact the Demographic Statistical Methods Division via e-mail at <dsmd.source.and.accuracy@census.gov>.

[11] The estimates for data year 2016 come from the 2017 CPS ASEC Research files. The estimates from these files are not provided in Table B-7.

[12] The estimates for data year 2017 come from the CPS ASEC 2018 Bridge files.