

**Source of the Data and Accuracy of the Estimates for the
2016 Annual Social and Economic Supplement Split-panel Sample Microdata
File**

Table of Contents

SOURCE OF THE DATA.....	1
Basic CPS.....	1
The 2016 Annual Social and Economic Supplement Split-panel Sample	2
Estimation Procedure	2
ACCURACY OF THE ESTIMATES	3
Sampling Error.....	3
Nonsampling Error.....	3
Nonresponse.....	4
Sufficient Partial Interview	4
Undercoverage	4
Comparability of Data.....	5
A Nonsampling Error Warning.....	6
Standard Errors and Their Use.....	7
Estimating Standard Errors	7
Technical Assistance.....	7
REFERENCES.....	7

Tables

Table 1. 2016 ASEC Split-panel Sample Coverage Ratios	5
---	---

Source of the Data and Accuracy of the Estimates for the 2016 Annual Social and Economic Supplement Split-panel Sample Microdata File

SOURCE OF THE DATA

The data in this microdata file are from the 2016 Annual Social and Economic Supplement (ASEC) Split-panel Sample of the Current Population Survey (CPS). The U.S. Census Bureau conducted the 2016 ASEC Split-panel Sample data collection in the month of March of 2016. The 2016 ASEC Split-panel Sample uses two sets of questions, the basic CPS and the set of supplemental questions that were used in 2013 CPS ASEC, before the questionnaire redesign was implemented in the 2014 ASEC. 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 sponsored the 2016 ASEC Split-panel Sample.

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

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

Every ten years, the CPS first stage sample is redesigned¹ 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 2000 design, the United States was divided into 2,025 PSUs. These were then grouped into 824 strata and one PSU was selected for sample from each stratum. In the 2010 sample design, the United States was divided into 1,987 PSUs. These PSUs were then grouped into 852 strata. Within each stratum, a single PSU was chosen for the sample, with its probability of selection proportional to its population as of the most recent decennial census. In the case of strata consisting of only one PSU, the PSU was chosen with certainty.

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

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

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

design. Second, new PSUs selected for only the 2010 design gradually replace outgoing PSUs selected for only the 2000 design. By July 2015, the new 2010 sample design was completely implemented and the sample is entirely from the 2010 redesigned sample.

Approximately 74,000 housing units were selected for sample from the sampling frame for the basic CPS. Of all housing units in sample, about 61,100 were determined to be eligible for interview. Interviewers obtained interviews at about 52,000 of these units. Noninterviews occur when the occupants are not found at home after repeated calls or are unavailable for some other reason.

The 2016 Annual Social and Economic Supplement Split-panel Sample. In addition to the basic CPS questions, interviewers asked supplementary questions for the 2016 ASEC Split-panel Sample. They asked these questions of the civilian noninstitutional population and 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

These questions were the same questions asked in the 2013 ASEC before the 2014 questionnaire redesign.

The 2016 ASEC Split-panel Sample sampled housing units from the 74,000 basic CPS households. The 74,000 basic CPS households were ordered in the same sample selection order as the basic CPS and a systematic subsample of 5,000 housing units were identified to receive the Split-panel sample questionnaire. The remaining 69,000 CPS households received the new redesigned ASEC questionnaire. The cases that were identified in the 2016 ASEC sample as the Hispanic over-sample and CHIP samples were not eligible for this split-panel sample [2].

The sample for the 2016 ASEC Split-panel Sample contained 5,000 housing units. About 4,100 housing units were determined to be eligible for interview and about 3,500 interviews were obtained.

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 2016 ASEC Split-panel Sample, are prepared

monthly to agree with the most current set of population estimates that are released as part of the Census Bureau's population estimates and projections program.

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

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

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

The independent estimates by age, sex, race, and Hispanic origin, and for states by selected age groups and broad race categories, are developed using the basic demographic accounting formula whereby the population from the 2010 Decennial Census data is updated using data on the components of population change (births, deaths, and net international migration) with net internal migration as an additional component in the state population estimates.

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

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

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

ACCURACY OF THE ESTIMATES

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

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

Nonsampling Error. For a given estimator, the difference between the estimate that would result if the sample were to include the entire population and the true population value being

estimated is known as nonsampling error. There are several sources of nonsampling error that may occur during the development or execution of the survey. It can occur because of circumstances created by the interviewer, the respondent, the survey instrument, or the way the data are collected and processed. For example, errors could occur because:

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

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

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

Nonresponse. The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. For the cases eligible for the 2016 ASEC Split-panel Sample, the basic CPS household-level nonresponse rate was 14.6 percent. The household-level nonresponse rate for the ASEC was an additional 14.9 percent. These two nonresponse rates lead to a combined supplement nonresponse rate of 27.3 percent.

In accordance with Census Bureau and Office of Management and Budget Quality Standards, the Census Bureau will conduct a nonresponse bias analysis to assess nonresponse bias in the 2016 ASEC Split-panel.

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

Undercoverage. The concept of coverage in the survey sampling process is the extent to which the total population that could be selected for sample “covers” the survey’s target population. Missed housing units and missed people within sample households create undercoverage in the 2016 ASEC Split-panel Sample. Overall, 2016 ASEC Split-panel Sample undercoverage is estimated to be about 11 percent. Coverage varies with age, sex, and race. Generally, coverage

is larger for females than for males and larger for non-Blacks than for Blacks. This differential coverage is a general problem for most household-based surveys.

The 2016 ASEC Split-panel Sample 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 the 2016 ASEC Split-panel Sample coverage ratios by age and sex for certain race and Hispanic groups.

Age group	All people	Total		White only		Black only		Residual race		Hispanic	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-15	0.90	0.90	0.90	0.93	0.92	0.68	0.68	0.98	1.06	0.91	0.82
16-19	0.90	0.93	0.88	0.94	0.94	0.95	0.75	0.81	0.67	0.77	0.89
20-24	0.75	0.73	0.76	0.74	0.83	0.55	0.50	0.88	0.71	0.69	0.74
25-34	0.83	0.80	0.86	0.84	0.93	0.57	0.57	0.82	0.77	0.75	0.90
35-44	0.94	0.92	0.95	0.95	0.98	0.68	0.72	0.98	1.03	0.89	0.98
45-54	0.88	0.89	0.87	0.91	0.88	0.73	0.96	0.92	0.67	0.82	0.79
55-64	0.86	0.84	0.87	0.83	0.88	0.91	0.86	0.83	0.82	0.87	0.88
65+	0.99	0.96	1.01	0.98	1.03	0.86	0.98	0.82	0.88	0.94	0.86
15+	0.89	0.87	0.90	0.89	0.93	0.73	0.77	0.88	0.81	0.81	0.88
0+	0.89	0.88	0.90	0.90	0.93	0.72	0.75	0.90	0.87	0.84	0.86

Notes: (1) The Residual race group includes cases indicating a single race other than White or Black and cases indicating two or more races.
 (2) Hispanics may be any race.

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 use caution when comparing the data from this file, the 2016 ASEC Split-panel Sample, to the 2015 CPS ASEC because of the questionnaire differences, sample design differences, and data collection differences.

Data users should be careful when comparing estimates for 2016 ASEC Split-panel Sample (which reflect Census 2010-based controls) with estimates for 1999 to 2010 (from March 2000 CPS to March 2011 CPS), which reflect 2000 census-based controls, and to 1992 to 1998 (from March 1993 CPS to March 1999 CPS), which reflect 1990 census-based controls. Ideally, the same population controls should be used when comparing any estimates. In reality, the use of

the same population controls is not practical when comparing trend data over a period of 10 to 20 years. Thus, when it is necessary to combine or compare data based on different controls or different designs, data users should be aware that changes in weighting controls or weighting procedures could create small differences between estimates. See the following discussion for information on comparing estimates derived from different controls or different sample designs.

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

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

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

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

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

³ The phase-in process using the 2010 Census files began in April 2014.

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

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

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

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

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

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

Direct variance estimation methods should be used to calculate standard errors for the 2016 ASEC Split-panel Sample microdata file by using the replicate weights. For more information on calculating direct variance estimates, see reference [4].

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.

References

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

- [2] U.S. Census Bureau. 2006. *Current Population Survey: Design and Methodology*. Technical Paper 66. Washington, DC: Government Printing Office.
<http://www.census.gov/prod/2006pubs/tp-66.pdf>.
- [3] Brooks, C.A. and Bailar, B.A. 1978. *Statistical Policy Working Paper 3 - An Error Profile: Employment as Measured by the Current Population Survey*. Subcommittee on Nonsampling Errors, Federal Committee on Statistical Methodology, U.S. Department of Commerce, Washington, DC. <https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/spwp3.pdf>.
- [4] U.S. Census Bureau, July 15, 2009, “Estimating ASEC Variances with Replicate Weights Part I: Instructions for Using the ASEC Public Use Replicate Weight File to Create ASEC Variance Estimates.”
http://www.bls.census.gov/pub/cps/march/Use_of_the_Public_Use_Replicate_Weight_File_final_PR.doc.

All online resources accessed March 27, 2017