Attachment C

Source of the Data and Accuracy of the Estimates for the October 2020 Current Population Survey Microdata File on School Enrollment

Table of Contents

Tables

Source of the Data and Accuracy of the Estimates for the October 2020 Current Population Survey Microdata File on School Enrollment

SOURCE OF THE DATA

The data in this microdata file are from the October 2020 Current Population Survey (CPS). The U.S. Census Bureau conducts the CPS every month, although this file has only October data. The October 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 National Center for Educational Statistics jointly sponsor the supplemental questions for October.

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 universe, consists primarily of the population in correctional institutions and nursing homes (98 percent of the 4.0 million institutionalized people in the 2010 Census). Starting in August 2017, college and university dormitories were also excluded from the universe because most of the residents had usual residences elsewhere. Interviewers ask questions concerning labor force participation of each member 15 years old and older 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.

---

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.

Approximately 69,000 sampled addresses were selected from the sampling frame in October. Based on eligibility criteria, six percent of these sampled addresses were sent directly to computer-assisted telephone interviewing (CATI). The remaining sampled addresses were assigned to interviewers for computer-assisted personal interviewing (CAPI).[3] Of all addresses in sample, about 59,000 were determined to be eligible for interview. Interviewers obtained interviews at about 47,000 of the housing units at these addresses. Noninterviews occur when the occupants are not found at home after repeated calls or are unavailable for some other reason.

October 2020 Supplement. In October 2020, in addition to the basic CPS questions, interviewers asked supplementary questions of household members three years old and older on school enrollment.

Estimation Procedure. This survey's estimation procedure adjusts weighted sample results to agree with independently derived population controls of the civilian noninstitutionalized population of the United States, each state, and the District of Columbia. These population controls[4] are prepared monthly as part of the Census Bureau's Population Estimates 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).
- 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

[3] For further information on CATI and CAPI and the eligibility criteria, please see U.S. Census Bureau (2019).

[4] For additional information on population controls, including details on the demographic characteristics used and net international components, please see Chapters 1-3 and Appendix: History of the Current Population Survey of U.S. Census Bureau (2019).

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 controls.

The net international migration component of the population controls 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 behind the survey date, it is necessary to make short-term projections of these components to develop the estimate for the survey date.

ACCURACY OF THE ESTIMATES

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

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

- Measurement error: 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.

- Coverage error: Some individuals who should have been included in the survey frame were missed.
- Nonresponse error: Responses are not collected from all those in the sample or the respondent is unwilling to provide information.
- Imputation error: Values are estimated imprecisely for missing data.
- Processing error: Forms may be lost, data may be incorrectly keyed, coded, or recoded, etc.

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 October 2020 basic CPS, the household-level unweighted nonresponse rate was 19.7 percent. The person-level unweighted nonresponse rate for the School Enrollment supplement was an additional 11.0 percent.

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

In accordance with Census Bureau and Office of Management and Budget Quality Standards, the Census Bureau will conduct an analysis to assess nonresponse bias in the School Enrollment.

Responses are made up of complete interviews and sufficient partial interviews. 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 answer in order to be considered a sufficient partial interview.

As a result of sufficient partial interviews being considered responses, individual items/questions have their own response and refusal rates. As part of the nonsampling error analysis, the item response rates, item refusal rates, and edits are reviewed. For the School Enrollment supplement, the unweighted item refusal rates range from 0.25 percent to 4.86 percent. The unweighted item allocation rates range from 4.64 percent to 30.68 percent.

Undercoverage. The concept of coverage with a survey sampling process is defined as 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 October 2020 is estimated to be about ten 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 mitigates 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 October 2020 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: October 2020

Age group	Total			White alone		Black alone		Residual race[A]		Hispanic[B]	
	All people	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-15	0.86	0.87	0.86	0.89	0.89	0.81	0.78	0.80	0.79	0.83	0.82
16-19	0.89	0.92	0.87	0.95	0.89	0.85	0.80	0.80	0.86	0.87	0.81
20-24	0.79	0.82	0.76	0.84	0.78	0.74	0.67	0.80	0.75	0.80	0.75
25-34	0.82	0.80	0.83	0.84	0.87	0.65	0.70	0.75	0.79	0.74	0.82
35-44	0.90	0.88	0.92	0.92	0.95	0.72	0.83	0.84	0.87	0.82	0.88
45-54	0.90	0.90	0.90	0.92	0.93	0.82	0.80	0.87	0.84	0.83	0.86
55-64	0.96	0.95	0.97	0.97	0.98	0.88	0.89	0.88	0.91	0.85	0.90
65+	1.01	1.01	1.02	1.03	1.03	0.96	0.98	0.91	0.90	0.93	0.88
15+	0.91	0.90	0.92	0.93	0.94	0.79	0.82	0.83	0.85	0.82	0.85
0+	0.90	0.90	0.91	0.92	0.93	0.79	0.81	0.82	0.83	0.82	0.84

Source: U.S. Census Bureau, Current Population Survey, October 2020.

[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 is due to differences in interviewer training and experience and in differing survey processes.[5] These differences are examples 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 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 populations 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 2010 Census files (see "Basic CPS").[6] 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 estimates: (1) the temporary disruption of the rotation pattern from

---

[5] Survey processes include, but are not limited to, question wording, universe, sampling frame, interview modes, and weighting.

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

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 (2019) and Brooks & Bailar (1978).

Standard Errors and Their Use. A 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 the specified confidence that the interval includes the average estimate calculated from all possible samples. Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common type of hypothesis is that the population parameters are different. An example of this would be comparing the percentage of men who were part-time workers to the percentage of women who were part-time workers.

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

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

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

There are two ways to calculate standard errors for the CPS microdata file on School Enrollment.

1. Direct estimates created from replicate weighting methods;
2. Generalized variance estimates created from generalized variance function (GVF) parameters *a* and *b*.

While replicate weighting methods provide the most accurate variance estimates, this approach requires more computing resources and more expertise on the part of the user. The GVF parameters provide a method of balancing accuracy with resource usage as well as a smoothing effect on standard error estimates. For more information on calculating direct estimates, see U.S. Census Bureau (2009). For more information on GVF estimates, refer to the "Generalized Variance Parameters" section.

Generalized Variance Parameters. While it is possible to estimate 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 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 and 5 through 7 provide illustrations for calculating standard errors;
- Table 4 provides the October 2020 population totals for school enrollment groups;
- Table 8 provides the GVF parameters for labor force estimates; and
- Table 9 provides GVF parameters for characteristics from the October 2020 supplement.

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. Table 2 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 8). If the estimate is using School Enrollment supplement data, the GVF parameters will come from the School Enrollment supplement GVF table (Table 9).

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

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

Here x is the size of the estimate, and a and b are the parameters in Table 8 or 9 associated with the particular type of characteristic.

Illustration 1
Suppose there were 5,670,000 unemployed men (ages 16 and up) in the civilian labor force. Table 3 shows how to use the appropriate parameters from Table 8 and Formula (1) to estimate the standard error and confidence interval.

Table 3. Illustration of Standard Errors of Estimated Numbers

Number of unemployed males in the civilian labor force (x)	5,670,000
a-parameter (a)	-0.000031
b-parameter (b)	2,947
Standard error	125,000
90-percent confidence interval	5,464,000 to 5,876,000

Source: U.S. Census Bureau, Current Population Survey, School Enrollment, October 2020.

The standard error is calculated as

$$s_x = \sqrt{-0.000031 \times 5{,}670{,}000^2 + 2{,}947 \times 5{,}670{,}000},$$

which, rounded to the nearest thousand, is 125,000. The 90-percent confidence interval is calculated as 5,670,000 ± 1.645 × 125,000.

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

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

$$s_x = \sqrt{-\left(\frac{b}{T}\right)x^2 + bx} \qquad (2)$$

Here x is the size of the estimate, T is the population control from Table 4 for the age group of interest, and b is the parameter in Table 9 associated with the particular type of characteristic. If Table 4 does not explicitly contain the age group of interest, obtain T by summing the controls for the age groups available in the table that do contain the age group of interest. When calculating standard errors for numbers from cross-tabulations involving different characteristics, use the set of parameters for the characteristic that will give the largest standard error.

Table 4. Population Controls for School Enrollment Age Groups: October 2020

Age Group	Total	White Only	Black Only	Asian, AIAN, NHOPI[A]	Hispanic[B]
3-4	7,908,848	5,666,372	1,197,485	1,044,991	2,065,622
5	4,046,816	2,895,841	611,698	539,277	1,055,056
6	4,054,088	2,902,665	613,098	538,325	1,047,101
7-11	20,259,943	14,558,111	3,066,351	2,635,481	5,283,393
12-13	8,388,898	6,040,875	1,277,916	1,070,107	2,175,276
14	4,179,805	3,032,184	626,830	520,791	1,072,815
15-17	12,442,925	9,104,574	1,809,201	1,529,150	3,093,158
18	4,117,765	3,021,530	601,440	494,795	1,000,858
19	4,150,618	3,041,331	609,482	499,805	993,017
20-24	20,916,678	15,323,245	3,078,214	2,515,219	4,813,548
25-29	22,473,470	16,280,141	3,462,493	2,730,836	4,848,668
30-34	22,393,956	16,398,428	3,264,850	2,730,678	4,597,506
35-64	123,473,323	95,433,260	15,904,915	12,135,148	21,241,193
65+	55,116,473	46,040,122	5,334,800	3,741,551	4,939,893

Source: U.S. Census Bureau, Population Estimates, October 2020.

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

[B] Hispanics may be any race.

Notes: White, Black, and Asian, AIAN, NHOPI totals include Hispanics. The Asian, AIAN, NHOPI parameters are to be used for Asian, AIAN, NHOPI alone and all race in-combination group estimates.

Illustration 2

Suppose there were 3,189,000 three- and four-year-olds enrolled in school and 7,908,848 total children in that age group. Table 5 shows how to use the appropriate *b* parameter from Table 9 and Formula (2) to estimate the standard error and confidence interval.

Table 5. Illustration of Standard Errors of School Enrollment Numbers

Number of three- and four-year-olds enrolled in school (x)	3,189,000
Total (*T*)	7,908,848
b-parameter (*b*)	4,256
Standard error	90,000
90-percent confidence interval	3,041,000 to 3,337,000

Source: U.S. Census Bureau, Current Population Survey, School Enrollment Supplement, October 2020.

The standard error is calculated as

$$s_x = \sqrt{-\left(\frac{4{,}256}{7{,}908{,}848}\right) \times 3{,}189{,}000^2 + 4{,}256 \times 3{,}189{,}000}$$

which, rounded to the nearest thousand, is 90,000. The 90-percent confidence interval is calculated as 3,189,000 ± 1.645 × 90,000.
A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

Standard Errors of Estimated Percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on both the size of the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories, use the parameter from Table 8 or 9 as indicated by the numerator.

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

$$s_{y,p} = \sqrt{\frac{b}{y} p(100 - p)} \qquad (3)$$

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 8 or 9 associated with the characteristic in the numerator of the percentage.

Illustration 3
Suppose there were 16,995,000 people aged 18 to 21, and 50.0 percent were enrolled in college. Table 6 shows how to use the appropriate parameters from Table 9 and Formula (3) to estimate the standard error and confidence interval.

Table 6. Illustration of Standard Errors of Estimated Percentages

Percentage of people aged 18-21 enrolled in college (p)	50.0
Base (y)	16,995,000
b-parameter (b)	4,256
Standard error	0.79
90-percent confidence interval	48.7 to 51.3

Source: U.S. Census Bureau, Current Population Survey, School Enrollment, October 2020.

The standard error is calculated as

$$s_{y,p} = \sqrt{\frac{4{,}256}{16{,}995{,}000} \times 50.0 \times (100.0 - 50.0)} = 0.79$$

and the 90-percent confidence interval for the estimated percentage of people aged 18 to 21 enrolled in college is from 48.7 to 51.3 percent (i.e., 50.0 ± 1.645 × 0.79).

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

$$s_{x_1 - x_2} = \sqrt{\left(s_{x_1}\right)^2 + \left(s_{x_2}\right)^2} \qquad (4)$$

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 4

Suppose that of the 6,720,000 employed men between 20-24 years of age, 28.0 percent were part-time workers, and of the 6,505,000 employed women between 20-24 years of age, 40.3 percent were part-time workers. Table 7 shows how to use the appropriate parameters from Table 8 and Formulas (3) and (4) to estimate the standard error and confidence interval.

Table 7. Illustration of Standard Errors of Estimated Differences

	Men (x_1)	Women (x_2)	Difference
Percentage working part-time (p)	28.0	40.3	12.3
Base (y)	6,720,000	6,505,000	-
b-parameter (b)	2,947	2,788	-
Standard error	0.94	1.02	1.39
90-percent confidence interval	26.5 to 29.5	38.6 to 42.0	10.0 to 14.6

Source: U.S. Census Bureau, Current Population Survey, School Enrollment, October 2020.

The standard error of the difference is calculated as

$$s_{x_1-x_2} = \sqrt{0.94^2 + 1.02^2} = 1.39$$

and the 90-percent confidence interval around the difference is calculated as 12.3 ± 1.645 × 1.39. Since this interval does not include zero, we can conclude with 90-percent confidence that the percentage of part-time women workers between 20-24 years of age is greater than the percentage of part-time men workers between 20-24 years of age.

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).

Year-to-Year Factors.
In past years, the Census Bureau published a table of year factors for the School Enrollment Supplement in the Source and Accuracy Statement. User demand for these factors has diminished with the introduction of replicate weights. Data users producing estimates from prior years should consult the Source and Accuracy Statements covering the years of their analysis to estimate standard errors.

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 8. Parameters for Computation of Standard Errors for Labor Force Characteristics: October 2020

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 9. Parameters for Computation of Standard Errors for School Enrollment Characteristics: October 2020

Characteristics	*b*				
	Total	White	Black	Asian, AIAN, NHOPI[A]	Hispanic[B]
PEOPLE					
Total	3,995	3,638	4,278	5,260	5,495
Male	2,732	2,591	2,707	3,468	4,065
Female	2,684	2,930	3,051	3,629	3,601
Age					
3 to 5	4,256	4,216	4,044	4,401	4,334
6 to 14	3,674	3,483	4,044	6,024	4,334
15 to 24	4,256	4,216	4,456	4,401	4,334
25 and over	3,674	3,483	4,044	6,024	5,220
Level of enrollment below college (ages 3 to 24)					
Total enrolled	3,615	3,615	4,459	4,631	4,644
Nursery School	4,144	4,142	4,459	3,953	4,644
Kindergarten	3,615	3,615	3,919	3,953	4,644
Elementary School	3,615	4,142	4,459	6,093	4,644
High School	3,615	3,615	3,919	3,505	4,644
Marital status, household and family characteristics					
Some household members	6,973	5,844	6,934	7,077	7,435
All household members	4,707	4,165	5,061	4,862	5,657
FAMILIES, HOUSEHOLDS, OR UNRELATED INDIVIDUALS					
Income, earnings	6,618	5,580	6,367	6,519	6,706
Marital status, household, and family characteristics, educational attainment, population by age/sex	5,045	4,516	4,786	5,352	5,487

Source: U.S. Census Bureau, Current Population Survey, Internal data from the School Enrollment, October 2020.

[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 School Enrollment 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.

DRB Clearance Number - CBDRB-FY21-154

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.

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. (2006). *Household Data ("A" tables, monthly; "D" tables, quarterly).* https://www.bls.gov/cps/eetech_methods.pdf

Bureau of Labor Statistics. (2014). *Redesign of the Sample for the Current Population Survey*. http://www.bls.gov/cps/sample_redesign_2014.pdf

U.S. Census Bureau. (2009). *Estimating ASEC Variances with Replicate Weights Part I: Instructions for Using the ASEC Public Use Replicate Weight File to Create ASEC Variance Estimates*. http://usa.ipums.org/usa/resources/repwt/Use_of_the_Public_Use_Replicate_Weight_File_final_PR.doc

U.S. Census Bureau. (2019). *Current Population Survey: Design and Methodology*. Technical Paper 77. Washington, DC: Government Printing Office. https://www2.census.gov/programs-surveys/cps/methodology/CPS-Tech-Paper-77.pdf

All online references accessed June 2, 2021.