Source and Accuracy of Estimates for *Income, Poverty, and Health Insurance Coverage in the United States: 2009*

SOURCE OF DATA

The estimates in the report *Income, Poverty, and Health Insurance Coverage in the United States: 2009* come from the 2010 Current Population Survey Annual Social and Economic Supplement (CPS ASEC). The U.S. Census Bureau conducts the ASEC over a 3-month period—in February, March, and April—with most data collection occurring in the month of March. The 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 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 (91 percent of the 4.1 million institutionalized people in Census 2000). 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.[1]

In April 2004, the Census Bureau began phasing out the 1990 sample and replacing it with the 2000 sample, creating a mixed sampling frame. Two simultaneous changes occurred during this phase-in period. First, primary sampling units (PSUs)[2] selected for only the 2000 design gradually replaced those selected for the 1990 design. This involved 10 percent of the sample. Second, within PSUs selected for both the 1990 and 2000 designs, sample households from the 2000 design gradually replaced sample households from the 1990 design. This involved about 90 percent of the sample. The new sample design was completely implemented by July 2005.

In the first stage of the sampling process, PSUs are selected for sample. The United States is divided into 2,025 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 824 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.

[1] For detailed information on the 1990 sample redesign, see the report *Employment and Earnings*, Volume 41 Number 5, U.S. Department of Labor, U.S. Bureau of Labor Statistics, May 1994.

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

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[1] + basic CPS) housing units eligible	
		Interviewed	Not interviewed	Interviewed	Not interviewed
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	[2]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	[3]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	[4]449/461	45,000	2,000	45,000	2,000
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

[1] The ASEC was referred to as the Annual Demographic Survey (ADS) until 2002.
[2] The Census Bureau redesigned the CPS following 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.
Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Approximately 72,000 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).[3] Of all housing units in sample, about 58,700 were determined to be eligible for interview. Interviewers obtained interviews at about 54,100 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 changes in the CPS design for the years in which data appear in this report.

[3] For further information on CATI and CAPI and the eligibility criteria, please see Technical Paper 66, *Current Population Survey: Design and Methodology*, U.S. Department of Commerce, U.S. Census Bureau, 2002, <www.census.gov/prod/2006pubs/tp66.pdf>.

The Annual Social and Economic Supplement. In addition to the basic CPS questions, interviewers asked supplementary questions for the 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 97,300 housing units were in sample for the ASEC. About 82,700 housing units were determined to be eligible for interview, and about 77,000 interviews were obtained (see Table 1).

The additional sample for the 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 ASEC, please refer to:

> Technical Paper 66, *Current Population Survey: Design and Methodology*, U.S. Department of Commerce, U.S. Census Bureau, 2006, <www.census.gov/prod/2006pubs/tp66.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 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 latest decennial 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.

The estimation procedure of the ASEC includes a further adjustment so the husband and wife of a household receive 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 or 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* at <www.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. For the cases eligible for the 2010 ASEC, the basic CPS household-level nonresponse rate was 7.8 percent. The household-level nonresponse rate for the ASEC was an additional 6.9 percent. These two nonresponse rates lead to a combined supplement nonresponse rate of 14.1 percent.

Coverage. 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 2010 is estimated to be about 12.0 percent. CPS coverage varies with age, sex, and race. Generally, coverage is larger for females than males and larger for non-Blacks than Blacks. This differential coverage is a general problem for most household-based surveys.

Table 2.
CPS Coverage Ratios: March 2010

Age	Total	All people		White only		Black only		Residual race		Hispanic[1]	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0 to 15 years	0.87	0.88	0.87	0.89	0.88	0.81	0.82	0.87	0.89	0.86	0.83
16 to 19 years	0.87	0.85	0.88	0.86	0.89	0.78	0.84	0.91	0.91	0.92	0.99
20 to 24 years	0.78	0.77	0.79	0.77	0.80	0.69	0.73	0.92	0.89	0.83	0.88
25 to 34 years	0.84	0.81	0.86	0.83	0.87	0.70	0.81	0.81	0.86	0.76	0.88
35 to 44 years	0.88	0.86	0.91	0.87	0.92	0.81	0.89	0.83	0.87	0.76	0.95
45 to 54 years	0.91	0.91	0.91	0.92	0.91	0.84	0.87	0.90	0.99	0.90	0.96
55 to 64 years	0.89	0.88	0.90	0.88	0.90	0.83	0.89	0.89	0.95	0.80	0.90
65 years and older	0.93	0.93	0.92	0.94	0.93	0.96	0.95	0.79	0.86	0.81	0.79
15 years and older	0.88	0.87	0.89	0.88	0.89	0.80	0.86	0.86	0.90	0.82	0.91
0 years and older	0.88	0.87	0.89	0.88	0.89	0.80	0.85	0.86	0.90	0.83	0.89

[1] 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: The Residual race group includes cases indicating a single race other than White or Black and cases indicating two or more races.

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

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 2010 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.

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 1999 to 2009 in *Income, Poverty, and Health Insurance Coverage in the United States: 2009* (which reflect Census 2000-based controls) with estimates for 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 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 2000-based controls results in about a 1 percent increase from the 1990 census-based controls in the civilian noninstitutionalized population and in the number of families and households. Thus, estimates of levels for data collected in 2003 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.

Note that certain microdata files from 2002—namely June, October, November, and the 2002 ASEC—contain both Census 2000-based estimates and 1990 census-based estimates and are subject to the comparability issues discussed previously. All other microdata files from 2002 reflect the 1990 census-based controls.

Users should also exercise caution because of changes caused by the phase-in of the Census 2000 files (see "Basic CPS"). 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 2004 through June 2005 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 (formerly called metropolitan areas). Most of the known effect on estimates during and after the sample redesign will be the result of changing from 1990 to 2000 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:

- Statistical Policy Working Paper 3, *An Error Profile: Employment as Measured by the Current Population Survey*, U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, 1978, <www.fcsm.gov/working-papers/spp.html>.

- Technical Paper 66, *Current Population Survey: Design and Methodology*, U.S. Department of Commerce, U.S. Census Bureau, 2006, <www.census.gov/prod/2006pubs/tp66.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. 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. This is because the width of the income interval containing the estimate is greater than $2,500.

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 2009 (2010 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* at <www.census.gov/prod2/popscan/p60-114.pdf>.

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 AOIC, or White non-Hispanic population	Total or White
Total Black, Black AOIC, or Black non-Hispanic population	Black
Total Asian, AIAN, NHOPI; Asian, AIAN, NHOPI AOIC; or Asian, AIAN, NHOPI non-Hispanic population	Asian, AIAN, NHOPI
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:
AIAN and NHOPI are American Indian and Alaska Native and Native Hawaiian and Other Pacific Islander, respectively.
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).
Hispanics may be any race.
Two or more races refers to the group of cases self-classified as having two or more races.
Source: U.S. Census Bureau, Demographic Statistical Methods Division.

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 tables in *Income, Poverty, and Health Insurance Coverage in the United States: 2009* list estimates followed by a number labeled "90 percent confidence interval (±)." 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 8 in *Income, Poverty, and Health Insurance Coverage in the United States: 2009* shows the numbers for health insurance. For the statement "the percentage of people without health insurance was 16.7 percent in 2009," the 90 percent confidence interval for the estimate 16.7 percent is 16.7 (± 0.2) percent, or 16.5 percent to 16.9 percent. Some tables also display asterisks in the last columns for significant differences between years.

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.

Table 4.
Parameters for Computation of *Income, Poverty, and Health Insurance Coverage in the United States: 2009* Standard Errors

Characteristic	Total or White		Black		Asian, AIAN, NHOPI[1]		Hispanic[2]	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*
BELOW POVERTY LEVEL								
People								
Total	–0.000017	5,282	–0.000086	5,282	–0.000227	5,282	–0.000108	5,282
Male	–0.000036	5,282	–0.000181	5,282	–0.000466	5,282	–0.000211	5,282
Female	–0.000034	5,282	–0.000162	5,282	–0.000442	5,282	–0.000222	5,282
Age								
Under 15	–0.000065	4,072	–0.000261	4,072	–0.000641	4,072	–0.000264	4,072
Under 18	–0.000049	4,072	–0.000201	4,072	–0.000529	4,072	–0.000218	4,072
15 and older	–0.000022	5,282	–0.000113	5,282	–0.000300	5,282	–0.000138	5,282
15 to 24	–0.000047	1,998	–0.000198	1,998	–0.000541	1,998	–0.000167	1,998
25 to 44	–0.000025	1,998	–0.000114	1,998	–0.000286	1,998	–0.000132	1,998
45 to 64	–0.000025	1,998	–0.000142	1,998	–0.000399	1,998	–0.000238	1,998
65 and older	–0.000052	1,998	–0.000382	1,998	–0.001034	1,998	–0.000710	1,998
Households, Families, and Unrelated Individuals								
Total	0.000052	1,243	0.000052	1,243	0.000052	1,243	0.000052	1,243
ALL INCOME LEVELS								
People								
Total	–0.000005	1,249	–0.000031	1,430	–0.000081	1,430	–0.000037	1,430
Male	–0.000011	1,249	–0.000066	1,430	–0.000169	1,430	–0.000073	1,430
Female	–0.000010	1,249	–0.000057	1,430	–0.000156	1,430	–0.000077	1,430
Age								
15 to 24	–0.000029	1,249	–0.000142	1,430	–0.000387	1,430	–0.000119	1,430
25 to 44	–0.000015	1,249	–0.000082	1,430	–0.000205	1,430	–0.000095	1,430
45 to 64	–0.000016	1,249	–0.000102	1,430	–0.000286	1,430	–0.000170	1,430
65 and older	–0.000032	1,249	–0.000273	1,430	–0.000740	1,430	–0.000508	1,430
People by family income	–0.000010	2,494	–0.000061	2,855	–0.000162	2,855	–0.000074	2,855
Households, Families, and Unrelated Individuals								
Total	–0.000005	1,140	–0.000027	1,245	–0.000071	1,245	–0.000032	1,245
NONINCOME CHARACTERISTICS								
People								
Employment status	–0.000016	3,068	–0.000151	3,455	–0.000346	3,198	–0.000141	3,455
Educational attainment	–0.000005	1,206	–0.000029	1,364	–0.000063	1,101	–0.000024	922
Health insurance	–0.000009	2,652	–0.000062	3,909	–0.000164	3,809	–0.000078	3,809
Total, Marital Status, Other								
Some household members	–0.000009	2,652	–0.000062	3,909	–0.000164	3,809	–0.000078	3,809
All household members	–0.000011	3,222	–0.000091	5,617	–0.000241	5,617	–0.000115	5,617
Households, Families, and Unrelated Individuals								
Total	–0.000004	1,052	–0.000020	952	–0.000054	952	–0.000025	952

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

[2] Hispanics may be any race.

Notes:

To obtain parameters prior to 2008, multiply by the appropriate factor in Table 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 ethnicity, the *a* and *b* parameters are zero.

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, Demographic Statistical Methods Division.

Table 5.
Year Factors for ASEC Estimates (1959 to 2008)[1]

Year of estimate	Total or White	Black[2]		Hispanic[3]
	a and *b*	*a* and *b*	*a*	*a* and *b*
2002 to 2008	1.00	1.00	1.00	1.00
2000 (expanded) to 2001	1.00	1.00	1.53	1.00
1995 to 2000 (basic)	1.97	1.97	3.00	1.97
1989 to 1994	1.82	1.82	2.78	1.82
1988	2.02	2.02	3.09	2.12
1984 to 1987	1.70	1.70	2.60	1.70
1981 to 1983	1.70	1.70	2.60	2.38
1972 to 1980	1.52	1.52	2.32	2.13
1966 to 1971	1.52	1.52	2.32	3.58
1959 to 1965	2.28	2.28	3.48	5.38

[1] Due to a change in the population control definitions, the parameters published in the source and accuracy statements for the *Income, Poverty, and Health Insurance Coverage in the United States* reports for 2002 and 2003 may not be identical to the product of the 2008 parameters (Table 4) and the year factors for 2002 and 2003 in this table.

[2] Blacks have two separate factors due to the revised race definitions introduced in 2003 (which apply to estimates from 2002 to the present) and their effect on the population control totals. Use the factors in the second Black column to get *a* parameters for all estimates of the Black population except those for Black families, households, and unrelated individuals in poverty—in which case, use the factor from the first Black column.

[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.

Note: For races not listed, use the factors for Total or White.

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

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 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. In this source and accuracy statement, Table 4 provides generalized variance parameters for characteristics from the 2010 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. Table 5 provides factors to derive prior year parameters. Tables 6 and 7 contain correlation coefficients for comparing estimates from consecutive years. Table 8 contains the correlation coefficients for comparing race categories that are subsets of one another. Tables 9 and 10 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 Alaska 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.

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.

Table 6.
CPS Year-to-Year Correlation Coefficients for Poverty Estimates: 1970 to 2009[1]

Characteristic	1972–1983, 1984–2000 (basic) or 2000 (expanded)–2009		1999 (basic)–2000 (expanded)		1983–1984		1971–1972		1970–1971	
	People	Families[2]	People	Families[2]	People	Families[2]	People	Families[2]	People	Families[2]
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[3]	0.65	0.55	0.52	0.40	0.56	0.47	0.17	0.16	0.35	0.32

[1] Correlation coefficients are not available for poverty estimates before 1970.
[2] For households and unrelated individuals, use the correlation coefficient for families.
[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.
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.

Standard Errors of Estimated Numbers. The approximate standard error, s_x, of an estimated number shown in *Income, Poverty, and Health Insurance Coverage in the United States: 2009* can be obtained using the formula

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

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.

Illustration 1

In *Income, Poverty, and Health Insurance Coverage in the United States: 2009*, Table 1 shows that there were 117,538,000 households in the United States in 2009. Use the appropriate parameters from Table 4 and Formula (1) to get:

Number of households (x)	117,538,000
a parameter (a)	–0.000004
b parameter (b)	1,052
Standard error	262,000
90 percent confidence interval	117,107,000 to 117,969,000

The standard error is calculated as

$$s_x = \sqrt{-0.000004 \times 117{,}538{,}000^2 + 1{,}052 \times 117{,}538{,}000} = 262{,}000$$

and the 90 percent confidence interval is calculated as 117,538,000 ± 1.645 × 262,000.

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.

Table 7.
CPS Year-to-Year Correlation Coefficients for Income and Health Insurance Estimates: 1960 to 2009[1]

Characteristic	1960–2000 (basic) or 2000 (expanded)–2009		1999 (basic)–2000 (expanded)	
	People	Families[2]	People	Families[2]
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[3]	0.45	0.55	0.36	0.28

[1] Correlation coefficients are not available for income and health insurance estimates before 1960.

[2] For households and unrelated individuals, use the correlation coefficient for families.

[3] Hispanics may be any race. For a more detailed discusson 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.

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:

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

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 *Income, Poverty, and Health Insurance Coverage in the United States: 2009*, Table C-1 shows that there were 50,674,000 out of 304,280,000 people, or 16.7 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)	16.7
Base (y)	304,280,000
b parameter (b)	2,652
Standard error	0.11
90 percent confidence interval	16.5 to 16.9

The standard error is calculated as

$$s_{y,p} = \sqrt{\frac{2{,}652}{304{,}280{,}000} \times 16.7 \times (100 - 16.7)} = 0.11$$

Table 8.
CPS Correlation Coefficients Between Race Groups and Subgroups: 2009

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.

The 90 percent confidence interval of the percentage of people without health insurance is calculated as 16.7 ± 1.645 × 0.11.

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_{x1} and s_{x2} are the standard errors of the estimates x_1 and x_2. The estimates can be numbers, percentages, ratios, etc. Tables 6 and 7 contain the correlation coefficient r for year-to-year comparisons for CPS poverty, income, and health insurance estimates of numbers and proportions. Table 8 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 $r = 0$. 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 3

In *Income, Poverty, and Health Insurance Coverage in the United States: 2009*, Table C-1 shows that 50,674,000 out of 304,280,000 people, or 16.7 percent, were not covered by health insurance in 2009, and that 46,340,000 out of 301,483,000 people, or 15.4 percent, were not covered by health insurance in 2008. The apparent difference is 1.3 percent. Use the appropriate parameters, year factor, and correlation coefficient from Tables 4, 5, and 7 and Formulas (2) and (3) to get:

	2009 (x_1)	2008 (x_2)	Difference
Percentage of people without health insurance (p)	16.7	15.4	1.3
Base (y)	304,280,000	301,483,000	–
b parameter (b)	2,652	[1]2,652	–
Correlation (r)	–	–	0.30
Standard error	0.11	0.11	0.13
90 percent confidence interval	16.5 to 16.9	15.2 to 15.6	1.1 to 1.5

[1] This parameter is calculated by multiplying the year factor for 2008 (from Table 5), 1.0, by the current b parameter.

The standard error of the difference is calculated as

$$s_{x_1-x_2} = \sqrt{0.11^2 + 0.11^2 - 2 \times 0.30 \times 0.11 \times 0.11} = 0.13$$

and the 90 percent confidence interval around the difference is calculated as 1.3 ± 1.645 × 0.13. Since this interval does not include zero, we can conclude with 90 percent confidence that the percentage of people without health insurance in 2009 was higher than the percentage of people without health insurance in 2008.

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, Poverty, and Health Insurance Coverage in the United States: 2009*, Table B-3 shows that the number of families below the poverty level, x, was 8,792,000 and the total number of families, y, was 78,867,000. The ratio of families below the poverty level to the total number of families would be 0.111 or 11.1 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	8,792,000	78,867,000	11.1
a parameter (a)	+0.000052	–0.000004	–
b parameter (b)	1,243	1,052	–
Standard error	112,000	241,000	0.16
90 percent confidence interval	8,591,000 to 8,993,000	78,471,000 to 79,263,000	10.8 to 11.4

The standard error is calculated as

$$s_{x/y} = \frac{8{,}792{,}000}{78{,}867{,}000}\sqrt{\left(\frac{112{,}000}{8{,}792{,}000}\right)^2 + \left(\frac{241{,}000}{78{,}867{,}000}\right)^2} = 0.0016 = 0.16\%$$

and the 90 percent confidence interval of the percentage is calculated as 11.1 ± 1.645 × 0.16.

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. Determine, using Formula (2), the standard error of the estimate of 50 percent from the distribution.

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 ($0 \leq p \leq 1$). 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 2 to obtain the standard error of the median.

   Note: Median incomes and their standard errors as calculated below 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	3,747,000	3,747,000	3.19
$5,000 to $9,999	4,823,000	8,570,000	7.29
$10,000 to $14,999	6,759,000	15,329,000	13.04
$15,000 to $24,999	14,023,000	29,352,000	24.97
$25,000 to $34,999	13,003,000	42,355,000	36.03
$35,000 to $49,999	16,607,000	58,962,000	50.16
$50,000 to $74,999	21,279,000	80,241,000	68.27
$75,000 to $99,999	13,549,000	93,790,000	79.79
$100,000 and over	23,749,000	[1]117,539,000	100.00
Total number of households	117,538,000		

[1] This number does not equal the total number of households because of rounding.

1. Using Formula (2) with b = 1,140 from Table 4, the standard error of 50 percent on a base of 117,538,000 is about 0.16 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 49.84 and 50.16.

3. The lower and upper limits for the interval in which the percentage limits falls are $35,000 and $49,999, respectively.

   Therefore, the estimated numbers of households with an income less than or equal to $35,000 and $49,999 are 42,354,000 and 58,961,000, respectively.

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

$$X_{pN} = \frac{0.4984 \times 117{,}538{,}000 - 42{,}354{,}000}{58{,}961{,}000 - 42{,}354{,}000}(50{,}000 - 35{,}000) + 35{,}000 = 49{,}657$$

   Similarly, the upper limit is found to be about

$$X_{pN} = \frac{0.5016 \times 117{,}538{,}000 - 42{,}354{,}000}{58{,}961{,}000 - 42{,}354{,}000}(50{,}000 - 35{,}000) + 35{,}000 = 49{,}996$$

   Thus, a 68 percent confidence interval for the median income for households is from $49,657to $49,996.

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

$$\frac{49{,}996 - 49{,}657}{2} = 169.5$$

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

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

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.

$\bar{x}_i$ = $(Z_{Li} + Z_{Ui})/2$ 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

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

Illustration 6

In *Income, Poverty, and Health Insurance Coverage in the United States: 2009*, Table 7 shows that there were 8,792,000 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 $500	268,000	3.0	250
$500 to $999	306,000	3.5	750
$1,000 to $1,999	667,000	7.6	1,500
$2,000 to $2,999	657,000	7.5	2,500
$3,000 to $3,999	509,000	5.8	3,500
$4,000 to $4,999	563,000	6.4	4,500
$5,000 to $5,999	542,000	6.2	5,500
$6,000 to $6,999	526,000	6.0	6,500
$7,000 to $7,999	472,000	5.4	7,500
$8,000 or more	4,282,000	48.7	12,000
Total number of families	[1]8,792,000		

[1] This value doesn't equal the sum of the number of families due to rounding.

Using Formula (8),

$$\bar{x} = (0.030 \times 250) + (0.035 \times 750) + (0.076 \times 1{,}500) + (0.075 \times 2{,}500) + (0.058 \times 3{,}500) + (0.064 \times 4{,}500) + (0.062 \times 5{,}500) + (0.060 \times 6{,}500) + (0.054 \times 7{,}500) + (0.487 \times 12{,}000) = 7{,}806$$

and Formula (7),

$$S^2 = (0.030 \times 250^2) + (0.035 \times 750^2) + (0.076 \times 1{,}500^2) + (0.075 \times 2{,}500^2) + (0.058 \times 3{,}500^2) + (0.064 \times 4{,}500^2) + (0.062 \times 5{,}500^2) + (0.060 \times 6{,}500^2) + (0.054 \times 7{,}500) + (0.487 \times 12{,}000) - 7{,}806^2 = 19{,}310{,}000$$

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

Average income deficit for families in poverty $(\bar{x})$	$7,806
Variance (S^2)	19,310,000
Base (y)	8,792,000
b parameter (b)	1,243
Standard error	$52
90 percent confidence interval	$7,720 to $7,892

Note: This result is different from the average deficit for families in poverty and its 90 percent confidence interval in Table 7 of the report because the report value is calculated using all (ungrouped) weighted data points.

The standard error is calculated as

$$s_{\bar{x}} = \sqrt{\frac{1{,}243}{8{,}792{,}000}(19{,}310{,}000)} = 52$$

and the 90 percent confidence interval is calculated as $7,806 ± 1.645 × $52.

Standard Errors of Estimated Per Capita Deficits. Certain average values in reports associated with the 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 (7), r represents the correlation between p and h.

Table 9.
Factors and Populations for State Standard Errors and Parameters: 2009

State	Factor	Population	State	Factor	Population
Alabama	1.05	4,645,546	Montana	0.24	962,760
Alaska	0.18	683,794	Nebraska	0.46	1,774,347
Arizona	1.23	6,558,845	Nevada	0.67	2,624,964
Arkansas	0.68	2,850,634	New Hampshire	0.34	1,309,649
California	1.25	36,633,824	New Jersey	1.12	8,620,926
Colorado	1.20	4,991,493	New Mexico	0.58	1,990,273
Connecticut	0.88	3,468,282	New York	1.17	19,334,965
Delaware	0.22	875,390	North Carolina	1.11	9,242,116
District of Columbia	0.18	595,918	North Dakota	0.16	634,281
Florida	1.12	18,260,440	Ohio	1.09	11,375,790
Georgia	1.08	9,690,023	Oklahoma	0.91	3,624,901
Hawaii	0.29	1,255,124	Oregon	1.01	3,806,331
Idaho	0.36	1,531,854	Pennsylvania	1.09	12,414,017
Illinois	1.13	12,757,268	Rhode Island	0.30	1,036,241
Indiana	1.08	6,343,904	South Carolina	1.06	4,486,745
Iowa	0.77	2,972,621	South Dakota	0.17	798,436
Kansas	0.73	2,772,546	Tennessee	1.08	6,227,479
Kentucky	1.05	4,243,338	Texas	1.28	24,592,116
Louisiana	1.05	4,419,866	Utah	0.54	2,793,918
Maine	0.39	1,300,497	Vermont	0.18	615,606
Maryland	1.13	5,624,897	Virginia	1.08	7,704,704
Massachusetts	1.06	6,533,444	Washington	1.15	6,616,376
Michigan	1.09	9,820,392	West Virginia	0.39	1,799,079
Minnesota	1.07	5,220,758	Wisconsin	1.10	5,588,842
Mississippi	0.71	2,888,230	Wyoming	0.15	542,006
Missouri	1.11	5,897,451			

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

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, Poverty and Health Insurance Coverage in the United States: 2009*, Table 4, there are 31,197,000 people living in families in poverty and 8,792,000 families in poverty. Continuing with Illustration 6, the average deficit income for families in poverty was $7,806 with a standard error of $52. 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	8,792,000	31,197,000	$7,806	$2,200
a parameter (a)	+0.000052	–0.000018	–	–
b parameter (*b*)	1,243	5,282	–	–
Correlation (*r*)	–	–	–	0.7
Standard error	122,000	384,000	$52	$27
90 percent confidence interval	8,591,000 to 8,993,000	30,565,000 to 31,829,000	$7,720 to $7,892	$2,156 to $2,244

Note: This result is different than the average per capita deficit for families in poverty and its standard error in Table 7 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

$$x = \frac{8{,}792{,}000 \times 7{,}806}{31{,}197{,}000} = 2{,}200$$

and the estimate of the standard error is calculated as

$$s_x = \frac{8{,}792{,}000 \times 7{,}806}{31{,}197{,}000} \sqrt{\left(\frac{52}{7{,}806}\right)^2 + \left(\frac{384{,}000}{31{,}197{,}000}\right)^2 + \left(\frac{122{,}000}{8{,}792{,}000}\right)^2 - 2 \times 0.7 \times \left(\frac{384{,}000}{31{,}197{,}000}\right) \times \left(\frac{122{,}000}{8{,}792{,}000}\right)} = 27$$

The 90 percent confidence interval is calculated as \$2,200 ± 1.645 × \$27.

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, Poverty, and Health Insurance Coverage in the United States* report series no longer presents state estimates for income, poverty, and health insurance coverage. The American Community Survey now provides those estimates. For ASEC estimates, 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."

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 9. 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 9.

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

$$a_{state} = \frac{-b_{state}}{State\ Population} \tag{12}$$

The state population is found in Table 9.

Table 10.
Factors and Populations for Regional Standard Errors and Parameters: 2009

Region	Factor	Population
Northeast	1.05	54,633,627
Midwest	1.03	65,956,636
South	1.08	111,761,422
West	1.10	70,991,562

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

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 10.

Illustration 8

In *Income, Poverty, and Health Insurance Coverage in the United States: 2009*, Table 4 shows that there were 17,609,000 people living in poverty in the South. Use the appropriate parameter, factor, and population from Tables 4 and 10 and Formulas (1) and (12) to get:

Number of people living in poverty in the South	17,609,000
b parameter (*b*)	5,282
South factor	1.08
South population	111,761,422
South *a* parameter (a_{region})	–0.000051
South *b* parameter (b_{region})	5,705
Standard error	291,000
90 percent confidence interval	17,130,000 to 18,088,000

Obtain the region-level *b* parameter by multiplying the *b* parameter, 5,282, by the South regional factor, 1.08. This gives b_{region} = 5,282 × 1.08 = 5,705. Obtain the needed region-level *a* parameter by

$$a_{region} = \frac{-5{,}705}{111{,}761{,}422} = -0.000051$$

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, –0.000051 and 5,705, respectively. The standard error is given by

$$s_x = \sqrt{-0.000051 \times 17{,}609{,}000^2 + 5{,}705 \times 17{,}609{,}000} = 291{,}000$$

and the 90 percent confidence interval of the number of people living in poverty in the South is calculated as 17,609,000 ± 1.645 × 291,000.

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:

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

where POP_i (the state population for state i) and the state factors are from Table 9.

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:

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

Illustration 9

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

$$\text{state group factor} = \frac{12{,}757{,}268 \times 1.13 + 6{,}343{,}904 \times 1.08 + 9{,}820{,}392 \times 1.09}{12{,}757{,}268 + 6{,}343{,}904 + 9{,}820{,}392} = 1.11$$

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_{xi} 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_{xi} 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, 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.

Illustration 10

Suppose the 2007–2009 3-year-average percentage of the AIAN population[4] without health insurance is 30.1, and the percentages and bases for 2007, 2008, and 2009 are 32.1, 29.2, and 29.1 percent and 2,745,000, 2,852,000, and 2,681,000, respectively. Use the appropriate parameters, factors, and correlation coefficients from Tables 4, 5, and 7 and Formulas (15) and (16) to get:

	2007	2008	2009	2007–2009 average
Percentage of AIAN population without health insurance (x)	32.1	29.2	29.1	30.1
Base (*x*)	2,745,000	2,852,000	2,681,000	–
b parameter (*b*)	[1]3,809	[1]3,809	3,809	–
Correlation (*r*)	–	–	–	0.30, 0.30
Standard error	1.74	1.66	1.71	1.16
90 percent confidence interval	29.2 to 35.0	26.5to 31.9	26.3 to 31.9	28.2 to 32.0

[1] These parameters are calculated by multiplying the year factor from Table 5, 1.0, by the current parameter.

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

$$s_{\bar{x}} = \frac{3.48}{3} = 1.16$$

where

$$s_x = \sqrt{1.74^2 + 1.66^2 + 1.71^2 + (2 \times 0.30 \times 1.74 \times 1.66) + (2 \times 0.30 \times 1.66 \times 1.71)} = 3.48$$

The 90 percent confidence interval for the 3-year-average percentage of the AIAN population without health insurance is $30.1 \pm 1.645 \times 1.16$.

Standard Errors of Differences of 2-Year Averages. Comparing two nonoverlapping 2-year averages also improves precision for comparisons across years. Use the formulas described previously to estimate the standard error s_{xi} 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 nonoverlapping 2-year averages $\bar{x}_{1,2} - \bar{x}_{3,4}$ is

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

[4] 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.

Illustration 11

Suppose that you want to calculate the standard error of the difference between the 2006, 2007, 2008, and 2009 2-year averages of the percentage of people in California without health insurance. Use the following information along with the appropriate parameters and factors from Tables 4, 5, 9, and Formula (2) to get:

	2006	2007	2008	2009
Percentage of people in California without health insurance (p)	18.8	18.2	18.6	20.0
Base (y)	36,208,000	36,295,000	36,691,000	36,749,000
b parameter (b)	[1]2,652	[1]2,652	[1]2,652	2,652
California state factor	1.25	1.25	1.25	1.25
State b parameter (b_{state})	3,315	3,315	3,315	3,315
Standard error[2]	0.37	0.37	0.37	0.38

[1] These parameters are calculated by multiplying the year factors from Table 6 by the current parameter.
[2] See "Standard Errors of State Estimates" for instructions and illustrations on calculating state standard errors.

Use this information, Formulas (15), (16), and (17), and the appropriate correlation coefficient from Table 7 to get:

	2006, 2007	2007, 2008	2008, 2009	Average (2006, 2007)–average (2008, 2009)
Average percentage of people in California without health insurance ($\bar{x}$)	18.5	–	19.3	0.8
Correlation (r)	0.30	0.30	0.30	–
Standard error	[1]0.30	–	[1]0.30	0.40
90 percent confidence interval	18.0 to 19.0	–	18.8 to 19.8	0.1 to 1.5

[1] 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

$$s_{x_{1,2}-x_{3,4}} = \sqrt{0.30^2 + 0.30^2 - \frac{1}{2} \times 0.30 \times 0.37 \times 0.37} = 0.40$$

and the 90 percent confidence interval around the difference of the 2-year averages is calculated as $0.8 \pm 1.645 \times 0.40$. Since this interval does include zero, we cannot conclude with 90 percent confidence that the 2008–2009 average percentage of people in California without health insurance was higher than the 2006–2007 average percentage of people in California without health insurance.

Other Standard Errors. In the report *Income, Poverty, and Health Insurance Coverage in the United States: 2009*, eight tables provide confidence intervals for most of the estimates discussed in the text. For other estimates, the standard errors can be calculated using the formulas in this source and accuracy statement. For more information or if you have questions on calculating standard errors, please contact the Demographic Statistical Methods Division via e-mail at <dsmd.source.and.accuracy@census.gov>.