Appendix I

MEMORANDUM FOR: Cheryl Landman
Chief, Demographic Surveys Division

From: Ruth Ann Killion
Chief, Demographic Statistical Methods Division

Subject: Source and Accuracy Statement for the May 2008 CPS Microdata File on Public Participation in the Arts

Attached is the statement on the source of the data and accuracy of the estimates for the May 2008 CPS Microdata File on Public Participation in the Arts.

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:

L. Clement (DSD)
G. Weyland
R. Schwartz
B. Kominski (HHES)
W. Savino (ACSD)
A. Shields

email:

P. Flanagan (DSMD)
J. Scott
S. Adeshiyan
X. Liu
B. Tran
T. Moore
HSSB (9)

Source of the Data and Accuracy of the Estimates for the May 2008 CPS Microdata File on Public Participation in the Arts

Table of Contents

Tables

Source of the Data and Accuracy of the Estimates for the May 2008 CPS Microdata File on Public Participation in the Arts

SOURCE OF THE DATA

The data in this microdata file are from the May 2008Current Population Survey (CPS). The U.S. Census Bureau conducts the CPS every month, although this file has only May 2008 data. The May 2008 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 May 2008.

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 (91 percent of the 4.1 million institutionalized people in Census 2000). Interviewers ask questions concerning labor force participation about each member 15 years old and over in sample households. Typically, the week containing the nineteenth of the month is the interview week. The week containing the twelfth is the reference week (i.e., the week about which the labor force questions are asked).

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

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, within PSUs selected for both the 1990 and 2000 designs, sample households from the 2000 design gradually replaced sample households from the 1990 design. This involved about 90 percent of the sample. The new sample design was completely implemented by July 2005.

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

---

1 For detailed information on the 1990 sample redesign, 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 May 2008. 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.

May 2008 Supplement. In May 2008, 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-fourth 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 May 1, 2007 to May 1, 2008. 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 a core set of questions and four modules. Module A was titled Reading and Music Preference, module B was titled Participation Via Internet and Other Media, Module C was titled Leisure Activities, and Module D was titled Arts Learning. Each module was administered to only a portion of the sampled cases. Interviews were conducted during the period of May 18 - 24, 2008.

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.

---

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

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.

PPAS Estimation Procedure. The PPAS 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 are18 years old or older.

The questionnaire modules and the special core question were originally assigned to households so that half of the sample would receive each module and the special core question. Problems occurred during the selection of respondents that changed the probabilities used to assign the modules, the special core question, and the selection of sample cases within the household. The selection probabilities were corrected in the weighting. The module factor, as described later in this section, was modified to account for the assignment of the modules and the special core questions.

Each sampled person receives one or two weights for the PPA survey depending on the modules asked. The first weight should be used to create estimates from the core and module C since questions were asked about the respondent's spouse/partner in these sections. The second weight should be used to create estimates from modules A, B, and D since these sections did not include questions about the respondent's spouse/partner. Both 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	
Analysis of Module	**Module Factor to Assign**
Core Questions Only	1.000000
A or B	2.222222
C or D or Special Core Question	1.818182
A and B in combination	12.000000
A and C or A and D or B and C or B and D or C and D in combination	5.454545

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 May 2008 basic CPS, the household-level nonresponse rate was 7.8 percent. The person-level nonresponse rate for the Public Participation in the Arts supplement was an additional 18.4 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 May 2008 is estimated to be about 12 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 May 2008 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: May 2008											
Age group	Total			White only		Black only		Residual race		Hispanic	
	All people	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-15	0.89	0.89	0.90	0.90	0.91	0.80	0.80	0.95	0.93	0.89	0.89
16-19	0.89	0.88	0.89	0.89	0.91	0.82	0.83	0.94	0.85	0.95	0.89
20-24	0.77	0.76	0.79	0.77	0.80	0.67	0.73	0.73	0.81	0.85	0.86
25-34	0.83	0.80	0.86	0.83	0.86	0.64	0.80	0.78	0.88	0.78	0.91
35-44	0.88	0.85	0.90	0.87	0.93	0.73	0.80	0.78	0.84	0.78	0.91
45-54	0.90	0.89	0.91	0.90	0.92	0.85	0.87	0.79	0.84	0.83	0.88
55-64	0.90	0.90	0.90	0.91	0.91	0.84	0.88	0.86	0.86	0.89	0.89
65+	0.93	0.92	0.94	0.91	0.94	1.01	0.97	0.89	0.82	0.86	0.93
15+	0.88	0.86	0.89	0.88	0.90	0.78	0.84	0.82	0.85	0.83	0.90
0+	0.88	0.87	0.89	0.88	0.91	0.79	0.83	0.85	0.87	0.84	0.89

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 2000-based controls, with microdata files from March 1994 through December 2002, 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 can create small differences between estimates. See the discussion following for information on comparing estimates derived from different controls or different sample designs.

Microdata files from previous years reflect the latest available census-based controls. Although the most recent change in population controls had relatively little impact on summary measures such as averages, medians, and percentage distributions, it did have a significant impact on levels. For example, use of Census 2000-based controls results in about a 1 percent increase from the 1990 census-based controls in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 2003 and later years will differ from those for earlier years by more than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain population subgroups than for the total population.

Note that certain microdata files from 2002, namely June, October, November, and the 2002 ASEC, contain both Census 2000-based estimates and 1990 census-based estimates and are

subject to the comparability issues discussed previously. All other microdata files from 2002 reflect the 1990 census-based controls.

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

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

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

For additional information on nonsampling error including the possible impact on CPS data when known, refer to 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 May 2008 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 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 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 4,508,000 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*)	4,508,000
a parameter (*a*)	-0.000032
b parameter (*b*)	2,971
Standard error	113,000
90-percent confidence interval	4,322,000 to 4,694,000

The standard error is calculated as

$$s_x = \sqrt{-0.000032 \times 4{,}508{,}000^2 + 2{,}971 \times 4{,}508{,}000} = 113{,}000$$

The 90-percent confidence interval is calculated as 4,508,000 ± 1.645 × 113,000.

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*:

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

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

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

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

$$r = \frac{x \cdot s_y}{y \cdot s_x} \quad (4)$$

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 116,300,000 women aged 18 and over, and 13.6 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)	13.6
Base (y)	116,300,000
b parameter (b)	35,647
Standard error	0.6
90-percent confidence interval	12.6 to 14.6

The standard error is calculated as

$$s_{y,p} = \sqrt{\frac{35{,}647}{116{,}300{,}000} \times 13.6 \times (100 - 13.6)} = 0.6$$

The 90-percent confidence interval for the estimated percentage of women aged 18 years old or older who listen to jazz is from 12.61 to 14.59 percent (i.e., $13.6 \pm 1.645 \times 0.6$).

Illustration 3

Suppose the ratio of men to women working part-time was 9,223,000 to 17,667,000, or 0.52. 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	9,223,000	17,667,000	0.52
a parameter (a)	-0.000032	-0.000031	-
b parameter (b)	2,971	2,782	-
Standard error	157,000	199,000	0.01
90-percent confidence interval	8,965,000 to 9,481,000	17,340,000 to 17,994,000	0.50 to 0.54

The standard error is calculated as

$$s_{x/y} = \frac{9{,}223{,}000}{17{,}667{,}000}\sqrt{\left(\frac{157{,}000}{9{,}223{,}000}\right)^2 + \left(\frac{199{,}000}{17{,}667{,}000}\right)^2} = 0.01$$

and the 90-percent confidence interval is calculated as $0.52 \pm 1.645 \times 0.01$.

Illustration 4

Suppose that the number of unemployed males was 4,508,000 and the total number unemployed was 8,193,000. The ratio of unemployed males to the total number unemployed would be 0.55 or 55 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	4,508,000	8,193,000	55.0
a parameter (a)	-0.000032	-0.000016	-
b parameter (b)	2,971	3,096	-
correlation (r)	-	-	0.76
Standard error	113,000	156,000	0.9
90-percent confidence interval	4,322,000 to 4,694,000	7,936,000 to 8,450,000	53.5 to 56.5

The correlation is calculated as

$$r = \frac{4{,}508{,}000 \times 156{,}000}{8{,}193{,}000 \times 113{,}000} = 0.76$$

The standard error is calculated as

$$s_{x/y} = \frac{4{,}508{,}000}{8{,}193{,}000}\sqrt{\left(\frac{113{,}000}{4{,}508{,}000}\right)^2 + \left(\frac{156{,}000}{8{,}193{,}000}\right)^2 - \left(2 \times 0.76 \times \frac{113{,}000 \times 156{,}000}{4{,}508{,}000 \times 8{,}193{,}000}\right)} = 0.009$$

and the 90-percent confidence interval is calculated as 0.55 ± 1.645 × 0.009.

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

$$s_{x_1 - x_2} = \sqrt{{s_{x_1}}^2 + {s_{x_2}}^2} \qquad (3)$$

where s_{x_1} and s_{x_2} are the standard errors of the estimates, x_1 and x_2. The estimates can be numbers, percentages, ratios, etc. This will result in accurate estimates of the standard error of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. However, if there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration 5

Suppose that of the 68,300,000 people with a high school diploma but no college, 9.5 percent attended a live opera, and of the 61,400,000 people with some college or associate degree, 21.3 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 working part-time (p)	9.5	21.3	11.8
Base	68,300,000	61,400,000	-
b parameter (b)	40,263	40,263	-
Standard error	0.71	1.05	1.27
90-percent confidence interval	8.3 to 10.7	19.6 to 23.0	9.7 to 13.9

The standard error of the difference is calculated as

$$s_{x-y} = \sqrt{0.71^2 + 1.05^2} = 1.27$$

The 90-percent confidence interval around the difference is calculated as 11.8 ± 1.645 × 1.27. 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 Module A and B is 12.0.

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: May 2008

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: May 2008[1]

Characteristic	Core		Modules A or B		Module C or Special Core Question		Module D		Population
	a	b	a	b	a	b	a	b	
All Adults	-0.000118	26,532	-0.000266	59,862	-0.000170	38,332	-0.000220	49,404	224,826,742
Sex									
Male	-0.000186	20,220	-0.000551	59,862	-0.000301	32,699	-0.000455	49,404	108,545,640
Female	-0.000174	20,220	-0.000515	59,862	-0.000281	32,699	-0.000425	49,404	116,281,102
Ethnicity and Race									
Hispanic[2]	-0.001017	30,967	-0.002524	76,829	-0.001635	49,784	-0.001892	57,588	30,444,019
Nonhispanic White	-0.000152	23,545	-0.000388	59,862	-0.000248	38,332	-0.000320	49,404	154,461,582
Nonhispanic African American	-0.001210	30,967	-0.003001	76,829	-0.001945	49,784	-0.002250	57,588	25,597,094
Nonhispanic Other	-0.002162	30,967	-0.005845	83,729	-0.003476	49,784	-0.004784	68,532	14,324,047
Age	-0.000075	16,951	-0.000226	50,822	-0.000129	28,929	-0.000173	38,917	224,826,742
Income	-0.000118	26,532	-0.000266	59,862	-0.000170	38,332	-0.000220	49,404	224,826,742
Education	-0.000118	26,532	-0.000266	59,862	-0.000170	38,332	-0.000220	49,404	224,826,742
State and Region									
New England	-0.001406	17,961	-0.003052	38,994	-0.003000	38,332	-0.002448	31,280	12,775,516
Connecticut	-0.003866	13,328	-0.012019	41,439	-0.006152	21,211	-0.008411	28,999	3,447,696
Maine	-0.002795	3,642	-0.008512	11,091	-0.004929	6,422	-0.006483	8,447	1,302,995
Massachusetts	-0.004164	26,532	-0.009921	63,217	-0.006833	43,540	-0.007753	49,404	6,371,844
Rhode Island	-0.003511	3,642	-0.010693	11,091	-0.006191	6,422	-0.008144	8,447	1,037,252
Remainder New England[3]	-0.005915	3,642	-0.018013	11,091	-0.010430	6,422	-0.013719	8,447	615,729
Mid-Atlantic	-0.000818	32,608	-0.002100	83,729	-0.001421	56,653	-0.001719	68,532	39,861,959
New Jersey	-0.003088	26,532	-0.007358	63,217	-0.005067	43,540	-0.005750	49,404	8,592,019

Table 5. Parameters for Computation of Standard Errors for Public Participation in the Arts Characteristics: May 2008[1]

Characteristic	Core		Modules A or B		Module C or Special Core Question		Module D		Population
	a	b	a	b	a	b	a	b	
New York	-0.001883	35,847	-0.004397	83,729	-0.003530	67,209	-0.004409	83,952	19,041,198
Pennsylvania	-0.002170	26,532	-0.005170	63,217	-0.003560	43,540	-0.004040	49,404	12,228,742
South Atlantic	-0.000464	26,532	-0.001046	59,862	-0.000670	38,332	-0.000863	49,404	57,236,836
Florida	-0.001323	23,885	-0.003107	56,111	-0.001811	32,699	-0.002736	49,404	18,059,796
Georgia	-0.003065	29,092	-0.006943	65,909	-0.004856	46,095	-0.005205	49,404	9,492,256
Maryland	-0.003328	18,440	-0.007907	43,814	-0.005416	30,011	-0.006378	35,346	5,541,450
North Carolina	-0.003229	29,092	-0.007317	65,909	-0.005117	46,095	-0.005484	49,404	9,008,211
South Carolina	-0.005477	23,885	-0.012867	56,111	-0.008790	38,332	-0.011329	49,404	4,360,741
Virginia	-0.003164	23,885	-0.007433	56,111	-0.005078	38,332	-0.006544	49,404	7,549,167
West Virginia	-0.005539	9,901	-0.013857	24,772	-0.008752	15,645	-0.009666	17,279	1,787,633
Remainder S. Atlantic[4]	-0.002264	3,254	-0.005811	8,354	-0.004053	5,826	-0.004085	5,873	1,437,582
East North Central	-0.000511	23,408	-0.001536	70,275	-0.001082	49,525	-0.001079	49,404	45,765,789
Illinois	-0.001840	23,408	-0.005004	63,666	-0.003013	38,332	-0.003883	49,404	12,721,800
Michigan	-0.002360	23,408	-0.006419	63,666	-0.003865	38,332	-0.004981	49,404	9,918,880
Ohio	-0.002072	23,408	-0.005635	63,666	-0.003392	38,332	-0.004372	49,404	11,299,174
Remainder E.N. Central[5]	-0.001648	19,487	-0.005474	64,735	-0.002765	32,699	-0.003526	41,702	11,825,935
West North Central	-0.000702	13,901	-0.002021	40,039	-0.001170	23,173	-0.001682	33,315	19,811,330
Iowa	-0.002638	7,786	-0.007308	21,568	-0.004657	13,745	-0.006111	18,036	2,951,442
Kansas	-0.003924	10,729	-0.012003	32,819	-0.007602	20,786	-0.008428	23,044	2,734,129
Minnesota	-0.002077	10,729	-0.006355	32,819	-0.004025	20,786	-0.004462	23,044	5,164,487
Missouri	-0.004036	23,408	-0.010977	63,666	-0.006609	38,332	-0.008518	49,404	5,800,136
Nebraska	-0.004446	7,786	-0.012316	21,568	-0.007849	13,745	-0.010299	18,036	1,751,178
North Dakota	-0.004249	2,655	-0.011724	7,325	-0.007113	4,444	-0.008641	5,399	624,786

Table 5. Parameters for Computation of Standard Errors for Public Participation in the Arts Characteristics: May 2008[1]

Characteristic	Core		Modules A or B		Module C or Special Core Question		Module D		Population
	a	b	a	b	a	b	a	b	
South Dakota	-0.003381	2,655	-0.009329	7,325	-0.005660	4,444	-0.006876	5,399	785,172
East South Central	-0.001495	26,532	-0.003374	59,862	-0.002160	38,332	-0.002784	49,404	17,743,068
Alabama	-0.006352	29,092	-0.014392	65,909	-0.010065	46,095	-0.010788	49,404	4,579,659
Remainder East South Central[6]	-0.001814	23,885	-0.004263	56,111	-0.002484	32,699	-0.003753	49,404	13,163,409
West South Central	-0.000771	26,532	-0.001739	59,862	-0.001114	38,332	-0.001436	49,404	34,414,531
Texas	-0.001221	29,092	-0.002766	65,909	-0.001935	46,095	-0.002073	49,404	23,827,505
Remainder W.S. Central[7]	-0.002256	23,885	-0.005300	56,111	-0.003621	38,332	-0.004666	49,404	10,587,026
Mountain	-0.000853	18,281	-0.002373	50,822	-0.001790	38,332	-0.002306	49,404	21,419,886
Colorado	-0.003768	18,281	-0.007677	37,246	-0.006016	29,188	-0.006245	30,295	4,851,354
Nevada	-0.004013	10,394	-0.011315	29,310	-0.007298	18,905	-0.008509	22,041	2,590,269
Wyoming	-0.004597	2,400	-0.021254	11,097	-0.008513	4,445	-0.010235	5,344	522,125
Remainder Mountain[8]	-0.001359	18,281	-0.003777	50,822	-0.002849	38,332	-0.003671	49,404	13,456,138
Pacific	-0.000549	26,532	-0.001405	67,885	-0.000793	38,332	-0.001248	60,282	48,321,085
California	-0.000733	26,532	-0.001814	65,718	-0.001189	43,075	-0.001364	49,404	36,220,464
Oregon	-0.004889	18,281	-0.013591	50,820	-0.010251	38,332	-0.013212	49,404	3,739,264
Washington	-0.004117	26,532	-0.010196	65,718	-0.006683	43,075	-0.007665	49,404	6,445,194
Remainder Pacific[9]	-0.003298	6,319	-0.008149	15,615	-0.005088	9,750	-0.009870	18,913	1,916,163
Metropolitan Areas									
Boston-Worcester-Manchester, MA-NH	-0.002385	26,532	-0.004037	44,900	-0.003968	44,138	-0.004442	49,404	11,122,535
Chicago-Naperville-Michigan City, IL-IN	-0.001763	33,497	-0.003748	71,232	-0.002838	53,925	-0.003030	57,587	19,003,804
Dallas-Fort Worth, TX	-0.001406	33,497	-0.002989	71,232	-0.002263	53,925	-0.002417	57,587	23,827,505
Denver-Aurora-	-0.003494	16,951	-0.007140	34,641	-0.005380	26,100	-0.005941	28,823	4,851,354

Table 5. Parameters for Computation of Standard Errors for Public Participation in the Arts Characteristics: May 2008[1]

Characteristic	Core		Modules A or B		Module C or Special Core Question		Module D		Population
	a	b	a	b	a	b	a	b	
Boulder, CO									
Detroit-Warren-Flint, MI	-0.000972	33,497	-0.002067	71,232	-0.001565	53,925	-0.001671	57,587	34,466,615
Los Angeles-Long Beach-Riverside, CA	-0.001101	39,872	-0.002204	79,815	-0.001489	53,925	-0.001794	64,962	36,220,464
Miami-Fort Lauderdale-Miami Beach, FL	-0.002208	39,872	-0.004419	79,815	-0.002986	53,925	-0.003597	64,962	18,059,796
NY-Newark-Bridgeport, NY-NJ-CT-PA	-0.000921	39,872	-0.001843	79,815	-0.001245	53,925	-0.001500	64,962	43,309,655
Philadelphia-Camden-Vineland, PA-NJ-DE-MD	-0.001271	33,497	-0.002702	71,232	-0.002046	53,925	-0.002184	57,587	26,362,211
San Jose-Francisco-Oakland, CA	-0.001101	39,872	-0.002204	79,815	-0.001489	53,925	-0.001794	64,962	36,220,464
Washington-Baltimore-Northern Virginia , DC-MD-VA-WV	-0.000151	23,407	-0.000333	51,440	-0.000212	32,699	-0.000186	28,823	154,557,079
Occupation	-0.000075	16,951	-0.000244	54,832	-0.000130	29,237	-0.000220	49,404	224,826,742

Notes: (1) These parameters are to be applied to the May 2008 Public Participation in the Arts Supplement data.
(2) Hispanics may be any race.
(3) Remainder New England includes New Hampshire and Vermont.
(4) Remainder S. Atlantic includes Delaware and the District of Columbia.
(5) Remainder E.N. Central includes Indiana and Wisconsin.
(6) Remainder E. S. Central includes Kentucky, Mississippi, and Tennessee.
(7) Remainder W.S. Central includes Arkansas, Louisiana, and Oklahoma.
(8) Remainder Mountain includes Arizona, Idaho, New Mexico, Montana, and Utah.
(9) 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)