UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

Oct 15, 2014

MEMORANDUM FOR: Lisa Clement
Survey Director for CPS and Time Use
Associate Directorate for Demographic Programs

From: Ruth Ann Killion
Chief, Demographic Statistical Methods Division

Subject: Source and Accuracy Statement for the July 2012 CPS Microdata File on Public Participation in the Arts - Revision

Attached is the statement on the source of the data and accuracy of the estimates for the July 2012 CPS Microdata File on Public Participation in the Arts (PPA). This revision accounts for corrections to the eligible universe using corrected edits by the sponsor.

If you have any questions or need additional information, please contact David Hornick of the Demographic Statistical Methods Division via email at dsmd.source.and.accuracy@census.gov.

Attachment

cc:

G. Weyland (ADDP)
K. Linse
T. Hicks
J. Holland Ray
V. Ashton
B. Kominski (SEHSD)
W. Savino (ACSD)
D. Smallwood
J. Sweeney
C. Bennett
J. Farber (DSMD)
Y. Cheng
J. Scott
I. Irizarry-Pietri
T. Kennel
D. Hornick
S. Clark
M. Herbstritt
R. Hoop

Source of the Data and Accuracy of the Estimates for the July 2012 CPS Microdata File on Public Participation in the Arts

Table of Contents

Tables

Source of the Data and Accuracy of the Estimates for the July 2012 CPS Microdata File on Public Participation in the Arts

SOURCE OF THE DATA

The data in this microdata file are from the July 2012 Current Population Survey (CPS). The U.S. Census Bureau conducts the CPS every month, although this file has only July 2012 data. The July 2012 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 entire population. The National Endowment of the Arts sponsored the supplemental questions for July 2012.

Basic CPS. The monthly CPS collects primarily labor force data about the civilian noninstitutional 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.

In April 2004, the Census Bureau began phasing out the 1990 sample[1] and replacing it with the 2000 sample, creating a mixed sampling frame. Two simultaneous changes occurred during this phase-in period. First, primary sampling units (PSUs)[2] selected for only the 2000 design gradually replaced those selected for the 1990 design. This involved 10 percent of the sample. Second, for PSUs selected for both designs, sample households from the 2000 design gradually replaced sample households from the 1990 design. This involved about 90 percent of the sample. The new sample design was completely implemented by July 2005.

In the first stage of the sampling process, PSUs are selected for sample. The United States is divided into 2,025 PSUs. The PSUs were redefined for this design to correspond to the Office of Management and Budget definitions of Core-Based Statistical Area definitions and to improve efficiency in field operations. These PSUs are grouped into 824 strata. Within each stratum, a single PSU is chosen for the sample, with its probability of selection proportional to its population as of the most recent decennial census. This PSU represents the entire stratum from which it was selected. In the case of strata consisting of only one PSU, the PSU is chosen with certainty.

---

[1] For detailed information on the 1990 sample redesign, please see reference [1].

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

Approximately 72,000 housing units were selected for sample from the sampling frame in July 2012. Based on eligibility criteria, 11 percent of these housing units were sent directly to computer-assisted telephone interviewing (CATI). The remaining units were assigned to interviewers for computer-assisted personal interviewing (CAPI).[3] Of all housing units in sample, about 59,000 were determined to be eligible for interview. Interviewers obtained interviews at about 54,000 of these units. Noninterviews occur when the occupants are not found at home after repeated calls or are unavailable for some other reason.

July 2012 Supplement. In July 2012, in addition to the basic CPS questions, interviewers asked supplementary questions on public participation in the arts of two randomly selected household members aged 18 or older from about one-half the sampled CPS households. If the selected person had a spouse or partner then questions were also asked of their spouse/partner. The supplement contained questions about the sampled member's participation in various artistic activities from July 1, 2011 to July 1, 2012. It asked about the type of artistic activity, the frequency of participation, training and exposure, musical and artistic preferences, school-age socialization, and computer usage related to artistic information. These topics were separated into two core sets of questions and five modules. Interviews were conducted during the period of July 15-21, 2012.

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

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

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

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

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

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

---

[3] For further information on CATI and CAPI and the eligibility criteria, please see reference [2].

- 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.

PPA Estimation Procedure. The PPA adjusts weighted sample results to agree with the same independently derived population estimates of the civilian noninstitutional population of the United States as the CPS. However, the age groups were modified to include only those who are 18 years old or older.

The questionnaire modules and core questions were assigned to households so that half of the sample would receive each core and 40% would receive each module.

Each sampled person receives up to four of five produced weights for the PPA survey depending on the modules asked. There is a weight for each of Core 1 questions; Core 2 questions; Modules A1 and D; Module A2; and Modules B, C, and E.[4] The first weight, PWOWGT, should be used to create estimates from Core 1. The second, PWTWGT, should be used to create estimates from Core 2. Modules A1 and D should receive the third weight, PWSWGT. The fourth weight, PWAWGT, should be used to create estimates from Module A2. Modules B, C, and E should receive the fifth weight, PWNWGT. All weights were created using the same weighting procedure but different person selection factors.

To account for the assignment of modules to a portion of the respondents, the data user must apply a module factor to determine the final weight. The value of the factor is based on the analysis the data user is conducting. Table 1 provides the factors for each module or combination of modules (cross analysis of variables from two modules). These factors are determined by summing the proportion of cases that were asked the module or combination of modules of interest. The factor is the inverse of the proportion of cases receiving the module or combination of modules.

Table 1. Module Factors to Assign to Each Case in Analysis to Calculate the Final Weight

Core/Module Weight Used	Module Factor to Assign
PWOWGT, PWTWGT, or PWAWGT	1.000000
PWSWGT	2.25
PWNWGT	1.75

[4] Module A1 refers to question 1-4 in Module A, which were asked of respondent and spouse. Module A2 refers to question 5-8 in Module A, which were asked of respondent only.

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 that 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 July 2012 basic CPS, the household-level nonresponse rate was 9.3 percent. The person-level nonresponse rate for the Public Participation in the Arts supplement was an additional 25.2 percent.

Since the basic CPS nonresponse rate is a household-level rate and the Public Participation in the Arts 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 Public Participation in the Arts supplement.

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 July 2012 is estimated to be about 14 percent. CPS coverage varies with age, sex, and race. Generally, coverage is larger for females than for males and larger for non-Blacks than for Blacks. This differential coverage is a general problem for most household-based surveys.

The CPS weighting procedure partially corrects for bias from undercoverage, but biases may still be present when people who are missed by the survey differ from those interviewed in ways other than age, race, sex, Hispanic origin, and state of residence. How this weighting procedure affects other variables in the survey is not precisely known. All of these considerations affect comparisons across different surveys or data sources.

A common measure of survey coverage is the coverage ratio, calculated as the estimated population before poststratification divided by the independent population control. Table 2 shows July 2012 CPS coverage ratios by age and sex for certain race and Hispanic groups. The CPS coverage ratios can exhibit some variability from month to month.

Table 2. CPS Coverage Ratios: July 2012

Age group	All people	Total		White only		Black only		Residual race		Hispanic	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-15	0.87	0.88	0.86	0.92	0.89	0.75	0.77	0.78	0.78	0.85	0.82
16-19	0.85	0.84	0.85	0.87	0.89	0.76	0.68	0.78	0.85	0.87	0.87
20-24	0.74	0.74	0.75	0.75	0.77	0.68	0.70	0.71	0.67	0.72	0.70
25-34	0.84	0.82	0.85	0.84	0.88	0.73	0.80	0.74	0.76	0.74	0.82
35-44	0.86	0.85	0.88	0.87	0.91	0.76	0.75	0.75	0.77	0.81	0.90
45-54	0.88	0.87	0.89	0.89	0.91	0.75	0.78	0.83	0.80	0.82	0.87
55-64	0.89	0.88	0.89	0.90	0.90	0.76	0.84	0.84	0.88	0.77	0.81
65+	0.92	0.93	0.91	0.93	0.92	0.89	0.93	0.89	0.81	0.86	0.87
15+	0.86	0.85	0.87	0.88	0.89	0.75	0.79	0.78	0.79	0.79	0.84
0+	0.86	0.86	0.87	0.88	0.89	0.75	0.78	0.78	0.79	0.81	0.83

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. 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 Census 2010-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 Census 2010-based controls results in about a 0.2 percent increase from the 2000 census-based controls in the civilian noninstitutional 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 2000 files (see "Basic CPS"). During this time period, CPS data were collected from sample designs based on different censuses. Three features of the new CPS design have the potential of affecting published estimates: (1) the temporary disruption of the rotation pattern from August 2004 through June 2005 for a comparatively small portion of the sample, (2) the change in sample areas, and (3) the introduction of the new Core-Based Statistical Areas (formerly called metropolitan areas). Most of the known effect on estimates during and after the sample redesign will be the result of changing from 1990 to 2000 geographic definitions. Research has shown that the national-level estimates of the metropolitan and nonmetropolitan populations should not change appreciably because of the new sample design. However, users should still exercise caution when comparing metropolitan and nonmetropolitan estimates across years with a design change, especially at the state level.

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

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

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

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.

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 is a simple model that expresses the variance as a function of the expected value of the survey estimate. The parameters of the generalized variance function are estimated using direct replicate variances. These generalized variance parameters provide a relatively easy method to obtain approximate standard errors for numerous characteristics. In this source and accuracy statement, Table 4 provides the generalized variance parameters for labor force estimates, and Table 5 provides generalized variance parameters for characteristics from the July 2012 Public Participation in the Arts 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 generalized variance 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 generalized variance parameters to use in standard error calculations for the basic CPS. For PPAS, the race/ethnicity parameters are given in Table 5.

Table 3. Estimation Groups of Interest and Generalized Variance Parameters

Race/ethnicity group of interest	Generalized variance parameters to use in standard error calculations
Total population	Total or White
Total White, White AOIC, or White non-Hispanic population	Total or White
Total Black, Black AOIC, or Black non-Hispanic population	Black
Asian alone, Asian AOIC, or Asian non-Hispanic population	Asian, AIAN, NHOPI
AIAN alone, AIAN AOIC, or AIAN non-Hispanic population	
NHOPI alone, NHOPI AOIC, or NHOPI non-Hispanic population	
Populations from other race groups	Asian, AIAN, NHOPI
Hispanic population	Hispanic
Two or more races – employment/unemployment and educational attainment characteristics	Black
Two or more races – all other characteristics	API, AIAN, NHOPI

Notes: (1) API, AIAN, NHOPI are Asian and Pacific Islander, American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander, respectively.
(2) AOIC is an abbreviation for alone or in combination. The AOIC population for a race group of interest includes people reporting only the race group of interest (alone) and people reporting multiple race categories including the race group of interest (in combination).
(3) Hispanics may be any race.
(4) Two or more races refers to the group of cases self-classified as having two or more races.

Standard Errors of Estimated Totals and Means. The approximate standard error, s_x, of an total or mean 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 4 or 5 associated with the particular type of characteristic. When calculating standard errors from cross-tabulations involving different characteristics, use the set of parameters for the characteristic that will give the largest standard error.

Illustration 1
Suppose there were 6,863,000 unemployed men (ages 16 and up) in the civilian labor force. Use the appropriate parameters from Table 4 and Formula (1) to get

Illustration 1	
Number of unemployed males in the civilian labor force (x)	6,863,000
a parameter (a)	-0.000032
b parameter (b)	2,971
Standard error	137,000
90-percent confidence interval	6,638,000 to 7,088,000

The standard error is calculated as

$$s_x = \sqrt{-0.000032 * 6{,}863{,}000^2 + 2{,}971 * 6{,}863{,}000} = 137{,}000$$

The 90-percent confidence interval is calculated as 6,863,000 ± 1.645 × 137,000.

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 and Ratios. The reliability of an estimated percentage or ratio using sample data depends on the size of both the numerator, x, and denominator, y. This section presents two equations to calculate standard errors of estimated percentages and ratios. The first equation is simplified and can be used for most percentage estimates; the second equation can be used for all percentage and ratio estimates but is more complex. Use the following questions to determine if the simplified equation can be used to calculate the standard error of a percentage:

1) Do both the numerator and denominator use the same parameters from Table 4 or 5?
2) Is the denominator a CPS population control - a total by race/ethnicity (excluding the group self-classified as having two or more races), sex, or age group, or state? See "CPS Estimation Procedure" for more information on the specific CPS population controls and "PPAS Estimation Procedure" for more information on the specific PPAS population controls.

If the answer to either question is yes, then use the following simplified formula to find the approximate standard error, $s_{y,p}$, of the estimated percentage p:

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

Here y is the total number of people, families, households, or unrelated individuals in the denominator of the percentage, p is the percentage, and b is the parameter in Table 4 or 5 associated with the characteristic in the numerator of the percentage.

If the answer to both questions is no, or the estimate is not a percentage, compute the standard error of the ratio using

$$s_{x/y} = \frac{x}{y} \sqrt{\left(\frac{s_x}{x}\right)^2 + \left(\frac{s_y}{y}\right)^2 - 2r\frac{s_x s_y}{xy}} \tag{3}$$

The standard error of the numerator, s_x, and that of the denominator, s_y, may be calculated using standard error formulas described in this document. In Formula (3), r represents the correlation between the numerator and the denominator of the estimate. If r has not been previously calculated for a specific estimate, consider the type of ratio being estimated. For ratios where the numerator is a subset of the denominator use

$$r = \frac{x \cdot s_y}{y \cdot s_x} \tag{4}$$

For ratios where the denominator is a count of families or households and the numerator is a count of people in those families or households with a certain characteristic and there is at least one person with the characteristic in every family or household, use 0.7 as an estimate of r. An example of this type is the average number of children per family with children. For all other types of ratios, r is assumed to be zero. Examples are the average number of children per family. 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 (3) by 100 or 1,000, respectively, to obtain the standard error.

Illustration 2

Suppose there were 121,467,000 women aged 18 and over, and 8.3 percent indicate they listen to jazz. Use the appropriate parameter from Table 5 and Formula (2), since the denominator in this percentage is treated as a CPS population control, to get

Illustration 2	
Percentage of women 18+ who indicate they listen to jazz (*p*)	8.3
Base (*y*)	121,467,000
b parameter (*b*)	31,194
Standard error	0.44
90-percent confidence interval	7.6 to 9.0

The standard error is calculated as

$$s_{y,x} = \sqrt{\frac{31{,}194}{121{,}467{,}000} \times 8.3 \times (100 - 8.3)} = 0.44$$

The 90-percent confidence interval for the estimated percentage of women aged 18 years old or older who listen to jazz is from 7.6 to 9.0 percent (i.e., 8.3 ± 1.645 × 0.44).

Illustration 3

Suppose the ratio of men to women working part-time was 10,683,000 to 18,329,000, or 0.58. Use Formulas (1) and (3) with $r = 0$ and the appropriate parameters from Table 4 to get

Illustration 3			
	Males (*x*)	Females (*y*)	Ratio
Number who work part-time	10,683,000	18,329,000	0.58
a parameter (*a*)	-0.000032	-0.000031	-
b parameter (*b*)	2,971	2,782	-
Standard error	168,000	201,000	0.01
90-percent confidence interval	10,407,000 to 10,959,000	17,998,000 to 18,660,000	0.56 to 0.60

The standard error is calculated as

$$s_{x/y} = \frac{10{,}683{,}000}{18{,}329{,}000} \sqrt{\left(\frac{168{,}000}{10{,}683{,}000}\right)^2 + \left(\frac{201{,}000}{18{,}329{,}000}\right)^2} = 0.01$$

and the 90-percent confidence interval is calculated as 0.58 ± 1.645 × 0.01.

Illustration 4

Suppose that the number of unemployed males was 6,863,000 and the total number unemployed was 13,400,000. The ratio of unemployed males to the total number unemployed would be 0.50 or 50 percent. The numerator and denominator in this percentage do not use the same parameters from Table 4, and the denominator is not a CPS population control. Therefore, use Formulas (3) and (4) for the standard error and correlation, *r*, along with Formula (1) and the appropriate parameters from Table 4 to get

Illustration 4			
	Unemployed Males (x)	Unemployed Total (y)	Ratio
Number Unemployed	6,863,000	13,400,000	.51
a parameter (a)	-0.000032	-0.000016	-
b parameter (b)	2,971	3,096	-
correlation (r)	-	-	0.74
Standard error	137,000	197,000	0.01
90-percent confidence interval	6,638,000 to 7,088,000	13,076,000 to 13,724,000	0.49 to 0.53

The correlation is calculated as

$$r = \frac{6{,}863{,}000 * 197{,}000}{13{,}400{,}000 \times 137{,}000} = 0.74$$

The standard error is calculated as

$$s_{x/y} = \frac{6{,}863{,}000}{13{,}400{,}000}\sqrt{\left(\frac{137{,}000}{6{,}863{,}000}\right)^2 + \left(\frac{197{,}000}{13{,}400{,}000}\right)^2 - \left(2 * 0.74 * \frac{137{,}000 * 197{,}000}{6{,}863{,}000 * 13{,}400{,}000}\right)} = 0.01$$

and the 90-percent confidence interval is calculated as 0.51± 1.645 × 0.01.

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

$$s_{x_1 - x_2} = \sqrt{s_{x_1}^2 + s_{x_2}^2} \qquad (5)$$

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 5

Suppose that of the 69,947,000 people with a high school diploma but no college, 0.4 percent attended a live opera, and of the 67,903,000 people with some college or associate degree, 1.7 percent attended a live opera. Use the appropriate parameters from Table 5 and Formulas (2) and (3) to get

Illustration 5			
	High School Diploma (x_1)	Some College or Associates (x_2)	Difference
Percentage who attended live opera (p)	0.4	1.7	1.3
Base	69,974,000	67,903,000	-
b parameter (b)	20,357	20,357	-
Standard error	0.11	0.22	0.25
90-percent confidence interval	0.2 to 0.6	1.3 to 2.1	0.9 to 1.7

The standard error of the difference is calculated as

$$s_{x-y} = \sqrt{0.11^2 + 0.22^2} = 0.25$$

The 90-percent confidence interval around the difference is calculated as $1.3 \pm 1.645 \times 0.25$. Since this interval does not include zero, we can conclude with 90 percent confidence that the percentage of people with some college or associate degree who attended a live opera is greater than the percentage of people with a high school diploma who attended a live opera.

Standard Errors for Cross-Module Analysis. The standard errors of estimates from cross-module analysis may be obtained by determining new a and b parameters and using these adjusted parameters in the standard error formulas mentioned previously. To determine a new cross-module b parameter, multiply the Core b parameter from Table 5 by the factor provided in Table 1. For example, the cross-module factor to apply to tables weighted by PWSWGT and PWNWGT is 3.9375.

To determine the new a parameter, use the following formula:

$$a_{cross-module} = \frac{-b_{cross-module}}{POP_{item}}$$

where POP_{item} is the population found in Table 5.

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 the "Explanatory Notes and Estimates of Error: Household Data" section in *Employment and Earnings*, a monthly report published by the U.S. Bureau of Labor Statistics.

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 4. Parameters for Computation of Standard Errors for Labor Force Characteristics: July 2012		
Characteristic	a	b
Total or White		
Civilian labor force, employed	-0.000016	3,068
Not in labor force	-0.000009	1,833
Unemployed	-0.000016	3,096
Civilian labor force, employed, not in labor force, and unemployed		
Men	-0.000032	2,971
Women	-0.000031	2,782
Both sexes, 16 to 19 years	-0.000022	3,096
Black		
Civilian labor force, employed, not in labor force, and unemployed		
Total	-0.000151	3,455
Men	-0.000311	3,357
Women	-0.000252	3,062
Both sexes, 16 to 19 years	-0.001632	3,455
Hispanic		
Civilian labor force, employed, not in labor force, and unemployed		
Total	-0.000141	3,455
Men	-0.000253	3,357
Women	-0.000266	3,062
Both sexes, 16 to 19 years	-0.001528	3,455
Asian, AIAN, NHOPI		
Civilian labor force, employed, not in labor force, and unemployed		
Total	-0.000346	3,198
Men	-0.000729	3,198
Women	-0.000659	3,198
Both sexes, 16 to 19 years	-0.004146	3,198

Notes: (1) These parameters are to be applied to basic CPS monthly labor force estimates.
(2) API, AIAN, NHOPI are Asian and Pacific Islander, American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander, respectively.
(3) 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.
(4) Hispanics may be any race. For a more detailed discussion on the use of parameters for race and ethnicity, please see the "Generalized Variance Parameters" section.
(5) For nonmetropolitan characteristics, multiply the *a* and *b* parameters by 1.5. If the characteristic of interest is total state population, not subtotaled by race or ethnicity, the *a* and *b* parameters are zero.

Table 5. Parameters for Computation of Standard Errors for Public Participation in the Arts Characteristics: July 2012[1]

Characteristic	Core 1 (PWOWGT)		Core 2 (PWTWGT)		Module A2 (PWAWGT)[2]		Modules A1 and D (PWSWGT)		Modules B, C, and E (PWNWGT)		Population
	a	b	a	b	a	b	a	b	a	b	
All Adults	-0.000108	25,331	-0.000113	26,633	-0.000150	35,192	-0.000119	27,985	-0.000146	34,265	234,993,544
Sex											
Male	-0.000175	19,757	-0.000184	20,769			-0.000218	24,699	-0.000339	38,290	113,111,098
Female	-0.000162	19,757	-0.000170	20,769			-0.000203	24,699	-0.000256	31,194	121,882,446
Ethnicity and Race											
Hispanic[3]	-0.000895	31,294	-0.000902	31,561			-0.001107	38,728	-0.001329	46,472	34,978,765
Nonhispanic White	-0.000147	22,914	-0.000156	24,277			-0.000178	27,758	-0.000227	35,388	155,746,242
Nonhispanic African American	-0.001167	31,294	-0.001177	31,561			-0.001588	42,582	-0.001320	35,388	26,818,640
Nonhispanic Other	-0.001501	26,198	-0.001515	26,444			-0.002219	38,728	-0.002233	38,970	17,449,898
Age	-0.000079	18,554	-0.000084	19,651			-0.000090	21,237	-0.000126	29,535	234,993,544
Income	-0.000107	25,097	-0.000104	24,425			-0.000129	30,198	-0.000142	33,305	234,993,544
Education	-0.000087	20,357	-0.000088	20,577			-0.000108	25,330	-0.000139	32,614	234,993,544
State and Region											
New England	-0.001165	16,719	-0.001171	16,805			-0.001326	19,030	-0.001480	21,238	14,354,570
Connecticut	-0.003449	12,190	-0.003188	11,265			-0.005385	19,030	-0.005287	18,684	3,534,000
Maine	-0.003531	4,648	-0.003839	5,053			-0.004845	6,377	-0.006107	8,038	1,316,141
Massachusetts	-0.003991	26,124	-0.004276	27,986			-0.005129	33,567	-0.005189	33,965	6,545,152
Rhode Island	-0.003971	4,109	-0.003782	3,914			-0.004801	4,969	-0.005283	5,467	1,034,917
Remainder New England[4]	-0.001913	3,682	-0.002034	3,914			-0.002582	4,969	-0.002841	5,467	1,924,360
Mid-Atlantic	-0.000751	30,519	-0.000767	31,157			-0.000933	37,925	-0.001193	48,478	40,644,243
New Jersey	-0.003491	30,519	-0.003983	34,825			-0.004338	37,925	-0.004907	42,900	8,742,479

Table 5. Parameters for Computation of Standard Errors for Public Participation in the Arts Characteristics: July 2012[1]

Characteristic	Core 1 (PWOWGT)		Core 2 (PWTWGT)		Module A2 (PWAWGT)[2]		Modules A1 and D (PWSWGT)		Modules B, C, and E (PWNWGT)		Population
	a	b	a	b	a	b	a	b	a	b	
New York	-0.001581	30,519	-0.001804	34,825			-0.001965	37,925	-0.002511	48,478	19,303,797
Pennsylvania	-0.002140	26,956	-0.002178	27,441			-0.003010	37,925	-0.004126	51,980	12,597,967
South Atlantic	-0.000402	24,067	-0.000449	26,856			-0.000621	37,137	-0.000646	38,646	59,834,763
Florida	-0.001624	30,778	-0.001891	35,833			-0.002804	53,151	-0.002215	41,977	18,953,952
Georgia	-0.003181	30,778	-0.003703	35,833			-0.004138	40,039	-0.004338	41,977	9,676,062
Maryland	-0.003013	17,395	-0.002986	17,239			-0.004015	23,177	-0.004605	26,585	5,773,145
North Carolina	-0.003230	30,778	-0.003181	30,308			-0.004203	40,039	-0.004406	41,977	9,527,385
South Carolina	-0.005121	23,601	-0.004459	20,551			-0.005029	23,177	-0.006617	30,493	4,608,503
Virginia	-0.002970	23,601	-0.003380	26,856			-0.003965	31,503	-0.004864	38,646	7,945,311
West Virginia	-0.004124	7,550	-0.005698	10,431			-0.007126	13,045	-0.007812	14,300	1,830,636
Remainder S. Atlantic[5]	-0.002184	3,318	-0.002015	3,062			-0.002666	4,052	-0.003754	5,706	1,519,769
East North Central	-0.000543	24,969	-0.000529	24,312			-0.000758	34,876	-0.000919	42,288	45,999,618
Illinois	-0.002132	27,125	-0.002040	25,947			-0.002742	34,876	-0.003325	42,288	12,719,872
Michigan	-0.002776	27,125	-0.002655	25,947			-0.003569	34,876	-0.004328	42,288	9,771,598
Ohio	-0.002119	24,152	-0.002277	25,947			-0.002588	29,495	-0.003176	36,200	11,396,362
Remainder E.N. Central[6]	-0.001865	22,592	-0.001807	21,887			-0.002879	34,876	-0.002989	36,200	12,111,786
West North Central	-0.000751	15,347	-0.000665	13,586			-0.000861	17,594	-0.001021	20,856	20,424,312
Iowa	-0.004732	14,373	-0.003413	10,369			-0.005725	17,392	-0.004769	14,487	3,037,761
Kansas	-0.005086	14,373	-0.003669	10,369			-0.005603	15,833	-0.007380	20,856	2,826,041
Minnesota	-0.002329	12,391	-0.002554	13,586			-0.002976	15,833	-0.003352	17,834	5,319,676
Missouri	-0.004255	25,194	-0.003895	23,062			-0.005077	30,058	-0.005034	29,805	5,920,434
Nebraska	-0.003781	6,913	-0.004682	8,560			-0.005189	9,487	-0.005290	9,671	1,828,151
North Dakota	-0.004920	3,337	-0.005164	3,502			-0.006430	4,361	-0.008782	5,956	678,139

Table 5. Parameters for Computation of Standard Errors for Public Participation in the Arts Characteristics: July 2012[1]

Characteristic	Core 1 (PWOWGT)		Core 2 (PWTWGT)		Module A2 (PWAWGT)[2]		Modules A1 and D (PWSWGT)		Modules B, C, and E (PWNWGT)		Population
	a	b	a	b	a	b	a	b	a	b	
South Dakota	-0.003283	2,673	-0.005255	4,278			-0.004916	4,002	-0.006097	4,964	814,110
East South Central	-0.000519	23,870	-0.000510	23,444			-0.000743	34,167	-0.000721	33,144	45,999,618
Alabama	-0.006165	29,249	-0.005856	27,780			-0.009182	43,560	-0.009654	45,800	4,744,286
Remainder East South Central[7]	-0.001635	22,171	-0.001646	22,322			-0.002520	34,167	-0.002189	29,686	13,559,940
West South Central	-0.000776	28,490	-0.000761	27,931			-0.000999	36,667	-0.001098	40,313	36,715,430
Texas	-0.001114	28,490	-0.001202	30,731			-0.001434	36,667	-0.001576	40,313	25,574,581
Remainder W.S. Central[8]	-0.002065	23,002	-0.001942	21,637			-0.003291	36,667	-0.003854	42,932	11,140,849
Mountain	-0.000740	16,433	-0.000850	18,885			-0.001244	27,642	-0.001284	28,526	22,217,388
Colorado	-0.002799	14,216	-0.004162	21,142			-0.005442	27,642	-0.004656	23,651	5,079,448
Nevada	-0.003935	10,617	-0.004362	11,769			-0.005214	14,069	-0.005187	13,997	2,698,332
Wyoming	-0.005525	3,101	-0.004924	2,764			-0.006316	3,545	-0.007894	4,430	561,194
Remainder Mountain[9]	-0.001386	19,239	-0.001523	21,142			-0.002124	29,481	-0.002406	33,396	13,878,414
Pacific	-0.000601	30,136	-0.000615	30,842			-0.000713	35,734	-0.000825	41,383	50,143,307
California	-0.000939	35,178	-0.000951	35,640			-0.001107	41,485	-0.001252	46,902	37,459,067
Oregon	-0.004575	17,667	-0.004765	18,400			-0.005697	22,000	-0.006319	24,400	3,861,658
Washington	-0.003357	22,787	-0.003700	25,113			-0.005264	35,734	-0.005458	37,046	6,787,833
Remainder Pacific[10]	-0.002612	5,316	-0.002251	4,581			-0.003172	6,454	-0.003071	6,249	2,034,749
	#DIV/0!	0	0.000000	0			0.000000	0	0.000000	0	
Metropolitan Areas	-0.003164	24,834	-0.003257	25,563			-0.003770	29,595	-0.003874	30,407	7,849,226
Boston-Worcester-Manchester, MA-NH	-0.001534	29,416	-0.001528	29,298			-0.001813	34,754	-0.002254	43,213	19,170,852
Chicago-Naperville-Michigan City, IL-IN	-0.001150	29,416	-0.001146	29,298			-0.001359	34,754	-0.001690	43,213	25,574,581
Dallas-Fort Worth, TX	-0.002870	14,576	-0.003710	18,847			-0.003852	19,564	-0.004284	21,759	5,079,448
Denver-Aurora-Boulder, CO	-0.003283	2,673	-0.005255	4,278			-0.004916	4,002	-0.006097	4,964	814,110

Table 5. Parameters for Computation of Standard Errors for Public Participation in the Arts Characteristics: July 2012[1]

Characteristic	Core 1 (PWOWGT)		Core 2 (PWTWGT)		Module A2 (PWAWGT)[2]		Modules A1 and D (PWSWGT)		Modules B, C, and E (PWNWGT)		Population
	a	b	a	b	a	b	a	b	a	b	
Detroit-Warren-Flint, MI	-0.003010	29,416	-0.002998	29,298			-0.003205	31,316	-0.004422	43,213	9,771,598
Los Angeles-Long Beach-Riverside, CA	-0.000966	36,179	-0.000977	36,615			-0.001011	37,857	-0.001312	49,159	37,459,067
Miami-Fort Lauderdale-Miami Beach, FL	-0.001552	29,416	-0.001784	33,820			-0.003347	63,441	-0.002009	38,084	18,953,952
NY-Newark-Bridgeport, NY-NJ-CT-PA	-0.000666	29,416	-0.000829	36,615			-0.000857	37,857	-0.001113	49,159	44,178,243
Philadelphia-Camden-Vineland, PA-NJ-DE-MD	-0.001050	29,416	-0.001046	29,298			-0.000997	27,926	-0.001543	43,213	28,011,945
San Jose-Francisco-Oakland, CA	-0.000785	29,416	-0.000903	33,820			-0.001243	46,544	-0.001312	49,159	37,459,067
Washington-Baltimore-Northern Virginia , DC-MD-VA-WV	-0.001340	21,674	-0.001248	20,181			-0.001415	22,874	-0.001880	30,407	16,170,507
Occupation	-0.000072	16,904	-0.000083	19,555			-0.000098	23,124	-0.000126	29,619	234,993,544

Notes: (1) These parameters are to be applied to the July 2012 Public Participation in the Arts Supplement data.
(2) PWAWGT does not give breakdowns for parameters because it is only used for an overall estimate.
(3) Hispanics may be any race.
(4) Remainder New England includes New Hampshire and Vermont.
(5) Remainder S. Atlantic includes Delaware and the District of Columbia.
(6) Remainder E.N. Central includes Indiana and Wisconsin.
(7) Remainder E. S. Central includes Kentucky, Mississippi, and Tennessee.
(8) Remainder W.S. Central includes Arkansas, Louisiana, and Oklahoma.
(9) Remainder Mountain includes Arizona, Idaho, New Mexico, Montana, and Utah.
(10) Remainder Pacific includes Alaska and Hawaii.

References

[1] Bureau of Labor Statistics. 1994. *Employment and Earnings*. Volume 41 Number 5, May 1994. Washington, DC: Government Printing Office.

[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. (http://www.fcsm.gov/working-papers/spp.html)