Source of the Data and Accuracy of the Estimates for the June 2018 Current Population Survey Microdata File on Fertility and Birth Expectation

Table of Contents

Tables

Source of the Data and Accuracy of the Estimates for the June 2018 Current Population Survey Microdata File on Fertility and Birth Expectation

SOURCE OF THE DATA

The data in this microdata file are from the June 2018 Current Population Survey (CPS). The U.S. Census Bureau conducts the CPS every month, although this file has only June data. The June survey 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 civilian noninstitutionalized population. The Census Bureau and the U.S. Bureau of Labor Statistics also jointly sponsor the supplemental questions for June.

Basic CPS. The monthly CPS collects primarily labor force data about the civilian noninstitutionalized population living in the United States. The institutionalized population, which is excluded from the population universe, is composed primarily of the population in correctional institutions and nursing homes (98 percent of the 4.0 million institutionalized people in Census 2010). Starting August 2017, college and university dormitories were also excluded from the population universe because the majority of the residents had usual residences elsewhere. Interviewers ask questions concerning labor force participation about each member 15 years old and over in sample households. Typically, the week containing the nineteenth of the month is the interview week. The week containing the twelfth is the reference week (i.e., the week about which the labor force questions are asked).

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

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

Approximately 71,000 housing units were selected for sample from the sampling frame in June. Based on eligibility criteria, nine percent of these housing units were sent directly to

---

1 For detailed information on the 2010 sample redesign, please see Bureau of Labor Statistics (2014).

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

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 60,000 were determined to be eligible for interview. Interviewers obtained interviews at about 51,000 of these units. Noninterviews occur when the occupants are not found at home after repeated calls or are unavailable for some other reason.

June 2018 Supplement. In June 2018, in addition to the basic CPS questions, interviewers asked supplementary questions of women 15 to 44 years of age on fertility.

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

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

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

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

The independent estimates by age, sex, race, and Hispanic origin, and for states by selected age groups and broad race categories, are developed using the basic demographic accounting formula whereby the population from the 2010 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 of the population estimates includes:

- 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; and
- Net movement of the Armed Forces population to and from the United States.

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.

---

[3] For further information on CATI and CAPI and the eligibility criteria, please see U.S. Census Bureau (2006).

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

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

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

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

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

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

Nonresponse. The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. For the June 2018 basic CPS, the household-level nonresponse rate was 15.7 percent. The person-level nonresponse rate for the Fertility supplement was an additional 8.6 percent.

Since the basic CPS nonresponse rate is a household-level rate and the Fertility supplement nonresponse rate is a person-level rate, we cannot combine these rates to derive an overall nonresponse rate. Nonresponding households may have fewer persons than interviewed ones, so combining these rates may lead to an overestimate of the true overall nonresponse rate for persons for the Fertility supplement.

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

As part of the nonsampling error analysis, the item response rates, item refusal rates, and edits are reviewed. For the Fertility supplement, the item refusal rates range from 1.4 percent to 4.6 percent. This survey is fully allocated. The item nonresponse rates range from 18.1 percent to 27.9 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 June 2018 is estimated to be about 11 percent. CPS coverage varies with age, sex, and race. Generally, coverage is higher for females than for males and higher 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 1 shows June 2018 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 1. Current Population Survey Coverage Ratios: June 2018

Age group	All people	Total		White only		Black only		Residual race[A]		Hispanic[B]	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-15	0.87	0.88	0.87	0.92	0.91	0.71	0.71	0.86	0.80	0.81	0.82
16-19	0.86	0.88	0.84	0.90	0.88	0.77	0.65	0.88	0.89	0.82	0.83
20-24	0.75	0.75	0.75	0.80	0.78	0.59	0.68	0.68	0.69	0.69	0.72
25-34	0.82	0.80	0.85	0.84	0.89	0.63	0.73	0.75	0.73	0.70	0.83
35-44	0.90	0.88	0.92	0.90	0.96	0.76	0.78	0.84	0.85	0.79	0.90
45-54	0.90	0.90	0.91	0.94	0.94	0.70	0.77	0.81	0.86	0.80	0.87
55-64	0.93	0.92	0.93	0.94	0.95	0.77	0.86	0.85	0.83	0.84	0.84
65+	0.97	0.97	0.97	0.98	0.99	1.00	0.90	0.82	0.83	0.85	0.90
15+	0.89	0.88	0.90	0.91	0.93	0.74	0.78	0.80	0.81	0.78	0.85
0+	0.89	0.88	0.89	0.91	0.93	0.73	0.77	0.82	0.81	0.79	0.84

Source: U.S. Census Bureau, Current Population Survey, June 2018.

[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.

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

Microdata files from previous years reflect the latest available census-based controls. Although the most recent change in population controls had relatively little impact on summary measures such as averages, medians, and percentage distributions, it did have a significant impact on levels. For example, use of 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.

These differences could be disproportionately greater for certain population subgroups than for the total population.

Users should also exercise caution because of changes caused by the phase-in of the Census 2010 files (see "Basic CPS").[4] 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.

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 U.S. Census Bureau (2006) and Brooks & Bailar (1978).

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.

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

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

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

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

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

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

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

In this source and accuracy statement, Tables 3 through 8 provide illustrations for calculating standard errors. Table 9 provides the GVF parameters for labor force estimates, and Tables 10 and 11 provides GVF parameters for characteristics from the June 2018 supplement. Tables 12 and 13 provide factors and population controls to derive state and regional parameters.

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

The GVF parameters to use in computing standard errors are dependent upon the race/ethnicity group of interest. The following table summarizes the relationship between the race/ethnicity group of interest and the GVF parameters to use in standard error calculations.

Table 2. 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
White alone, White alone or in combination (AOIC), or White non-Hispanic population	Total or White
Black alone, Black AOIC, or Black non-Hispanic population	Black
Asian alone, Asian AOIC, or Asian non-Hispanic population	Asian, American Indian and Alaska Native (AIAN), Native Hawaiian and Other Pacific Islander (NHOPI)
AIAN alone, AIAN AOIC, or AIAN non-Hispanic population	Asian, AIAN, NHOPI
NHOPI alone, 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

Source: U.S. Census Bureau, Current Population Survey, internal data files.
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.

When calculating standard errors for an estimate of interest from cross-tabulations involving different characteristics, use the set of GVF parameters for the characteristic that will give the largest standard error. If the estimate of interest is strictly from basic CPS data, the GVF parameters will come from the CPS GVF table (Table 9). If the estimate is using Fertility supplement data, the GVF parameters will come from the Fertility supplement GVF table (Table 10).

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

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

Here *x* is the size of the estimate, and *a* and *b* are the parameters in Table 9 or 10 associated with the particular type of characteristic.

<u>Illustration 1</u>
Suppose there were 2,210,000 unemployed women of ages 15 to 44 in the civilian labor force. Use the appropriate parameters from Table 9 and Formula (1) to get

Table 3. Illustration of Standard Errors of Estimated Numbers

Number of unemployed males in the civilian labor force (x)	2,210,000
a-parameter (a)	-0.000028
b-parameter (b)	2,788
Standard error	78,000
90-percent confidence interval	2,082,000 to 2,338,000

Source: U.S. Census Bureau, Current Population Survey, Fertility Supplement, June 2018.

The standard error is calculated as

$$s_x = \sqrt{-0.000028 \times 2{,}210{,}000^2 + 2{,}788 \times 2{,}210{,}000},$$

which, rounded to the nearest thousand, is 78,000. The 90-percent confidence interval is calculated as 2,210,000 ± 1.645 × 78,000.

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.

<u>Standard Errors of Estimated Percentages</u>. 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 9 or 10 as indicated by the numerator.

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

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

Here y is the total number of people, families, households, or unrelated individuals in the base or denominator of the percentage, p is the percentage $100*x/y$ $(0 \le p \le 100)$, and b is the parameter in Table 9 or 10 associated with the characteristic in the numerator of the percentage.

Illustration 2
Suppose that 29.0 percent of the 63,890,000 women 15 to 44 years old were married when the first child was born. Use the appropriate parameter from Table 10 and Formula (2) to get

Table 4. Illustration of Standard Errors of Estimated Percentages

Percentage of women aged 15-44 who were married when the first child was born (p)	29.0
Base (y)	63,890,000
b-parameter (b)	5,564
Standard error	0.42
90-percent confidence interval	28.3 to 29.7

Source: U.S. Census Bureau, Current Population Survey, Fertility Supplement, June 2018.

The standard error is calculated as

$$s_{y,p} = \sqrt{\frac{5{,}564}{63{,}890{,}000} \times 29.0 \times (100.0 - 29.0)} = 0.42$$

The 90-percent confidence interval for the estimated percentage of women aged 15 to 44 who were married when the first child was born is from 28.3 to 29.7 percent (i.e., 29.0 ± 1.645 × 0.42).

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

$$s_{x_1-x_2} = \sqrt{{s_{x_1}}^2 + {s_{x_2}}^2} \tag{3}$$

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. This will result in accurate estimates of the standard error 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

Suppose that of the 6,592,000 women in 2018 between 20-29 years of age who were ever married, 68.0 percent were in the labor force, and of the 6,597,000 women in 2016 between 20-29 years of age who were ever married, 66.0 percent were in the labor force. Use the appropriate parameters from Table 9 and Formulas (2) and (3) to get

Table 5. Illustration of Standard Errors of Estimated Differences

	2018 (x_1)	2016 (x_2)	Difference
Percentage women aged 20-29 ever married in the labor force (p)	68.0	66.0	2.0
Base (y)	6,592,000	6,597,000	-
b-parameter (b)	2,788	2,788	-
Standard error	0.96	0.97	1.36
90-percent confidence interval	66.4 to 69.6	64.4 to 67.6	-0.2 to 4.2

Source: U.S. Census Bureau, Current Population Survey, Fertility Supplement, June 2018.

The standard error of the difference is calculated as

$$s_{x_1-x_2} = \sqrt{0.96^2 + 0.97^2} = 1.36$$

The 90-percent confidence interval around the difference is calculated as 2.0 ± 1.645 × 1.36. Since this interval does include zero, we cannot conclude with 90 percent confidence that the percentage of women in 2018 between 20-29 years of age who were ever married, in the labor force, is different than the percentage of women in 2016 between 20-29 years of age who were ever married, in the labor force.

Standard Errors of Ratios. Certain estimates may be calculated as the ratio of two numbers. The standard error of a ratio, x/y, may be computed using

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

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 persons 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 mean number of children per family with children.

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

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

Illustration 4

Suppose there were 30,539,000 ever-married women 15-44 years old and 33,351,000 never-married women 15-44 years old. The ratio of ever-married women, x, to never-married women, y, is 0.92. Use the appropriate parameters from Table 10 and Formulas (1) and (4) to get

Table 6. Illustration of Standard Errors of Ratios

	Ever-married (x)	Never-married (y)	Ratio
Women 15-44	30,539,000	33,351,000	0.92
a parameter (a)	-0.000021	-0.000021	-
b parameter (b)	5,564	5,564	-
Standard error	388,000	403,000	0.016
90-percent confidence interval	29,901,000 to 31,177,000	32,688,000 to 34,014,000	0.89 to 0.95

Source: U.S. Census Bureau, Current Population Survey, Fertility Supplement, June 2018.

Using Formula (4) with $r = 0$, the estimate of the standard error is

$$s_{x/y} = \frac{30{,}539{,}000}{33{,}351{,}000}\sqrt{\left[\frac{388{,}000}{30{,}539{,}000}\right]^2 + \left[\frac{403{,}000}{33{,}351{,}000}\right]^2} = 0.016$$

The 90-percent confidence interval is calculated as 0.92 ± 1.645 × 0.016.

<u>Standard Errors of Fertility Ratios</u>. The standard error of a fertility ratio is a function of the number of children ever born per 1,000 women and the number of women in a given category. The formula for the standard error of a fertility ratio is

$$s_{x,y} = x\sqrt{a + \frac{b}{xy} + \frac{c}{1{,}000y}} \qquad (5)$$

where a, b, and c are the parameters from Table 11, x is the number of children ever born or expected per 1,000 women, and y is the number of women in thousands. This formula should be used when calculating standard errors for estimates involving the possibility of more than one event per women, i.e., number of children ever born. For data involving at most one event per woman, convert the ratio to a percentage and use Formula (2) and the parameters in Table 9 or 10 to calculate the standard errors.

Illustration 5
Suppose that 8,411,000 women 40-44 years old had 1,994 children ever born per 1,000 women. Use Formula (5) and the parameters in Table 11 to get

Table 7. Illustration of Standard Errors of Fertility Ratios

Children ever born (x)	1,994
Base (y) in Thousands	8,411
a parameter (a)	0.0000015
b parameter (b)	961
c parameter (c)	1,756
Standard error	33
90-percent confidence interval	1,940 to 2,048

Source: U.S. Census Bureau, Current Population Survey, Fertility Supplement, June 2018.

The standard error is calculated as

$$S_{x,y} = 1{,}994\sqrt{0.0000015 + \frac{961}{1{,}994 \times 8{,}411} + \frac{1{,}756}{1{,}000 \times 8{,}411}} = 33$$

The 90-percent confidence interval is from 1,940 to 2,048 children ever born per 1,000 women (i.e., 1,994 ± 1.645 × 33). 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 Quarterly or Yearly Averages. For information on calculating standard errors for labor force data from the CPS which involve quarterly or yearly averages, please see Bureau of Labor Statistics (2006).

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 was 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.

Standard Errors of 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 9 or 10 by the state factor from Table 12. To determine a new state-level a-parameter (a_{state}), use the following:

(1) If the a-parameter from Table 9 or 10 is positive, multiply it by the state factor from Table 12.

(2) If the a-parameter in Table 9 or 10 is negative, calculate the new state-level a-parameter as follows:

$$a_{state} = \frac{-b_{state}}{POP_{state}} \quad (6)$$

where POP_{state} is the state population found in Table 12.

To determine state-level parameters for the fertility ratio parameters found in Table 11, multiply all parameters by the state factor from Table 12.

Note: The Census Bureau recommends the use of 3-year averages to compare estimates across states and 2-year averages to evaluate changes in state estimates over time.

Standard Errors of Regional Estimates. To compute standard errors for regional estimates, follow the steps for computing standard errors for state estimates found in "Standard Errors of State Estimates" using the regional factors found in Table 13.

Illustration 6

Suppose that of 24,317,000 women 15-44 years old in the South, 47.5 percent remain childless. Use Formula (2) and the appropriate parameter and factor from Tables 10 and 13 to get:

Table 8. Illustration of Standard Errors of Regional Estimates

Percent of childless women in South (p)	47.5
Base (x)	24,317,000
b parameter (b)	4,364
South regional factor	1.11
Regional b parameter (b_{region})	4,844
Standard error	0.70
90-percent confidence interval	46.3 to 48.7

Source: U.S. Census Bureau, Current Population Survey, Fertility Supplement, June 2018.

Obtain the region-level b parameter by multiplying the b parameter in Table 10 by the regional factor in Table 13. This gives b_{region} = 4,364 × 1.11 = 4,844. The standard error of the estimate of the percentage of women 15-44 years old in the South who are childless can then be found by using Formula (2) and the new region-level b parameter. The standard error is calculated as

$$s_{x,y} = \sqrt{\frac{4{,}844}{24{,}317{,}000} \times 47.5(100 - 47.5)} = 0.70$$

and the 90-percent confidence interval for the percentage of women 15-44 years old in the South who are childless is calculated as 47.5 ± 1.645 × 0.70.

<u>Standard Errors of Groups of States</u>. 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:

$$state\ group\ factor = \frac{\sum_{i=1}^{n} (POP_i \times state\ factor_i)}{\sum_{i=1}^{n} POP_i} \qquad (7)$$

where POP_i and $state\ factor_i$ are the population and factor for state *i* from Table 22. To obtain a new state group b-parameter ($b_{state\ group}$), multiply the b-parameter from Table 18 or 19 by the state factor obtained by Formula (7). To determine a new state group a-parameter ($a_{state\ group}$), use the following:

(1) If the a-parameter from Table 9 or 10 is positive, multiply it by the state group factor determined by Formula (7).

(2) If the a-parameter from Table 9 or 10 is negative, calculate the new state group *a* parameter as follows:

$$a_{state\ group} = \frac{-b_{state\ group}}{\sum_{i=1}^{n} POP_i} \qquad (8)$$

To determine state group-level parameters for the fertility ratio parameters found in Table 11, multiply all parameters by the state group factor calculated by Formula (7).

<u>Illustration 7</u>
Suppose the state group factor for the state group Illinois-Indiana-Michigan was required. The appropriate factor would be

$$state\ group\ factor = \frac{12{,}580{,}894 \times 1.16 + 6{,}598{,}681 \times 1.14 + 9{,}879{,}776 \times 1.15}{12{,}580{,}894 + 6{,}598{,}681 + 9{,}879{,}776} = 1.15$$

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

Table 9. Parameters for Computation of Standard Errors for Labor Force Characteristics: June 2018

Characteristic	*a*	*b*
Total or White		
Civilian labor force, employed	-0.000013	2,481
Unemployed	-0.000017	3,244
Not in labor force	-0.000013	2,432
Civilian labor force, employed, not in labor force, and unemployed		
Men	-0.000031	2,947
Women	-0.000028	2,788
Both sexes, 16 to 19 years	-0.000261	3,244
Black		
Civilian labor force, employed, not in labor force, and unemployed		
Total	-0.000117	3,601
Men	-0.000249	3,465
Women	-0.000191	3,191
Both sexes, 16 to 19 years	-0.001425	3,601
Asian, American Indian and Alaska Native (AIAN), Native Hawaiian and Other Pacific Islander (NHOPI)		
Civilian labor force, employed, not in labor force, and unemployed		
Total	-0.000245	3,311
Men	-0.000537	3,397
Women	-0.000399	2,874
Both sexes, 16 to 19 years	-0.004078	3,311
Hispanic, may be of any race		
Civilian labor force, employed, not in labor force, and unemployed		
Total	-0.000087	3,316
Men	-0.000172	3,276
Women	-0.000158	3,001
Both sexes, 16 to 19 years	-0.000909	3,316

Source: U.S. Census Bureau, Internal Current Population Survey data files for the 2010 Design.

Notes: These parameters are to be applied to basic CPS monthly labor force estimates. The Total or White, Black, and Asian, AIAN, NHOPI parameters are to be used for both alone and in combination race group estimates. 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, the a- and b-parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Black, Hispanic, and Asian, AIAN, NHOPI parameters. For the groups self-classified as having two or more races, use the Asian, AIAN, NHOPI parameters for all employment characteristics.

Table 10. Parameters for Computation of Standard Errors for Fertility and Birth Expectation Characteristics: June 2018

Characteristic	Persons		Households, etc.	
	a	b	a	b
FERTILITY				
Total or White	-0.000037	2,393	(X)	(X)
Black	-0.000137	2,393	(X)	(X)
Hispanic	-0.000252	4,032	(X)	(X)
Asian, AIAN, NHOPI and two or more races	-0.000305	2,393	(X)	(X)
NUMBER OF BIRTHS				
Total or White	-0.000068	4,364	(X)	(X)
Black	-0.000250	4,356	(X)	(X)
Hispanic	-0.000459	7,343	(X)	(X)
Asian, AIAN, NHOPI and two or more races	-0.000554	4,356	(X)	(X)
MARITAL STATUS, HOUSEHOLD & FAMILY CHARACTERISTICS				
Total or White	-0.000021	5,564	-0.000008	2,208
Black	-0.000135	7,992	-0.000034	1,998
Hispanic	-0.000275	13,469	-0.000069	3,366
Asian, AIAN, NHOPI and two or more races	-0.000310	7,992	-0.000078	1,998
INCOME				
Total or White	-0.000010	2,620	-0.000009	2,393
Black	-0.000051	3,000	-0.000044	2,613
Hispanic	-0.000103	5,056	-0.000090	4,403
Asian, AIAN, NHOPI and two or more races	-0.000116	3,000	-0.000101	2,613
EDUCATIONAL ATTAINMENT				
Total or White	-0.000010	2,530	-0.000008	2,208
Black and two or more races	-0.000048	2,861	-0.000034	1,998
Hispanic	-0.000067	3,258	-0.000069	3,366
Asian, AIAN, NHOPI	-0.000111	2,861	-0.000078	1,998
NATIVITY – Born in:				
Mexico, other N. America, S. America	-0.000037	11,801	(X)	(X)
Europe	-0.000021	6,780	(X)	(X)
Asia, Africa, Oceania	-0.000035	11,051	(X)	(X)
United States	-0.000019	5,932	(X)	(X)

Source: U.S. Census Bureau, Current Population Survey, Internal data from the Fertility Supplement, June 2018.

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

[B] Hispanics may be any race.

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

Table 11. Parameters for Computation of Standard Errors for Fertility Ratios: June 2018

a	b	c
0.0000015	961	1,756

Source: U.S. Census Bureau, Current Population Survey, Internal data from the Fertility Supplement, June 2018.
Note: Multiply the parameters by 1.3 to get foreign-born parameters.

Table 12. Factors and Populations for State Parameters and Standard Errors: June 2018

State	Factor	Population	State	Factor	Population
Alabama	1.13	4,807,676	Montana	0.22	1,046,012
Alaska	0.18	710,126	Nebraska	0.51	1,901,477
Arizona	1.16	7,010,312	Nevada	0.72	3,013,211
Arkansas	0.73	2,963,855	New Hampshire	0.35	1,333,075
California	1.16	39,226,312	New Jersey	1.15	8,927,077
Colorado	1.17	5,582,566	New Mexico	0.44	2,053,490
Connecticut	0.88	3,538,202	New York	1.19	19,617,438
Delaware	0.23	955,498	North Carolina	1.18	10,171,904
District of Columbia	0.18	692,428	North Dakota	0.18	737,904
Florida	1.12	20,968,512	Ohio	1.15	11,512,634
Georgia	1.16	10,331,291	Oklahoma	1.07	3,858,271
Hawaii	0.33	1,368,587	Oregon	1.06	4,155,617
Idaho	0.40	1,730,355	Pennsylvania	1.16	12,619,954
Illinois	1.16	12,580,894	Rhode Island	0.28	1,045,930
Indiana	1.14	6,598,681	South Carolina	1.12	4,986,684
Iowa	0.78	3,115,966	South Dakota	0.23	858,971
Kansas	0.81	2,855,257	Tennessee	1.14	6,671,545
Kentucky	1.16	4,384,773	Texas	1.17	28,178,952
Louisiana	1.06	4,576,687	Utah	0.51	3,127,383
Maine	0.42	1,326,879	Vermont	0.20	617,898
Maryland	1.19	5,979,687	Virginia	1.19	8,312,350
Massachusetts	1.13	6,817,355	Washington	1.17	7,410,192
Michigan	1.15	9,879,776	West Virginia	0.50	1,775,910
Minnesota	1.16	5,566,518	Wisconsin	1.16	5,743,953
Mississippi	0.71	2,916,507	Wyoming	0.16	564,427
Missouri	1.18	6,023,367			

Source: U.S. Census Bureau, Current Population Survey, Internal data from the Fertility Supplement, June 2018.

Notes: These factors are for use with state-level fertility estimates for subpopulation groups. The state population counts in this table are for the 0+ population. For foreign-born and noncitizen characteristics for Total and White, the a- and b-parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks, Asians, American Indian and Alaska Natives, Native Hawaiian and Other Pacific Islanders, and Hispanics.

Table 13. Factors and Populations for Census Region Parameters and Standard Errors: June 2018

Region	Factor	Population
Northeast	1.08	67,375,398
Midwest	1.09	55,843,808
South	1.11	122,532,530
West	1.03	76,998,590
All Except South	1.06	200,217,796

Source: U.S. Census Bureau, Current Population Survey, Internal data from the Fertility Supplement, June 2018.

Notes: These factors are for use with census region-level fertility estimates for subpopulation groups. The census region population counts in this table are for the 0+ population. For foreign-born and noncitizen characteristics for Total and White, the a- and b-parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks, Asians, American Indian and Alaska Natives, Native Hawaiian and Other Pacific Islanders, and Hispanics.

REFERENCES

Brooks, C.A., & Bailar, B.A. 1978. *Statistical Policy Working Paper 3 - An Error Profile: Employment as Measured by the Current Population Survey*. Subcommittee on Nonsampling Errors, Federal Committee on Statistical Methodology, U.S. Department of Commerce, Washington, DC. https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/spwp3.pdf

Bureau of Labor Statistics, February 2006, "Household Data ("A" tables, monthly; "D" tables, quarterly)." https://www.bls.gov/cps/eetech_methods.pdf

Bureau of Labor Statistics, April 2014, "Redesign of the Sample for the Current Population Survey." http://www.bls.gov/cps/sample_redesign_2014.pdf

U.S. Census Bureau. 2006. *Current Population Survey: Design and Methodology*. Technical Paper 66. Washington, DC: Government Printing Office. http://www.census.gov/prod/2006pubs/tp-66.pdf

All online references accessed April 15, 2019.