Appendix C. Source and Accuracy of Estimates

SOURCE OF DATA

Most estimates in this report come from data obtained in October 1989 from the Current Population Survey (CPS). Some estimates are based on data obtained from the CPS in earlier years. The Bureau of the Census conducts the survey every month, although this report uses only October data for its estimates. The October survey uses two sets of questions, the basic CPS and the supplement.

Basic CPS. The basic 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.

The October 1989 CPS sample was selected from the 1980 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. It is located in 729 areas comprising 1,973 counties, independent cities, and minor civil divisions. About 56,100 occupied households are eligible for interview every month. Interviewers are unable to obtain interviews at about 2,500 of these units because the occupants are not home after repeated calls or are unavailable for some other reason.

Since the introduction of the CPS, the Bureau of the Census has redesigned the CPS sample several times to improve the quality and reliability of the data and to satisfy changing data needs. The most recent changes were completely implemented in July 1985.

The table in the next column summarizes changes in the CPS designs for the years for which data appear in this report.

October Supplement. In addition to the basic CPS questions, interviewers asked supplementary questions in October about school enrollment for all household members 3 years old and over.

Estimation Procedure. This survey's estimation procedure inflates weighted sample results to independent estimates of the civilian noninstitutional population of the United States by age, sex, race, and Hispanic/non-Hispanic categories. The independent estimates were based on statistics from decennial censuses of population; statistics on births, deaths, immigration and emigration; and statistics on the size of the Armed Forces. The independent population estimates used for 1981 (1980 for income estimates) to present were based on updates to controls established by the 1980 Decennial Census. Data previous to 1981 were based on independent population estimates from the most recent decennial census. For more details on the change in independent estimates, see the section entitled "Introduction of 1980 Census Population Controls" in an earlier report (Series P-60, No. 133).

Description of the October Current Population Survey

Time period	Number of sample areas	Housing units eligible	
		Interviewed	Not Interviewed
1988 to 1989	729	53,600	2,500
1985 to 1987	729	57,000	2,500
1984	[1]629/729	57,000	2,500
1981 to 1983	629	59,000	2,500
1980	629	65,500	3,000
1978 to 1979	614	55,000	3,000
1972 to 1977	461	46,500	2,500
1971	449	45,000	2,000
1967 to 1970	449	48,000	2,000
1963 to 1966	357	33,500	1,500
1960 to 1962	333	33,500	1,500
1957 to 1959	330	33,500	1,500
1954 to 1956	230	21,000	500-1,000
1947 to 1953	68	21,000	500-1,000

[1]The CPS was redesigned following the 1980 Decennial Census of Population and Housing. During phase-in of the new design, housing units from the new and old designs were in the sample.

The estimates in this report for 1985 and later also employ a revised survey weighting procedure for persons of Hispanic origin. In previous years, weighted sample results were inflated to independent estimates of the noninstitutional population by age, sex, and race. There was no specific control of the survey estimates for the Hispanic population. Since then, the Bureau of the Census developed independent population controls for the Hispanic population by sex and detailed age groups. Revised weighting procedures incorporate these new controls. The independent population estimates include some, but not all, undocumented immigrants.

ACCURACY OF THE ESTIMATES

Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. A sample survey estimate has two possible types of errors: nonsampling and sampling. The accuracy of an estimate depends on both types of errors, but the full extent of the nonsampling error is unknown. Consequently, one should be particularly careful when interpreting results based on a relatively small number of cases or on small differences between estimates. The standard errors for CPS estimates primarily indicate the magnitude of sampling error. They also partially measure the effect of some nonsampling errors in responses and enumeration, but do not measure systematic biases in the data. (Bias is the average over all possible samples of the differences between the sample estimates and the desired value.)

Nonsampling Variability. Nonsampling errors can be attributed to many sources. These sources include the inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, respondents' inability or unwillingness to provide correct information or to recall information, errors made in data collection such as in recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, and failure to represent all units with the sample (undercoverage).

CPS undercoverage results from missed housing units and missed persons within sample households. Compared to the level of the 1980 Decennial Census, overall CPS undercoverage is about 7 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. As described previously, ratio estimation to independent age-sexrace-Hispanic population controls partially corrects for the bias due to undercoverage. However, biases exist in the estimates to the extent that missed persons in missed households or missed persons in interviewed households have different characteristics from those of interviewed persons in the same age-sex-raceHispanic group. Furthermore, the independent population controls have not been adjusted for undercoverage in the 1980 census.

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 and Technical Paper 40, *The Current Population Survey: Design and Methodology,* Bureau of the Census, U.S. Department of Commerce.

Sampling Variability. Sampling variability is variation that occurred by chance because a sample was surveyed rather than the entire population. Standard errors, as calculated by methods described later in *"Standard Errors and Their Use,"* are primarily measures of sampling variability, although they may include some nonsampling error.

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. Use caution when comparing results from different sources.

Caution should also be used when comparing estimates in this report, which reflect 1980 census-based population controls, with estimates for 1980 and earlier years, which reflect 1970 censusbased population controls. This change in population controls had relatively little impact on summary measures such as means, medians, and percentage distributions, but did have a significant impact on levels. For example, use of 1980 based population controls results in about a 2-percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 1981 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.

Since no independent population control totals for persons of Hispanic origin were used before 1985, compare Hispanic estimates over time cautiously.

Note When Using Small Estimates. Summary measures (such as medians and percentage distributions) are shown only when the base is 75,000 or greater. Because of the large standard errors involved, summary measures would probably not reveal useful information when computed on a smaller base. However, estimated numbers are shown even though the relative standard errors of these numbers are larger than those for corresponding percentages. These smaller estimates permit combinations of the categories to suit data users' needs. Take care in the interpretation of small differences. For instance, even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Standard Errors and Their Use. A number of approximations are required to derive, at a moderate cost, standard errors applicable to all the estimates in this report. Instead of providing an individual standard error for each estimate, generalized sets of standard errors are provided for various types of characteristics. Thus, the tables show levels of magnitude of standard errors rather than the precise standard errors.

The sample estimate and its standard error enable one to construct a confidence interval, 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 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 1.6 standard errors below the estimate to 1.6 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.

Some statements in the report may contain estimates followed by a number in parentheses. This number can be added to and subtracted from the estimate to calculate upper and lower bounds of the 90-percent confidence interval. For example, if a statement contains the phrase "grew by 1.7 percent (±1.0)," the 90-percent confidence interval for the estimate, 1.7 percent, is 0.7 percent to 2.7 percent.

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 appearing in this report is that the population parameters are different. An example of this would be comparing the percentage of 18 to 24 year old high school graduates in 1989 to that in 1979.

Tests may be performed at various levels of significance, where a significance level is the probability of concluding that the characteristics are different when, in fact, they are the same. All statements of comparison in the text have passed a hypothesis test at the 0.10 level of significance or better. This means that the absolute value of the estimated difference between characteristics is greater than or equal to 1.6 times the standard error of the difference.

Standard Errors of Estimated Numbers. There are two ways to compute the approximate standard error, s_x, of an estimated number shown in this report. The first uses the formula

$$s_x = fs$$

where f is the appropriate factor from table C-4 and s is the standard error of the estimate obtained by interpolation from table C-1 or C-2. The second method uses formula (2), from which the standard errors in tables B-1 and B-2 were calculated. This formula will provide more accurate results than formula (1).

$$s_x = \sqrt{-(b/T)X^2 + bx}$$

Here x is the size of the estimate, T is the total number of persons in a specific age group and b is the parameter in table C-4 associated with the particular type of characteristic. If T is not known, for Total or White use 100,000,000; for Blacks and Hispanic use 10,000,000. 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.

Illustration. Table 1 shows there were 2,898,000 3 and 4 year olds enrolled in school and 7,405,000 children in that age group in October 1989. Using formula (1) with f = 1.1 from table C-4, and s = 67,000 from table C-1, the approximate standard error of the number of 3 and 4 year olds enrolled in school is

$$s_x = (1.1)(67{,}000) = 74{,}000$$

The value of s was obtained by linear interpolation in two directions. The first interpolation was between 5,000,000 and 10,000,000 total persons for both 2,000,000 and 3,000,000 estimated number of persons. The value for 2,000,000 estimated persons was 62.4 and for 3,000,000 estimated persons was 67.1. The second interpolation was between these two values to get the value corresponding to 2,898,000 persons.

Using the second method with b = 3,203 from table C-4, the approximate standard error is

$$s_x = \sqrt{-\frac{3{,}203}{7{,}405{,}000}(2{,}898{,}000)^2 + (3{,}203)(2{,}898{,}000)} = 75{,}000$$

The 90-percent confidence interval for this estimate is from 2,778,000 to 3,018,000, i.e., 2,898,000 ± 1.6(75,000). Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

Standard Errors of Estimated Percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on 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 C-4 indicated by the numerator.

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

$$s_{x,p} = fs$$

In this formula, f is the appropriate factor from table C-4, and s is the standard error of the estimate obtained by interpolation from table C-3.

Alternatively, formula (4) will provide more accurate results. The standard errors in table C-3 were calculated with this formula.

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

Here x is the total number of persons, families, households, or unrelated individuals in the base of the percentage, p is the percentage ($0 \leq p \leq 100$), and b is the parameter in table C-4 associated with the characteristic in the numerator of the percentage.

Illustration. As shown in Table 1, of the 14,189,000 persons aged 18 to 21, 39.7 percent were enrolled in college in 1989. Using formula (3) with f = 1.0, and s = 0.7 interpolating from table C-3, the approximate standard error is

$$s_{x,p} = (1.0)(0.7) = 0.7$$

Using the alternate method with b = 2,814 from table C-4, the approximate standard error on an estimate of the percent of 18 to 21 year olds enrolled in school is

$$s_{x,p} = \sqrt{\frac{2{,}814}{14{,}189{,}000}(39.7)(100.0-39.7)} = 0.7$$

The 90-percent confidence interval for the estimated percentage of persons aged 18 to 21 in 1989 enrolled in college is from 38.6 to 40.8 percent, i.e., 39.7 ± 1.6(0.7).

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

$$s_{x-y} = \sqrt{s_x^2x + s_y^2}$$

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 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. In October 1989, of the 6,995,000 males 18 to 21 years old, 37.9 percent were enrolled in college and of the 7,194,000 females of the same age group, 41.5 percent were enrolled in college. The apparent difference is 3.6 percent. Using formula (4) and b = 2,814 from table C-4, the standard error on the estimate of 37.9 percent is 1.0. Using formula (4) and b = 2,814 from table C-4, the standard error on the estimate of 41.5 percent is 1.0.

Therefore using formula (5) the standard error of the estimated difference of 3.6 percent is about:

$$s_{x-y} = \sqrt{(1.0)^2 + (1.0)^2} = 1.4$$

This means that the 90-percent confidence interval around the difference is from 1.4 to 5.8 percent, i.e., 3.6 ± 1.6(1.4). Since the interval does not contain zero, we can conclude with 90-percent confidence that the proportion of females aged 18 to 21 enrolled in college is greater than that for males.

Table C-1. **1989 Standard Errors of Estimated Numbers: Total or White**

(Numbers in thousands)

Size of estimate	Total persons in age group									
	100	250	500	1,000	2,500	5,000	10,000	25,000	50,000	100,000
10	5.0	5.2	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3
20	6.7	7.2	7.4	7.4	7.5	7.5	7.5	7.5	7.5	7.5
30	7.7	8.6	8.9	9.0	9.1	9.2	9.2	9.2	9.2	9.2
40	8.2	9.7	10.2	10.4	10.5	10.6	10.6	10.6	10.6	10.6
50	8.4	10.6	11.3	11.6	11.7	11.8	11.8	11.8	11.9	11.9
75	7.3	12.2	13.4	14.0	14.3	14.4	14.5	14.5	14.5	14.5
100	(X)	13.0	15.0	15.9	16.4	16.6	16.7	16.7	16.8	16.8
200	(X)	10.6	18.4	21.2	22.8	23.2	23.5	23.6	23.7	23.7
300	(X)	(X)	18.4	24.3	27.3	28.2	28.6	28.9	29.0	29.0
400	(X)	(X)	15.0	26.0	30.7	32.2	32.9	33.3	33.4	33.5
500	(X)	(X)	(X)	26.5	33.5	35.6	36.6	37.1	37.3	37.4
750	(X)	(X)	(X)	23.0	38.4	42.4	44.2	45.2	45.6	45.8
1,000	(X)	(X)	(X)	(X)	41.1	47.4	50.3	52.0	52.5	52.8
2,000	(X)	(X)	(X)	(X)	33.5	58.1	67.1	72.0	73.5	74.3
3,000	(X)	(X)	(X)	(X)	(X)	58.1	76.9	86.2	89.1	90.5
4,000	(X)	(X)	(X)	(X)	(X)	47.4	82.2	97.2	101.8	104.0
5,000	(X)	(X)	(X)	(X)	(X)	(X)	83.9	106.1	112.5	115.6
7,500	(X)	(X)	(X)	(X)	(X)	(X)	72.6	121.5	133.9	139.7
10,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	129.9	150.0	159.1
20,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	106.1	183.8	212.2
30,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	183.8	243.1
40,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	150.0	259.9
50,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	265.2
75,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	229.7

Note: For a particular characteristic, see table C-4 for the appropriate factor to apply to the above standard errors.

X Not applicable

Table C-2. **1989 Standard Errors of Estimated Numbers: Black or Other Races and Hispanic**

(Numbers in thousands)

Size of estimate	Total persons in age group						
	100	250	500	1,000	2,500	5,000	10,000
Black or Other Races							
10	5.9	6.0	6.1	6.1	6.2	6.2	6.2
20	7.8	8.4	8.5	8.6	8.7	8.7	8.7
30	8.9	10.0	10.4	10.5	10.6	10.7	10.7
40	9.6	11.3	11.8	12.1	12.2	12.3	12.3
50	9.8	12.3	13.1	13.4	13.7	13.7	13.8
75	8.4	14.1	15.6	16.3	16.6	16.8	16.8
100	(X)	15.1	17.5	18.5	19.1	19.3	19.4
200	(X)	12.3	21.4	24.7	26.5	27.0	27.3
300	(X)	(X)	21.4	28.3	31.7	32.8	33.3
400	(X)	(X)	17.5	30.2	35.8	37.4	38.2
500	(X)	(X)	(X)	30.9	39.0	41.4	42.5
750	(X)	(X)	(X)	26.7	44.7	49.3	51.4
1,000	(X)	(X)	(X)	(X)	47.8	55.2	58.5
2,000	(X)	(X)	(X)	(X)	39.0	67.6	78.0
3,000	(X)	(X)	(X)	(X)	(X)	67.6	89.4
4,000	(X)	(X)	(X)	(X)	(X)	55.2	95.6
5,000	(X)	(X)	(X)	(X)	(X)	(X)	97.6
7,500	(X)	(X)	(X)	(X)	(X)	(X)	84.5
Hispanic							
10	8.2	8.5	8.6	8.6	8.6	8.7	8.7
20	11.0	11.8	12.0	12.1	12.2	12.2	12.2
30	12.6	14.1	14.6	14.8	14.9	15.0	15.0
40	13.4	15.9	16.6	17.0	17.2	17.3	17.3
50	13.7	17.3	18.4	18.9	19.2	19.3	19.3
75	11.9	19.9	21.9	22.8	23.4	23.6	23.6
100	(X)	21.2	24.5	26.0	26.8	27.1	27.3
200	(X)	17.3	30.0	34.7	37.2	38.0	38.4
300	(X)	(X)	30.0	39.7	44.5	46.0	46.7
400	(X)	(X)	24.5	42.5	50.2	52.6	53.7
500	(X)	(X)	(X)	43.3	54.8	58.1	59.7
750	(X)	(X)	(X)	37.5	62.8	69.2	72.2
1,000	(X)	(X)	(X)	(X)	67.1	77.5	82.2
2,000	(X)	(X)	(X)	(X)	54.8	94.9	109.6
3,000	(X)	(X)	(X)	(X)	(X)	94.9	125.6
4,000	(X)	(X)	(X)	(X)	(X)	77.5	134.2
5,000	(X)	(X)	(X)	(X)	(X)	(X)	137.0
7,500	(X)	(X)	(X)	(X)	(X)	(X)	118.7

Note: For a particular characteristic, see table C-4 for the appropriate factor to apply to the above standard errors.

X Not applicable

Table C-3. **1989 Standard Errors of Estimated Percentages**

Base of percentage (thousands)	Estimated percentage				
	2 or 98	5 or 95	10 or 90	25 or 75	50
Total or White					
75	2.7	4.2	5.8	8.4	9.7
100	2.3	3.7	5.0	7.3	8.4
250	1.5	2.3	3.2	4.6	5.3
500	1.1	1.6	2.3	3.2	3.8
1,000	0.7	1.2	1.6	2.3	2.7
2,500	0.5	0.7	1.0	1.5	1.7
5,000	0.3	0.5	0.7	1.0	1.2
10,000	0.2	0.4	0.5	0.7	0.8
25,000	0.1	0.2	0.3	0.5	0.5
50,000	0.1	0.2	0.2	0.3	0.4
100,000	0.1	0.1	0.2	0.2	0.3
Black or Other Races					
25	5.5	8.5	11.7	16.9	19.5
50	3.9	6.0	8.3	11.9	13.8
75	3.2	4.9	6.8	9.8	11.3
100	2.7	4.3	5.9	8.4	9.8
250	1.7	2.7	3.7	5.3	6.2
500	1.2	1.9	2.6	3.8	4.4
1,000	1.0	1.6	2.1	3.1	3.6
2,500	0.5	0.9	1.2	1.7	2.0
5,000	0.4	0.6	0.8	1.2	1.4
10,000	0.3	0.4	0.6	0.8	1.0
20,000	0.2	0.3	0.4	0.6	0.7
Hispanic					
25	7.7	11.9	16.4	23.7	27.4
50	5.4	8.4	11.6	16.8	19.4
75	4.4	6.9	9.5	13.7	15.8
100	3.8	6.0	8.2	11.9	13.7
250	2.4	3.8	5.2	7.5	8.7
500	1.7	2.7	3.7	5.3	6.1
1,000	1.4	2.2	3.0	4.3	5.0
2,500	0.8	1.2	1.6	2.4	2.7
5,000	0.5	0.8	1.2	1.7	1.9
10,000	0.4	0.6	0.8	1.2	1.4
20,000	0.3	0.4	0.6	0.8	1.0

Note: For a particular characteristic, see table B-4 for the appropriate factor to apply to the above standard erxrors.

Table C-4. **1989 Standard Error Parameters for School Enrollment**

Characteristic	Total or White		Black		Hispanic	
	b	f	b	f	b	f
Persons Enrolled in School:						
Total	2,814	1.0	3,807	1.0	7,509	1.0
Children 13 and under	3,203	1.1	3,203	0.9	3,203	0.6
Marital Status	5,319	1.4	7,631	1.4	15,072	1.4
Household Characteristics:						
Head, Wife, or Primary Individual	2,111	0.9	1,907	0.7	3,767	0.7
Child or Other Relative in Primary Family, Secondary Family Member	5,319	1.4	7,631	1.4	15,072	1.4
Income, Earnings	2,505	0.9	2,864	0.9	5,657	0.9
Employment Status, Occupation:						
Both Sexes	2,762	1.0	2,762	0.9	2,912	0.6
Male	2,390	0.9	2,390	0.8	2,390	0.6
Female	2,048	0.9	2,048	0.7	2,048	0.5

Notes: To estimate standard errors for school enrollment prior to 1989, multiply the standard error for 1989 by the appropriate factor in table C-5.
The b parameters should be multiplied by 1.5 for nonmetropolitan residence categories.
The b parameters should be multiplied by 1.91 for farm parameters.
For regional data, multiply the standard error calculated by the appropriate factor in table C-6.

Table C-5. **Factors to Calculate School Enrollment Standard Errors Prior to 1989**

Year	Total or White	Black	Hispanic
1988	1.00	1.00	1.00
1985-1987	0.92	0.92	0.85
1982-1984	0.92	0.92	0.59
1977-1981	0.87	0.87	0.56
1967-1976	0.86	0.86	0.55
1957-1966	1.06	1.06	(X)
Before 1956	1.30	1.30	(X)

Notes: Apply the appropriate factor to the standard error calculated for 1989.

Use the Total or White factors to calculate standard errors for children 13 and under.

X Not applicable

Table C-6. **Regional Factors to Apply to 1989 Standard Errors**

Characteristic	Factor
U.S. Totals	1.00
Region	
Northeast	0.90
Midwest	0.94
South	0.95
West	1.16

Note: Multiply standard errors obtained using tables C-1 through C-4 by these factors.

☆ U.S. Government Printing Office : 1991 - 311-883/40072