Table of Contents

Tables

Source and Accuracy of the Data for the March 2002 Current Population Survey Microdata File

SOURCE OF DATA

The data in this microdata file came from the March 2002 Current Population Survey (CPS). The Census Bureau conducts the CPS every month, although this file has only March data. The March survey uses two sets of questions, the basic CPS and the supplement.

Basic CPS. The monthly CPS collects primarily labor force data about the civilian noninstitutional population. Interviewers ask questions concerning labor force participation about each member 15 years old and over in every sample household.

March Supplement. In March 2002, the interviewers asked additional questions to supplement the basic CPS questions. These additional questions covered the following topics:

- Household and Family Characteristics
- Marital Status
- Geographic Mobility
- Foreign Born Population
- Income from the previous calendar year
- Poverty
- Work Status/Occupation
- Health Insurance Coverage
- Noncash Benefits
- Educational Attainment

Basic CPS Sample Design. The present monthly CPS sample was selected from the 1990 Decennial Census files with coverage in all 50 states and the District of Columbia. The sample is continually updated to account for new residential construction. To obtain the sample, the United States was divided into 2,007 geographic areas. In most states, a geographic area consisted of a county or several contiguous counties. In some areas of New England and Hawaii, minor civil divisions are used instead of counties. These 2,007 geographic areas were then grouped into 754 strata, and one geographic area was selected from each stratum.

About 60,000 occupied households are eligible for interview every month out of the 754 strata. Interviewers are unable to obtain interviews at about 4,500 of these units. This occurs when the occupants are not found at home after repeated calls or are unavailable for some other reason.

The number of households that are eligible for interview in the basic CPS increased from 50,000 to 60,000 in July of 2001. This increase in the number of eligible households is due to the implementation of the State Children's Health Insurance Program (SCHIP) sample expansion. The SCHIP sample expansion increased the monthly CPS sample in states with high sampling errors for low-income uninsured children. With the increase in eligible households, the number of units where interviewers were unable to obtain an interview increased from 3,200 to 4,500.

March Supplement Sample. To obtain more reliable data for certain minority groups, the March Supplement sample includes 21,000 eligible housing units in addition to the 60,000 eligible housing units from the basic CPS. Included in this 21,000 housing unit increase are Hispanic households identified the previous November and following April, non-Hispanic non-White households identified the previous November, and non-Hispanic White households with children under 19 years of age identified in the previous November and following April. This March Supplement sample increase of 21,000 was first included in March 2001 for testing purposes and in March 2002 for reporting purposes.

For more information about the households eligible for the March supplement, please see Chapters 2 and 3 and Appendix J of:

> Technical Paper 63RV, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2002.

Sample Redesign. Since the introduction of the CPS, the Census Bureau has redesigned the CPS sample several times. These redesigns have improved the quality and accuracy of the data and have satisfied changing data needs. The most recent changes were phased in and implementation was completed in 1995.

Estimation Procedure. This survey's estimation procedure adjusts weighted sample results to agree with independent estimates of the civilian noninstitutional population of the United States by age, sex, race, Hispanic/non-Hispanic ancestry, and state of residence. The adjusted estimate is called the post-stratification ratio estimate. The independent estimates are calculated based on information from three primary sources:

- The 2000 Decennial Census of Population and Housing.
- Statistics on births, deaths, immigration, and emigration.
- Statistics on the size of the armed forces.

The estimation procedure for the March supplement included a further adjustment so husband and wife of a household received the same weight. The independent population estimates include some, but not all, undocumented immigrants.

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, however, is unknown.

Sampling Error. Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. This possible variation in the estimates due to sampling error is known as "sampling variability."

Nonsampling Error. All other sources of error in the survey estimates are collectively called nonsampling error. Sources of nonsampling error include the following:

- Inability to obtain information about all sample cases.
- Definitional difficulties.
- Differences in the interpretation of questions.
- Respondent inability or unwillingness to provide correct information.
- Respondent inability to recall information.
- Errors made in data collection, such as recording and coding data.
- Errors made in processing the data.
- Errors made in estimating values for missing data.
- Failure to represent all units with the sample (undercoverage).

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

Nonresponse. The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. For the March 2002 basic CPS, the nonresponse rate was 8.3%. The nonresponse rate for the March supplement was an additional 8.6%, for a total supplement nonresponse rate of 16.2%.

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. CPS undercoverage results from missed housing units and missed people within sample households. Overall CPS undercoverage is estimated to be about 8 percent. CPS undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks and other races combined than for Whites.

The Current Population Survey weighting procedure uses ratio estimation whereby sample estimates are adjusted to independent estimates of the national population by age, race, sex and Hispanic ancestry. This weighting partially corrects for bias due to 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, and Hispanic ancestry. 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, the estimated population before post-stratification divided by the independent population control. Table 1 shows CPS coverage ratios for age-sex-race groups for a typical month. The CPS coverage ratios can exhibit some variability from month to month. Other Census Bureau household surveys experience similar coverage.

Table 1. CPS Coverage Ratios

	Non-Black		Black		All People		
Age	M	F	M	F	M	F	Total
0-14	0.929	0.964	0.850	0.838	0.916	0.943	0.929
15	0.933	0.895	0.763	0.824	0.905	0.883	0.895
16-19	0.881	0.891	0.711	0.802	0.855	0.877	0.866
20-29	0.847	0.897	0.660	0.811	0.823	0.884	0.854
30-39	0.904	0.931	0.680	0.845	0.877	0.920	0.899
40-49	0.928	0.966	0.816	0.911	0.917	0.959	0.938
50-59	0.953	0.974	0.896	0.927	0.948	0.969	0.959
60-64	0.961	0.941	0.954	0.953	0.960	0.942	0.950
65-69	0.919	0.972	0.982	0.984	0.924	0.973	0.951
70+	0.993	1.004	0.996	0.979	0.993	1.002	0.998
15+	0.914	0.945	0.767	0.874	0.898	0.927	0.918
0+	0.918	0.949	0.793	0.864	0.902	0.931	0.921

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.

A number of changes were made in data collection and estimation procedures beginning with the January 1994 CPS. The major change was the use of a new questionnaire. The questionnaire was redesigned to measure the official labor force concepts more precisely, to expand the amount of data available, to implement several definitional changes, and to adapt to a computer-assisted interviewing environment. The March supplemental income questions were also modified for adaptation to computer-assisted interviewing, although there were no changes in definitions and concepts. See Appendix C of Report P-60 No. 188 on “Conversion to a Computer Assisted Questionnaire” for a description of these changes and the effect they had on the data. Due to these and other changes, one should use caution when comparing estimates from data collected before 1994 with estimates from data collected in 1994 and later.

Caution should also be used when comparing data from this microdata file, which reflects 2000 census-based population controls, with microdata files from March 1994-2001, which reflect 1990 census-based population controls. Microdata files from previous years reflect the latest available census-based population controls. Although this 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 2000 based population controls results in about a one percent increase from the 1990 based population controls in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 2002 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 subpopulation groups than for the total population.

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

Based on the results of each decennial census, the Census Bureau gradually introduces a new sample design for the CPS[1]. During this phase-in period, CPS data are collected from sample designs based on different censuses. While most CPS estimates were unaffected by this mixed sample, geographic estimates are subject to greater error and variability. Users should exercise caution when comparing estimates across years for metropolitan/ nonmetropolitan categories.

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. Even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test. Caution should also be used when interpreting results based on a relatively small number of cases. Summary measures probably do not reveal useful information when computed on a base[2] smaller than 75,000.

For additional information on nonsampling error including the possible impact on CPS data when known, refer to

- Statistical Policy Working Paper 3, *An Error Profile: Employment as Measured by the Current Population Survey*, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1978.

- Technical Paper 63RV, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2002.

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 that would include the average result of all possible samples with a known probability. For example, if all possible samples were surveyed under essentially the same general conditions and the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately

[1] For detailed information on the 1990 sample redesign, see the Department of Labor, Bureau of Labor Statistics report, *Employment and Earnings,* Volume 41 Number 5, May 1994.

[2] subpopulation

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. However, one can say with specified confidence that the interval includes the average estimate calculated from all possible samples.

Standard errors may be used to perform hypothesis testing. This is 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 Whites with a college education to the percentage of Blacks with a college education.

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 parameters 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 texts for alternative criteria.

Estimating Standard Errors. To estimate the standard error of a CPS estimate, the Census Bureau uses replicated variance estimation methods. 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 due to nonsampling error. Bias is the average over all possible samples of the differences between the sample estimates and the true value.

Generalized Variance Parameters. Consider all the possible estimates of characteristics of the population that are of interest to data users. Now consider all the subpopulations such as racial groups, age ranges, etc. Finally, consider every possible comparison or ratio combination. The list would be completely unmanageable. Similarly, a list of standard errors to go with every estimate would be unmanageable. Therefore, rather than providing an individual standard error for every possible estimate, we provide generalized variance parameters to allow for the calculation of standard errors.

Through experimentation, we have found that certain groups of estimates have similar relationships between their variances and expected values. We provide a generalized method for calculating standard errors for any of the characteristics of the population of interest. The generalized method uses generalized variance parameters for groups of estimates. These parameters are in Table 2, for basic CPS monthly labor force estimates, and Table 3, for March supplement data, including the Hispanic supplement.

Standard Errors of Estimated Numbers. The approximate standard error, s_x, of an estimated number from this microdata file can be obtained using this 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 2 or 3 associated with the particular type of characteristic. When calculating standard errors for numbers from cross-tabulations involving different characteristics, use the factor or set of parameters for the characteristic which will give the largest standard error.

For information on calculating standard errors for labor force data from the CPS which involve quarterly or yearly averages see "Explanatory Notes and Estimates of Error: Household Data" in *Employment and Earnings*, a monthly report published by the Bureau of Labor statistics.

Illustration No. 1

Suppose you want to calculate the standard error and a 90 percent confidence interval of the number of unemployed females in the civilian labor force when the number of unemployed females in the civilian labor force is about 3,773,000. Use Formula (1) and the appropriate parameters from Table 2 to get:

Number, x	3,773,000
a parameter	-0.000033
b parameter	2,693
standard error	98,000
90% conf. int.	3,612,000 to 3,934,000

where the standard error is calculated as

$$s_x = \sqrt{-0.000033 \times 3{,}773{,}000^2 + 2{,}693 \times 3{,}773{,}000} = 98{,}000$$

and the 90 percent confidence interval is calculated as 3,773,000 ± 1.645 × 98,000.

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.

Illustration No. 2

Suppose you want to calculate the standard error and a 90 percent confidence interval for the number of people aged 25 and over who held a bachelor's degree, when they numbered about 32,295,000. Use the appropriate parameters from Table 3 and Formula (1) to get:

Number, x	32,295,000
a parameter	-0.000005
b parameter	1,206
standard error	184,000
90% conf. int.	31,992,000 to 32,598,000

where the standard error is calculated as

$$s_x = \sqrt{-0.000005 \times 32{,}295{,}000^2 + 1{,}206 \times 32{,}295{,}000} = 184{,}000$$

and the 90 percent confidence interval is calculated as 32,295,000 ± 1.645 × 184,000.

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

<u>Standard Errors of Estimated Percentages</u>. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on the size of the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories, use the factor or parameter from Table 2 or 3 indicated by the numerator.

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

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

Here x is the total number of people, families, households, or unrelated individuals in the base of the percentage, p is the percentage ($0 \le p \le 100$) and b is the parameter in Table 2 or 3 associated with the characteristic in the numerator of the percentage.

<u>Illustration No. 3</u>

Suppose you want to calculate the standard error and confidence interval for the percentage of people aged 25 and over with a bachelor's degree who were Black when there were about 32,295,000 people aged 25 and over with a bachelor's degree, of which about 7.5 percent were Black. Use the appropriate parameter from Table 3 and Formula (2) to get:

Percentage, p	7.5
Base, x	32,295,000
b parameter	1,364
standard error	0.17
90% conf. int.	7.22 to 7.78

where the standard error is calculated as

$$s_{x,p} = \sqrt{\frac{1,364}{32,295,000} \times 7.5 \times 92.5} = 0.17$$

and the 90 percent confidence interval for the percentage of people aged 25 and over with a bachelor's degree who were Black is calculated as 7.5 ± 1.645 × 0.17.

Standard Error of a Difference. The standard error of the difference between two sample estimates is approximately equal to

$$s_{x-y} = \sqrt{s_x^2 + s_y^2} \qquad (3)$$

where s_x and s_y are the standard errors of the estimates, x and y. The estimates can be numbers, percentages, ratios, etc. This will represent the actual standard error quite accurately for the difference between estimates 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.

For information on calculating standard errors for labor force data from the CPS which involve differences in consecutive quarterly or yearly averages, consecutive month-to-month differences in estimates, and consecutive year-to-year differences in monthly estimates see "Explanatory Notes and Estimates of Error: Household Data" in *Employment and Earnings*, a monthly report published by the Bureau of Labor Statistics.

Illustration No. 4

Suppose you want to calculate the standard error and a 90 percent confidence interval for the difference in numbers of females and males living in the West[3] when they numbered about 32,365,000 and 32,031,000, respectively. Use the appropriate parameters from Table 3 and Formulas (2) and (3) to get:

	x	y	difference
Estimate	32,365,000	32,031,000	334,000
a parameter	-0.000014	-0.000014	-
b parameter	3,965	3,965	-
Standard error	337,000	336,000	476,000
90% conf. int.	31,811,000 to 32,919,000	31,478,000 to 32,584,000	-449,000 to 1,117,000

[3] The West region includes Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming.

where the standard error of the difference is calculated as

$$s_{x-y} = \sqrt{337{,}000^2 + 336{,}000^2} = 476{,}000$$

and the 90 percent confidence interval around the difference is calculated as
334,000 $\pm$ 1.645 × 476,000.

Since the 90 percent confidence interval contains zero, we cannot conclude, at the 10 percent significance level, that the number of females living in the West is different from the number of males.

Illustration No. 5
Suppose you want to calculate the standard error and a 90 percent confidence interval of the difference between the percentage of males and females aged 15 and over employed in agriculture (farming, forestry, and fishing). Suppose 2,391,000 of 71,565,000 employed males age 15 and over, or 3.34 percent, were employed in agriculture and about 683,000 of 63,697,000 employed females aged 15 and over, or 1.07 percent, were employed in agriculture. Use the appropriate parameters from Table 2 and Formulas (2) and (3) to get:

	x	y	difference
Percentage	3.34	1.07	2.27
Number, x	71,565,000	63,697,000	-
b parameter	2,989	2,989	-
Standard error	0.12	0.07	0.14
90% conf. int.	3.14 to 3.54	0.95 to 1.19	2.04 to 2.50

where the standard error of the difference is calculated as

$$s_{x-y} = \sqrt{0.12^2 + 0.07^2} = 0.14$$

and the 90 percent confidence interval around the difference is calculated as 2.27 ± 1.645 × 0.14.

Since this interval does not include zero, we can conclude with 90 percent confidence that the percentage of agriculturally employed females aged 15 and over is less than the percentage of agriculturally employed males aged 15 and over.

Standard Error of an Average for Grouped Data. The formula used to estimate the standard error of an average for grouped data is

$$s_{\bar{x}} = \sqrt{\frac{b}{y}\left(S^2\right)} \qquad (4)$$

In this formula, y is the size of the base of the distribution and b is a parameter from Table 2 or 3. The variance, S^2, is given by the following formula:

$$S^2 = \sum_{i=1}^{c} p_i \bar{x}_i^2 - \bar{x}^2 \qquad (5)$$

where $\bar{x}$, the average of the distribution, is estimated by

$$\bar{x} = \sum_{i=1}^{c} p_i \bar{x}_i \qquad (6)$$

c = the number of groups; i indicates a specific group, thus taking on values 1 through c.

p_i = estimated proportion of households, families or people whose values, for the characteristic (x-values) being considered, fall in group i.

$\bar{x}_i$ = $(Z_{i-1} + Z_i)/2$ where Z_{i-1} and Z_i are the lower and upper interval boundaries, respectively, for group i. $\bar{x}_i$ is assumed to be the most representative value for the characteristic for households, families, and unrelated individuals or people in group i. Group c is open-ended, i.e., no upper interval boundary exists. For this group the approximate average value is

$$\bar{x}_c = \frac{3}{2} Z_{c-1} \qquad (7)$$

<u>Standard Error of a Ratio</u>. Certain estimates may be calculated as the ratio of two numbers. The standard error of a ratio, x/y, may be computed using

$$s_{x/y} = \frac{x}{y}\sqrt{\left(\frac{s_x}{x}\right)^2 + \left(\frac{s_y}{y}\right)^2 - 2r\left(\frac{s_x s_y}{xy}\right)} \qquad (8)$$

The standard error of the numerator, s_x, and that of the denominator, s_y, may be calculated using formulas described earlier. In Formula (8), r represents the correlation between the numerator and the denominator of the estimate.

For one type of ratio, the denominator is a count of families or households and the numerator is a count of people in those families or households with a certain characteristic. If there is at least one person with the characteristic in every family or household, use 0.7 as an estimate of r. An example of this type is the average number of children per family with children.

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

Note: For estimates expressed as the ratio of x per 100 y or x per 1,000 y, multiply Formula (8) by 100 or 1,000, respectively, to obtain the standard error.

Illustration No. 6

Suppose you want to calculate the standard error and a 90 percent confidence interval for the ratio of males, x, to females, y, who make at least $50,000. Suppose there are 20,586,000 males who make at least $50,000 and about 7,244,000 females make the same, giving a ratio of x to y equal to 2.39.

Use the appropriate parameters from Table 3 to get:

	x	y	ratio
Estimate	20,586,000	7,244,000	2.84
a parameter	-0.000006	-0.000006	-
b parameter	1,249	1,249	-
Standard error	152,000	93,000	0.04
90% conf. int.	20,336,000 to 20,836,000	7,091,000 to 7,397,000	2.77 to 2.91

where the estimate of the standard error is calculated using Formula (8) and r = 0:

$$s_{x/y} = \frac{20{,}586{,}000}{7{,}244{,}000}\sqrt{\left[\frac{152{,}000}{20{,}586{,}000}\right]^2 + \left[\frac{93{,}000}{7{,}244{,}000}\right]^2} = 0.04$$

and the 90 percent confidence interval is calculated as 2.84 ± 1.645 × 0.04.

Standard Error of a Median. The sampling variability of an estimated median depends on the form of the distribution and the size of the base. One can approximate the reliability of an estimated median by determining a confidence interval about it. (See **Standard Errors and Their Use** for a general discussion of confidence intervals.)

Estimate the 68 percent confidence limits of a median based on sample data using the following procedure.

1. Determine, using Formula (2), the standard error of the estimate of 50 percent from the distribution.

2. Add to and subtract from 50 percent the standard error determined in step 1. These two numbers are the percentage limits corresponding to the 68 percent confidence about the estimated median.

3. Using the distribution of the characteristic, determine upper and lower limits of the 68 percent confidence interval by calculating values corresponding to the two points established in step 2.

Use the following formula to calculate the upper and lower limits.

$$X_{pN} = \frac{pN - N_1}{N_2 - N_1}(A_2 - A_1) + A_1 \qquad (9)$$

where

X_{pN} = estimated upper and lower bounds for the confidence interval ($0 \le p \le 1$). For purposes of calculating the confidence interval, p takes on the values determined in step 2. Note that X_{pN} estimates the median when p = 0.50.

N = for distribution of numbers: the total number of units (people, households, etc.) for the characteristic in the distribution.

= for distribution of percentages: the value 1.0.

p = the values obtained in Step 2.

A_1, A_2 = the lower and upper bounds, respectively, of the interval containing X_{pN} .

N_1, N_2 = for distribution of numbers: the estimated number of units (people, households, etc.) with values of the characteristic greater than or equal to A_1 and A_2, respectively.

= for distribution of percentages: the estimated percentage of units (people, households, etc.) having values of the characteristic greater than or equal to A_1 and A_2, respectively.

4. Divide the difference between the two points determined in step 3 by two to obtain the standard error of the median.

Note: Median incomes and their standard errors calculated as below may differ from those in published tables showing income since narrower income intervals were used in those calculations.

Illustration No. 7

Suppose you want to calculate the standard error of the median ot total money income for families with the following distribution.

Income level	Number of families	Cumulative Number of Families	Cumulative Percent of Families
Under $5,000	1,568,000	1,568,000	2.2%
$5,000 to $9,999	2,065,000	3,633,000	5.0%
$10,000 to $14,999	3,278,000	6,911,000	9.5%
$15,000 to $24,999	8,308,000	15,219,000	21.0%
$25,000 to $34,999	8,704,000	23,923,000	33.0%
$35,000 to $44,999	7,909,000	31,832,000	44.0%
$45,000 to $54,999	7,231,000	39,063,000	54.0%
$55,000 to $64,999	6,470,000	45,533,000	62.9%
$65,000 to $74,999	5,456,000	50,989,000	70.4%
$75,000 to $100,000	9,117,000	60,106,000	83.0%
$100,000 and over	12,282,000	72 ,388,000	100.0%

Total number of families . . .	72,388,000
Median income	$50,890

1. Using Formula (2) with b = 1,140, the standard error of 50 percent on a base of 72,388,000 is about 0.20 percent.

2. To obtain a 68 percent confidence interval on an estimated median, add to and subtract from 50 percent the standard error found in step 1. This yields percentage limits of 49.8 and 50.2.

3. The lower and upper limits for the interval in which the percentage limits falls are $45,000 and $55,000, respectively.

   Then, by addition, the estimated numbers of families with an income greater than or equal to $45,000 and $55,000 are 40,556,000 and 33,325,000, respectively.

   Using Formula (9), the upper limit for the confidence interval of the median is found to be about

$$\frac{0.498 \times 72{,}388{,}000 - 40{,}556{,}000}{33{,}325{,}000 - 40{,}556{,}000}(55{,}000 - 45{,}000) + 45{,}000 = 51{,}230$$

Similarly, the lower limit is found to be about

$$\frac{0.502 \times 72{,}388{,}000 - 40{,}556{,}000}{33{,}325{,}000 - 40{,}556{,}000}(55{,}000 - 45{,}000) + 45{,}000 = 50{,}830$$

Thus, a 68 percent confidence interval for the median income for families is from $50,830 to $51,230.

4. The standard error of the median is, therefore,

$$\frac{51{,}230 - 50{,}830}{2} = 200$$

Standard Error of Estimated Per Capita Deficit. Certain average values in this report represent the per capita deficit for households of a certain class. The average per capita deficit is approximately equal to

$$x = \frac{hm}{p} \qquad (10)$$

where

h = number of households in the class
m = average deficit for households in the class
p = number of people in households in the class
x = average per capita deficit of people in households in the class.

To approximate standard errors for these averages, use the formula

$$s_x = \frac{hm}{p}\sqrt{\left(\frac{s_m}{m}\right)^2 + \left(\frac{s_p}{p}\right)^2 + \left(\frac{s_h}{h}\right)^2 - 2r\left(\frac{s_p}{p}\right)\left(\frac{s_h}{h}\right)} \qquad (11)$$

In Formula (11), r represents the correlation between p and h.

For one type of average, the class represents households containing a fixed number of people. For example, h could be the number of three-person households. In this case, there is an exact correlation between the number of people in households and the number of households. Therefore, r = 1 for such households.

For other types of averages, the class represents households of other demographic types, for example, households in distinct regions, households in which the householder is of a certain age group, and owner-occupied and tenant-occupied households. In this and other cases in which the correlation between p and h is not perfect, use 0.7 as an estimate of r.

Accuracy of State Estimates. The redesign of the CPS following the 1980 census provided an opportunity to increase efficiency and accuracy of state data. All strata are now defined within state boundaries. The sample is allocated among the states to produce state and national estimates with the required accuracy while keeping total sample size to a minimum. Improved accuracy of state data was achieved with about the same sample size as in the 1970 design.

Since the CPS is designed to produce both state and national estimates, the proportion of the total population sampled and the sampling rates differ among the states. In general, the smaller the population of the state the larger the sampling proportion. For example, in Vermont approximately 1 in every 250 households is sampled each month. In New York the sample is about 1 in every 2,000 households. Nevertheless, the size of the sample in New York is four times larger than in Vermont because New York has a larger population.

Computation of Standard Errors for State Estimates. Standard errors for a state may be obtained by computing national standard errors, using formulas described earlier, and multiplying these by the appropriate f factor from Table 4. An alternative method for computing standard errors for a state is to multiply the a and b parameters in Table 2 or 3 by f^2 and then use these adjusted parameters in the standard error formulas.

Illustration No. 8

Suppose you want to calculate the standard error for the percentage of people 25 years old and over living in the state of New York who had completed a bachelor's degree or more. Suppose about 3,607,300 (26.3 percent) people had completed at least a bachelor's degree when there were about 13,716,000 people aged 18 and over living in New York. Following the first method mentioned above, use the appropriate parameter from Table 3 and Formula (2) to get:

Percentage, p	26.3
Base, x	13,716,000
b parameter	1,206
Standard error	0.41

Table 4 shows the f factor for New York to be 1.01. Thus, the standard error on the estimate of the percentage of people 18 and older in New York state who had completed college is approximately 1.01 × 0.41 = 0.41.

Following the alternative method mentioned above, obtain the needed state parameter by multiplying the parameter in Table 3 by the f ^{2} factor in Table 4 for the state of interest. For example, for educational attainment for total or white in New York this gives b = 1,206 × 1.02 = 1,230. The standard error of the estimate of the percentage of people 18 and older in New York state who had completed college can then be found by using formula (2), the base of 13,716,000 and the new b parameter, 1,230. This gives a standard error of 0.42. Differences are due to rounding.

Computation of a Factor for Groups of States. The factor adjusting standard errors for a group of states may be obtained by computing a weighted sum of the squared factors for the individual states in the group and taking the square root of the result. Depending on the combination of states, the resulting figure can be an overestimate.

The squared factor for a group of n states is given by

$$f^2 = \frac{\sum_{i=1}^{n} POP_i \times f_i^2}{\sum_{i=1}^{n} POP_i} \qquad (12)$$

where POP_i is the state population and f_i^2 is obtained from Table D. The 2001 civilian noninstitutionalized population from the CPS for each state is also given in Table D.

Illustration No. 9

Suppose the f^2 factor for the state group Illinois-Indiana-Michigan was required. The appropriate factor would be:

$$f^2 = \frac{9{,}612{,}000 \times 1.09 + 4{,}760{,}000 \times 0.90 + 7{,}791{,}000 \times 1.00}{9{,}612{,}000 + 4{,}760{,}000 + 7{,}791{,}000} = 1.02$$

Multiply the a and b parameters by f^2, 1.02, to obtain parameters for the state group, or use the original parameters and multiply the resulting standard errors by f, 1.01.

Computation of Standard Errors for Data for Combined Years. Sometimes estimates for multiple years are combined to improve precision. For example, suppose $\bar{x}$ is an average derived from n consecutive years' data, i.e., $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$ where the x_i are the estimates for the individual years.

Use the formulas described previously to estimate the standard error, s_x , of each year's estimate. Then the standard error of $\bar{x}$, $s_{\bar{x}}$, is

$$s_{\bar{x}} = \frac{s_x}{n} \qquad (13)$$

where

$$s_x = \sqrt{\sum_{i=1}^{n} s^2_{x_i} + 2r\sum_{i=1}^{n-1} s_{x_i} s_{x_{i+1}}} \qquad (14)$$

The correlation between consecutive years, r, is 0.35 for non-Hispanic households and 0.55 for Hispanic households. Correlation between nonconsecutive years is zero. The correlations were derived for income estimates but they can be used for other types of estimates where the year-to-year correlation between identical households is high.

<u>Illustration No. 10</u>

Suppose you want to calculate the standard error of the average number of children under the age of 18 without health insurance for 1997-2000 when the average is 9,541,000 and the standard errors for the individual years are 95,000, 139,000, and 153,000.

Using Formula (14), the standard error for the three years combined data is:

$$s_x = \sqrt{95{,}000^2 + 139{,}000^2 + 153{,}000^2 + (2\times0.35\times95{,}000\times139{,}000) + (2\times0.35\times139{,}000\times153{,}000)}$$

$$= 275{,}000$$

Therefore, the standard error of the average, using Formula (11), is

$$s_{\bar{x}} = \frac{275{,}000}{3} = 92{,}000.$$

Table 2. Parameters for Computation of Standard Errors for Labor Force Characteristics: March 2002

Characteristic	a	b
Labor Force and Not In Labor Force Data Other than Agricultural Employment and Unemployment		
Total or White	-0.000008	1,586
Men	-0.000035	2,927
Women	-0.000033	2,693
Both sexes, 16 to 19 years	-0.000244	3,005
Black	-0.000154	3,296
Men	-0.000336	3,332
Women	-0.000282	2,944
Both sexes, 16 to 19 years	-0.001531	3,296
Hispanic Ancestry	-0.000187	3,296
Men	-0.000363	3,332
Women	-0.000380	2,944
Both sexes, 16 to 19 years	-0.001822	3,296
Unemployment		
Total or White	-0.000017	3,005
Men	-0.000035	2,927
Women	-0.000033	2,693
Both sexes, 16 to 19 years	-0.000244	3,005
Black	-0.000154	3,296
Men	-0.000336	3,332
Women	-0.000282	2,944
Both sexes, 16 to 19 years	-0.001531	3,296
Hispanic Ancestry	-0.000187	3,296
Men	-0.000363	3,332
Women	-0.000380	2,944
Both sexes, 16 to 19 years	-0.001822	3,296
Agricultural Employment	0.001345	2,989

NOTE: These parameters are to be applied to basic CPS monthly labor force estimates.

For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics.

Table 3. a and b Parameters for Standard Error Estimates for People and Families: March 2002

Characteristics	Total or White		Black		Hispanic	
	a	b	a	b	a	b
PEOPLE						
Educational Attainment	-0.000005	1,206	-0.000052	1,364	-0.000035	922
Employment Characteristics	-0.000008	1,586	-0.000154	3,296	-0.000187	3,296
People by Family Income	-0.000011	2,494	-0.000110	2,855	-0.000109	2,855
Income	-0.000006	1,249	-0.000055	1,430	-0.000054	1,430
Health Insurance	-0.000004	1,115	-0.000038	1,354	-0.000027	997
Marital Status, Household and Family Characteristics						
Some household members	-0.000009	2,652	-0.000106	3,809	-0.000102	3,809
All household members	-0.000011	3,222	-0.000156	5,617	-0.000150	5,617
Mobility Characteristics (Movers)						
Educational Attainment, Labor Force, Marital Status, Household, Family, and Income	-0.000005	1,460	-0.000041	1,460	-0.000039	1,460
US, County, State, Region or MSA	-0.000014	3,965	-0.000110	3,965	-0.000106	3,965
Below Poverty						
Total	-0.000019	5,282	-0.000147	5,282	-0.000141	5,282
Male	-0.000038	5,282	-0.000317	5,282	-0.000269	5,282
Female	-0.000037	5,282	-0.000274	5,282	-0.000279	5,282
Age						
Under 15	-0.000067	4,072	-0.000413	4,072	-0.000367	4,072
Under 18	-0.000056	4,072	-0.000348	4,072	-0.000287	4,072
15 and over	-0.000024	5,282	-0.000203	5,282	-0.000201	5,282
15 to 24	-0.000051	1,998	-0.000345	1,998	-0.000197	1,998
25 to 44	-0.000024	1,998	-0.000191	1,998	-0.000112	1,998
45 to 64	-0.000031	1,998	-0.000285	1,998	-0.000124	1,998
65 and over	-0.000059	1,998	-0.000713	1,998	-0.000377	1,998
Unemployment	-0.000017	3,005	-0.000154	3,296	-0.000187	3,296
FAMILIES, HOUSEHOLDS, OR UNRELATED INDIVIDUALS						
Income	-0.000005	1,140	-0.000048	1,245	-0.000047	1,245
Marital Status, Household and Family Characteristics, Educational Attainment, Population by Age and/or Sex	-0.000005	1,052	-0.000037	952	-0.000036	952
Poverty	+0.000052	1,243	+0.000052	1,243	+0.000052	1,243

NOTE: These parameters are to be applied to March supplemental data including the Hispanic supplement.

For nonmetropolitan characteristics multiply a and b parameters by 1.5. If the characteristic of interest is total state population, not subtotaled by race or ancestry, the a and b parameters are zero.

For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics.

Table 4. Factors for State Standard Errors and Parameters and State Populations: 2002

State	f	f^2	Population
Alabama	0.95	0.90	3,378,000
Alaska	0.35	0.12	450,000
Arizona	1.11	1.24	3,926,000
Arkansas	0.79	0.62	2,030,000
California	1.28	1.63	25,334,000
Colorado	0.83	0.69	3,344,000
Connecticut	0.73	0.54	2,670,000
Delaware	0.41	0.17	609,000
District of Columbia	0.37	0.14	444,000
Florida	1.08	1.16	12,806,000
Georgia	1.28	1.65	6,224,000
Hawaii	0.50	0.25	906,000
Idaho	0.55	0.30	978,000
Illinois	1.04	1.09	9,600,000
Indiana	0.95	0.90	4,755,000
Iowa	0.71	0.51	2,233,000
Kansas	0.69	0.48	2,088,000
Kentucky	0.89	0.80	3,096,000
Louisiana	1.00	1.01	3,256,000
Maine	0.45	0.20	1,056,000
Maryland	0.95	0.90	4,040,000
Massachusetts	0.95	0.91	5,072,000
Michigan	1.00	1.00	7,783,000
Minnesota	0.90	0.81	3,934,000
Mississippi	0.84	0.70	2,102,000
Missouri	0.98	0.96	4,283,000
Montana	0.48	0.23	701,000
Nebraska	0.58	0.34	1,301,000
Nevada	0.61	0.37	1,602,000
New Hampshire	0.45	0.21	1,004,000
New Jersey	0.96	0.91	6,780,000
New Mexico	0.72	0.52	1,365,000
New York	1.01	1.02	14,708,000
North Carolina	1.05	1.09	6,133,000
North Dakota	0.35	0.12	504,000
Ohio	1.04	1.08	8,888,000
Oklahoma	0.83	0.70	2,604,000
Oregon	0.82	0.68	2,691,000
Pennsylvania	1.00	1.00	9,653,000
Rhode Island	0.40	0.16	824,000
South Carolina	0.89	0.79	3,074,000
South Dakota	0.36	0.13	588,000
Tennessee	1.13	1.28	4,413,000
Texas	1.22	1.50	15,514,000
Utah	0.68	0.46	1,612,000
Vermont	0.33	0.11	498,000
Virginia	1.13	1.29	5,361,000
Washington	1.08	1.16	4,572,000
West Virginia	0.56	0.32	1,425,000
Wisconsin	0.91	0.83	4,230,000
Wyoming	0.32	0.10	382,000

NOTE: For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics.