Appendix C. Source and Reliability of the Estimates

SOURCE OF DATA

Most of the estimates in this report are based on data obtained in October 1983 from the Current Population Survey (CPS) conducted by the Bureau of the Census and from supplementary questions to the CPS. Some estimates are based on data obtained from the CPS in earlier years. The monthly CPS deals mainly with labor force data for the civilian noninstitutional population. Questions relating to labor force participation are asked about each member in every sample household. In addition, supplementary questions regarding school enrollment are asked about all eligible household members 3 years old and over each October.

The present CPS sample was initially selected from the 1970 Census file and is updated continuously, where possible, to reflect new construction. The October 1983 sample was spread over 629 areas comprising 1,148 counties, independent cities, and minor civil divisions with coverage in each of the 50 States and the District of Columbia. The sample is composed of approximately 59,500 occupied households that are eligible for interview. Of this number, about 2,500 occupied units were visited, but interviews were not obtained because the occupants were not found at home after repeated calls or were unavailable for some other reason. The following table provides a description of some aspects of the CPS sample designs in use during the referenced data collection periods.

Description of the October Current Population Survey

Time period	Number of sample areas[1]	Housing units eligible	
		Interviewed	Not interviewed
October 1981 to 1983.	629	58,000	2,500
October 1980.....	629	63,000	3,000
October 1978 to 1979.	614	45,000	2,500
October 1972 to 1977.	461	45,000	2,000
October 1971.....	449	48,000	2,000
October 1967 to 1970.	449	48,000	2,000
October 1965 to 1966.	357	33,500	1,500
October 1960.....	[2]333	33,500	1,500
October 1955.....	230	21,000	500-1,000
October 1947 to 1950.	68	21,000	500-1,000

[1]Beginning in May 1956, these areas were chosen to provide coverage in each State and the District of Columbia.
[2]Three sample areas were added in 1960 to represent Alaska and Hawaii after statehood.

The estimation procedure used in this survey involves the inflation of the weighted sample results to independent estimates of the total civilian noninstitutional population of the United States by age, race, and sex. These independent estimates are based on statistics from the decennial censuses of population; statistics on births, deaths, immigration and emigration; and statistics on the strength of the Armed Forces. The independent population estimates used in this report to obtain data for 1981 and later are based on the 1980 Decennial Census. But, some of the data in this report for 1981 were also obtained using independent population estimates based on the 1970 Decennial Census. Any display of 1981 data using 1970 census-based population estimates is so noted in the report. In earlier reports in this series (P-20), data for 1972 through 1980 were obtained using independent population estimates based on the 1970 Decennial Census. Estimates for earlier years were based on earlier censuses. For more details on this change, see the section in Appendix B entitled, "Introduction of 1980 Census Population Controls."

RELIABILITY OF THE ESTIMATES

Since the CPS estimates were based on a sample, they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaires, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: sampling and nonsampling. The accuracy of a survey result depends on both types of errors, but the full extent of the nonsampling error is unknown. Consequently, particular care should be exercised in the interpretation of figures based on a relatively small number of cases or on small differences between estimates. The standard errors provided for the CPS estimates primarily indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in responses and enumeration, but do not measure any systematic biases in the data. (Bias is the difference, averaged over all possible samples, between the estimate and the desired value.)

Nonsampling Variability

Nonsampling errors can be attributed to many sources; e.g., inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, inability or unwillingness on the part of the

respondents to provide correct information, inability 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).

Undercoverage in the CPS results from missed housing units and missed persons within sample households. Overall undercoverage as compared to the level of the 1980 Decennial Census is about 7 percent. It is known that 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.

Ratio estimation to independent age-sex-race population controls, as described previously, partially corrects for the bias due to survey 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-race group. Further, the independent population controls used 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

The standard errors given in Tables C-1 through C-8 are primarily measures of sampling variability, that is, of the variation that occurred by chance because a sample rather than the entire population was surveyed. The sample estimate and its standard error enable one to construct confidence intervals, ranges that would include the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these being 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:

1. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

2. 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.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

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 percent of adults who were high school graduates in 1983 to those in 1973. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the parameters are different when, in fact, they are identical.

Table C-1. Standard Errors of Estimated Numbers for all Characteristics Other Than School Enrollment: Total or White

(Numbers in thousands)

Size of estimate[1]	Standard error	Size of estimate[1]	Standard error
25	10	5,000	136
50	14	10,000	189
100	19	25,000	288
250	31	50,000	379
500	43	100,000	444
1,000	61	150,000	401
2,500	97	200,000	184

[1]These values must be multiplied by the appropriate "f" factor to obtain the standard error for a specific characteristic other than school enrollment.

Note: (i) To estimate standard errors for years prior to 1956, multiply the above standard errors by 1.4; for 1956 to 1966, multiply by 1.14; and for 1967 to 1981, multiply 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{ax^2 = bx}$, where a = -0.000018 and b = 3,770 from table C-9.

Table C-2. Standard Errors of Estimated Numbers for all Characteristics Other Than School Enrollment: Black and Other Races

(Numbers in thousands)

Size of estimate	Standard error[1]	Size of estimate	Standard error[1]
25	10	2,500	99
50	15	5,000	133
100	21	10,000	166
250	33	15,000	173
500	46	25,000	107
1,000	64	30,000	(X)

X Not applicable.
[1]These values must be multiplied by the appropriate "f" factor to obtain the standard error for a specific characteristic other than school enrollment.

Note: (i) To estimate standard errors for years prior to 1956, multiply the above standard errors by 1.4; for 1956 to 1966, multiply by 1.14; and for 1967 to 1981, multiply by 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{ax^2 = bx}$, where a= -0.000154 and b = 4,310 from table C-9.

To perform the most common test, let x and y be sample estimates for two characteristics of interest. Let the standard error on the difference x-y be σ_{DIFF}. If the ratio $R = (x-y)/\sigma_{DIFF}$ is between -2 and +2, no conclusion about the parameters is justified at the 0.05 level of significance. If, on the other hand, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. In this event, it is commonly accepted practice to say that the parameters are different. Of course, sometimes this conclusion will be wrong. When the parameters are, in fact, the same, there is a 5 percent chance of concluding that they are different. All statements of comparison in the text have passed a hypothesis test at the 0.10 level of significance or better, and most have passed a hypothesis test at the 0.05 level of significance or better. This means that, for most differences cited in the text, the estimated difference between characteristics is greater than twice the standard error of the difference. For the other differences mentioned, the estimated difference between characteristics is between 1.6 and 2.0 times the standard error of the difference. When this is the case, the statement of comparison is qualified, e.g., by the use of the phrase "some evidence."

Comparability Of Data

In using metropolitan and nonmetropolitan data, caution should be used in comparing estimates for 1977 and 1978 to each other or to any other years. Methodological and sample design changes occurred in these years resulting in relatively large differences in the metropolitan and nonmetropolitan area estimates. However, estimates for 1979 and later are comparable as are estimates for 1976 and earlier.

Caution should also be used when comparing estimates for 1981 and later, which reflect 1980 census-based population controls, with estimates from earlier years. This change in population controls had relatively little impact on summary measures such as means, medians, and percent 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 1981 and later will differ from those for earlier years by more than what could be attributed to actual changes in the population and these differences could be disproportionately greater for certain subpopulation groups than for the total population.

Table C-3. Generalized Standard Errors for Estimated Numbers of Persons for School Enrollment Characteristics: Total or White

(Numbers in thousands)

Estimated number of persons	Total persons in age group									
	100	250	500	1,000	2,500	5,000	10,000	25,000	50,000	100,000
10	4.6	4.7	4.8	4.8	4.8	4.8	4.8	4.8	4.8	4.8
20	6.1	6.5	6.7	6.7	6.8	6.8	6.8	6.8	6.8	6.8
30	7.0	7.8	8.1	8.2	8.3	8.3	8.3	8.3	8.3	8.3
40	7.4	8.8	9.2	9.4	9.5	9.6	9.6	9.6	9.6	9.6
50	7.6	9.6	10.2	10.5	10.6	10.7	10.7	10.7	10.7	10.7
75	6.6	11.0	12.1	12.7	13.0	13.1	13.1	13.1	13.2	13.2
100	-	11.8	13.6	14.4	14.9	15.1	15.1	15.2	15.2	15.2
200	-	9.6	16.7	19.2	20.6	21.1	21.3	21.4	21.5	21.5
300	-	-	16.7	22.0	24.7	25.5	25.9	26.2	26.3	26.3
400	-	-	13.6	23.6	27.9	29.2	29.8	30.2	30.3	30.3
500	-	-	-	24.0	30.4	32.3	33.1	33.7	33.8	33.9
750	-	-	-	20.8	34.8	38.4	40.0	41.0	41.3	41.5
1,000	-	-	-	-	37.2	43.0	45.6	47.1	47.6	47.8
2,000	-	-	-	-	30.4	52.7	60.8	65.2	66.6	67.3
3,000	-	-	-	-	-	52.7	69.7	78.1	80.7	82.0
4,000	-	-	-	-	-	43.0	74.5	88.1	92.2	94.2
5,000	-	-	-	-	-	-	76.0	96.2	102.0	104.8
7,500	-	-	-	-	-	-	65.8	110.2	121.4	126.6
10,000	-	-	-	-	-	-	-	117.8	136.0	144.2
20,000	-	-	-	-	-	-	-	96.2	166.6	192.3
30,000	-	-	-	-	-	-	-	-	166.6	220.3
40,000	-	-	-	-	-	-	-	-	136.0	235.6
50,000	-	-	-	-	-	-	-	-	-	240.4
75,000	-	-	-	-	-	-	-	-	-	208.2
100,000	-	-	-	-	-	-	-	-	-	-

- Not applicable.

Note: (i) To estimate standard errors for years prior to 1956, multiply the above standard errors by 1.4; for 1956 to 1966, multiply by 1.14; and for 1967 to 1981, multiply by 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{-(b/T)\,x^2 + b\,x}$, where b = 2,312 (from table C-9) and T is the total number of persons in an age group.

Table C-4. Generalized Standard Errors for Estimated Numbers of Persons for School Enrollment Characteristics: Black and Other Races

(Numbers in thousands)

Estimated number of persons	Total persons in age group						
	100	250	500	1,000	2,500	5,000	10,000
10	4.8	5.0	5.0	5.1	5.1	5.1	5.1
20	6.4	6.9	7.1	7.1	7.2	7.2	7.2
30	7.4	8.3	8.6	8.7	8.8	8.8	8.8
40	7.9	9.3	9.8	10.0	10.1	10.2	10.2
50	8.1	10.2	10.8	11.1	11.3	11.3	11.4
75	7.0	11.7	12.9	13.4	13.8	13.9	13.9
100	-	12.5	14.4	15.3	15.8	16.0	16.0
200	-	10.2	17.7	20.4	21.9	22.3	22.6
300	-	-	17.7	23.4	26.2	27.1	27.5
400	-	-	14.4	25.0	29.6	30.9	31.6
500	-	-	-	25.5	32.2	34.2	35.1
750	-	-	-	22.1	36.9	40.7	42.5
1,000	-	-	-	-	39.5	45.6	48.4
2,000	-	-	-	-	32.2	55.9	64.5
3,000	-	-	-	-	-	55.9	73.9
4,000	-	-	-	-	-	45.6	79.0
5,000	-	-	-	-	-	-	80.6
7,500	-	-	-	-	-	-	69.8
10,000	-	-	-	-	-	-	-

- Not applicable.

Note: (i) To estimate standard errors for years prior to 1956, multiply the above standard errors by 1.4; for 1956 to 1966, multiply by 1.14; and for 1967 to 1981, multiply by 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{-(b/T)\ x^2 + b\ x}$, where b = 2,600 (from table C-9) and T is the total number of persons in an age group.

Table C-5. Standard Errors of Estimated Percentages for all Characteristics Other Than School Enrollment: Total or White

Base of percentage (thousands)	Estimated percentage[1]				
	2 or 98	5 or 95	10 or 90	25 or 75	50
100	2.7	4.2	5.8	8.4	9.7
250	1.7	2.7	3.7	5.3	6.1
500	1.2	1.9	2.6	3.8	4.3
1,000	0.9	1.3	1.8	2.7	3.1
2,500	0.5	0.8	1.2	1.7	1.9
5,000	0.4	0.6	0.8	1.2	1.4
10,000	0.3	0.4	0.6	0.8	1.0
25,000	0.2	0.3	0.4	0.5	0.6
50,000	0.1	0.2	0.3	0.4	0.4
100,000	0.09	0.1	0.2	0.3	0.3
150,000	0.07	0.1	0.2	0.2	0.3

[1]These factors must be multiplied by the appropriate "f" factors in table C-9 to obtain the standard error for a specific characteristic other than school enrollment.

Note: (i) To estimate standard errors for years prior to 1956, multiply the above standard errors by 1.4; for 1956 to 1966, multiply by 1.14; and for 1967 to 1981, multiply by 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{(b/x) - p(100 - p)}$, where b = 3,770 from table C-9.

Note When Using Small Estimates

Summary measures (such as medians, and percent distributions) are shown in the report only when the base is 75,000 or greater. Because of the large standard errors involved, there is little chance that summary measures would reveal useful information when computed on a smaller base. Estimated numbers are shown, however, even though the relative standard errors of these numbers are larger than those for corresponding percentages.

These smaller estimates are provided primarily to permit such combinations of the categories as serve each data user's needs. Also, care must be taken 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 Error Tables and Their Use

In order to derive standard errors that would be applicable to a large number of estimates and could be prepared at a moderate cost, a number of approximations were required. Therefore, instead of providing an individual standard error for each estimate, generalized sets of standard errors are provided for various types of characteristics. As a result, the sets of standard errors provided give an indication of the order of magnitude of the standard error of an estimate rather than the precise standard error.

The figures presented in tables C-1 through C-8 are approximations to the standard errors of various estimates for persons in the United States. To obtain the approximate standard error for a specific characteristic, the appropriate standard error in table C-1 through C-8 must be multiplied by the factor for that characteristic given in table C-9. These factors must be applied to the generalized standard errors in order to adjust for the combined effect of the sample design and the estimating procedure on the value of the characteristic. Standard errors for intermediate values not shown in the generalized tables of standard errors may be approximated by linear interpolation.

Two parameters (denoted "a" and "b") are used to calculate standard errors for each type of characteristic; they are presented in table C-9. The two parameters may be used directly to calculate the standard errors for estimated numbers and percentages. Methods for computation are given in the following sections.

Standard Errors of Estimated Numbers

The approximate standard error, σ_x, of an estimated number shown in this report can be obtained in two ways. It may be obtained by use of the formula

$$\sigma_x = f\sigma \qquad (1)$$

where f is the appropriate factor from table C-9 and σ is the standard error on the estimate obtained by interpolation from tables C-1, C-2, C-3, or C-4.

Alternatively, the standard error for estimates may be calculated directly using the parameters in table C-9. For all demographic characteristics other than school enrollment and education, use formula (2),

$$\sigma_x = \sqrt{ax^2 + bx} \qquad (2)$$

where x is the size of the estimate and a and b are the parameters in table C-9 associated with the particular characteristic. Standard errors for estimates of characteristics pertaining to school enrollment may be approximated by formula (3),

$$\sigma_x = \sqrt{-\frac{b}{T}x^2 + bx} \qquad (3)$$

where x is the size of the estimate, T is the total number of persons in the specific age group, and b is the parameter in table C-9 associated with the particular characteristic.

Illustrations of the Computation of the Standard Error of an Estimated Number

Table 17 shows that in 1983 there were 70,534,000 persons 35 years of age and older who were married. Using formula (1) with f=1.1 from table C-9 and σ=406,000 from table C-1, the standard error of 70,534,000 is (1.1)(406,000) = 446,600. Alternatively, using formula (2) with the appropriate a parameter of -0.000025 and b parameter of 4480 from table C-9, the approximate standard error is

$$438{,}000 \doteq \sqrt{(-0.000025)(70{,}534{,}000)^2 + (4480)(70{,}534{,}000)}.$$

Table C-6. Standard Errors of Estimated Percentages for all Characteristics Other Than School Enrollment: Black and Other Races

Base of percentage (thousands)	Estimated percentage[1]				
	2 or 98	5 or 95	10 or 90	25 or 75	50
100	2.9	4.5	6.2	9.0	10.4
250	1.8	2.9	3.9	5.7	6.6
500	1.3	2.0	2.8	4.0	4.6
1,000	0.9	1.4	2.0	2.8	3.3
2,500	0.6	0.9	1.2	1.8	2.1
5,000	0.4	0.6	0.9	1.3	1.5
10,000	0.3	0.5	0.6	0.9	1.0
15,000	0.2	0.4	0.5	0.7	0.8
20,000	0.2	0.3	0.4	0.6	0.7

[1]These factors must be multiplied by the appropriate "f" factors in table C-9 to obtain the standard error for a specific characteristic other than school enrollment.

Note: (i) To estimate standard errors for years prior to 1956, multiply the above standard errors by 1.4; for 1956 to 1966, multiply by 1.14; and for 1967 to 1981, multiply by 0.93.

(ii) The standard errors were calculated using formula, $\sqrt{(b/x) - p(100 - p)}$, where b = 4,310 from table C-9.

The 90-percent confidence interval for this estimate is from 69,833,000 to 71,235,000 (using 1.6 times the standard error). Similarly, the 95-percent confidence interval is from 69,658,000 to 71,410,000 (using twice the standard error). 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 95 percent of all possible samples.

Table C-7. Generalized Standard Errors of Estimated Percentages for School Enrollment Characteristics: Total or White

Base of percentage (thousands)	Estimated percentage[1]				
	2 or 98	5 or 95	10 or 90	25 or 75	50
100	2.1	3.3	4.6	6.6	7.6
250	1.3	2.1	2.9	4.2	4.8
500	1.0	1.5	2.0	2.9	3.4
1,000	0.7	1.0	1.4	2.1	2.4
2,500	0.4	0.7	0.9	1.3	1.5
5,000	0.3	0.5	0.6	0.9	1.1
10,000	0.2	0.3	0.5	0.7	0.8
25,000	0.1	0.2	0.3	0.4	0.5
50,000	0.1	0.1	0.2	0.3	0.3
100,000	0.07	0.1	0.1	0.2	0.2
150,000	0.05	0.09	0.1	0.2	0.2

[1]These values must be multiplied by the appropriate "f" factor in table C-9 to obtain the standard error for a specific school enrollment characteristic.

Note: (i) To estimate standard errors for years prior to 1956, multiply the above standard errors by 1.4; for 1956 to 1966, multiply by 1.14; and for 1967 to 1981, multiply by 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{(b/x) - p(100 - p)}$, where b = 2,312 from table C-9.

Table C-8. Generalized Standard Errors of Estimated Percentages for School Enrollment Characteristics: Black and Other Races

Base of percentage (thousands)	Estimated percentage[1]				
	2 or 98	5 or 95	10 or 90	25 or 75	50
75	2.6	4.1	5.6	8.1	9.3
100	2.3	3.5	4.8	7.0	8.1
250	1.4	2.2	3.1	4.4	5.1
500	1.0	1.6	2.2	3.1	3.6
1,000	0.7	1.1	1.5	2.2	2.5
2,000	0.5	0.7	1.0	1.4	1.6
5,000	0.3	0.5	0.7	1.0	1.1
10,000	0.2	0.4	0.5	0.7	0.8
15,000	0.2	0.3	0.4	0.6	0.7
20,000	0.2	0.2	0.3	0.5	0.6

[1]These factors must be multiplied by the appropriate "f" factors in table C-9 to obtain the standard error for a specific school enrollment characteristic.

Note: (i) To estimate standard errors for years prior to 1956, multiply the above standard errors by 1.4; for 1956 to 1966, multiply by 1.14; and for 1967 to 1981, multiply by 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{(b/x) - p(100 - p)}$, where b = 2,600 from table C-9.

Table 4 shows that in 1983 there were 2,111,000 persons 22 to 24 years old enrolled in college and 12,722,000 total persons in that age group. Using formula (1) with f=1.0 from table C-9 and σ=62,700 from table C-3, the standard error of 2,111,000 is (1.0)(62,700)= 62,700. The value of σ(=62,700) was obtained by linear interpolation in two directions in table C-3. The first interpolation was between 10,000,000 persons and 25,000,000 total persons for both 2,000,000 and 3,000,000 persons. The second interpolation was between these two values to get the value corresponding to 2,111,000 persons. Alternatively, table C-9 indicates that the appropriate "b" parameter to use in calculating a standard error for this estimate is b=2312. Using formula (3) since school enrollment is the characteristic of interest, the approximate standard error is

$$64{,}000 \doteq \sqrt{-\frac{2312}{12{,}722{,}000}\ (2{,}111{,}000)^2 + (2312)(2{,}111{,}000)}$$

The 90-percent confidence interval for this estimate is from 2,009,000 to 2,213,000, and the 95 percent confidence interval is from 1,983,000 to 2,239,000.

Standard Errors of Estimated Percentages

The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which this percentage is based. 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-9 indicated by the numerator. The approximate standard error, $\sigma_{(x,p)}$, of an estimated percentage can be obtained by use of the formula

$$\sigma_{(x,p)} = f\sigma \qquad (4)$$

In this formula, f is the appropriate factor from table C-9 and σ is the standard error on the estimate from table C-5, C-6, C-7, or C-8. For estimates pertaining to school enrollment characteristics of persons use tables C-7 or C-8; for any other characteristic use tables C-5 or C-6. Alternatively, the standard error may be approximated by the following formula from which the standard errors in Table C-5 through C-8 were calculated. Use of this formula will give more accurate results than use of formula (4) above.

$$\sigma_{(x,p)} = \sqrt{\frac{b}{x} \cdot p\ (100-p)} \qquad (5)$$

Here x is the size of the subclass of persons or households which is the base of the percentage, p is the percentage ($0 \geq p \geq 100$), and b is the parameter in table C-9 associated with the particular characteristic in the numerator of the percentage.

Illustration of the Computation of the Standard Error of an Estimated Percentage

Table 8 shows that an estimated 38.5 percent of the 397,000 Black persons who were graduated from high school in 1983 were enrolled in college. Using formula (4) with f=1.0 from table C-9 and σ=3.9 from table C-8, the standard error of 38.5 percent is (1.0)(3.9)=3.9. Alternatively, using formula (5) with the appropriate b parameter of 2,600 from Table C-9, the standard error of 38.5 percent is given by

$$3.9 \doteq \sqrt{\frac{2{,}600}{397{,}000}\ (38.5)(61.5)}$$

Thus, a 90-percent confidence interval for this estimate, using the standard error found by formula (4), is from 32.3 to 44.7, and the 95 percent confidence interval is from 30.7 to 46.3.

Standard Error of a Difference

For a difference between two sample estimates, the standard error is approximately equal to

$$\sigma_{(x,y)} = \sqrt{\sigma_x^2 + \sigma_y^2} \qquad (6)$$

where σ_x and σ_y are the standard errors of the estimates x and y, respectively. The estimates can be numbers, percents, etc. This will represent the actual standard error quite accurately for the difference between two estimates of the same characteristics in two different areas or for the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration of the Calculation of the Standard Error of a Difference

Table 1 of this report shows that in October 1983 an estimated 19.4 percent of 6,207,000 men 22 to 24 years old were enrolled in college as compared to 13.9 percent of 6,515,000 women 22 to 24 years old. Using formula (5) with the appropriate b parameter of 2312 from table C-9, the approximate standard error of 19.4 percent is 0.8, and the approximate standard error of 13.9 percent is 0.7. The apparent difference between these two estimates is 5.5 percent, and the standard error associated with the difference is:

$$1.1 \doteq \sqrt{(0.8)^2 + (0.7)^2}$$

The 90-percent confidence interval on the difference of 5.5 percent is from 3.7 to 7.3 percent. Similarly, the 95-percent confidence interval is from 3.3 to 7.7 percent. Therefore, a conclusion that the average estimate of the difference derived from all possible samples lies within a range computed in this way would be correct roughly 95 percent of the time. Since this interval does not contain zero, we can conclude with 95-percent confidence that college enrollment of males ages 22 to 24 was greater than that of females ages 22 to 24.

Table C-9. "a" and "b" Parameters and "f" Factors for Calculating Approximate Standard Errors of Estimated Numbers and Percentages

Characteristic	Persons			Families, family heads		
	a	b	f factor[1]	a	b	f factor[1]
SCHOOL ENROLLMENT CHARACTERISTICS						
Persons Enrolled in School						
14 to 34 years old:						
Total or White	(X)	2312	1.0	-0.000010	1778	0.7
Black	(X)	2600	1.0	-0.000066	1606	0.6
Spanish origin (levels)[3]	+0.001744	2131	(X)	-0.000042	3068	0.9
Spanish origin (percentages)[3]	(x)	3873	1.3	(X)	(X)	(X)
3 to 34 years old:						
Total or White	(X)	2312	1.0	-0.000010	1778	0.7
Black	(X)	2600	1.0	-0.000066	1606	0.6
Spanish origin (levels)[3]	+0.001744	2131	(x)	-0.000042	3068	0.9
Spanish origin (percentage)[3]	(X)	3873	1.3	(X)	(X)	(X)
35+ years old:						
Total or White	(X)	2312	1.0	(X)	(X)	(X)
Black	(X)	2600	1.0	(X)	(X)	(X)
Children Enrolled in School						
3 to 6 years old:						
Total or White	(X)	2698	1.1	(X)	(X)	(X)
Black	(X)	2698	1.0	(X)	(X)	(X)
Spanish origin[3]	-0.000194	2698	1.1	(X)	(X)	(X)
3 to 13 years old:						
Total or White	(X)	2698	1.1	(X)	(X)	(X)
Black	(X)	2698	1.0	(X)	(X)	(X)
Spanish origin[3]	-0.000073	2698	1.1	(X)	(X)	(X)
CHARACTERISTICS OTHER THAN SCHOOL ENROLLMENT						
Marital Status and Household and Family:						
Total or White	-0.000025	4480	1.1	(X)	(X)	(X)
Black	-0.000265	6426	1.2	(X)	(X)	(X)
Spanish origin[3]	-0.000055	9560	1.6	(X)	(X)	(X)
Income:						
Persons tabulated by family						
Total or White	-0.000018	3770	1.0	(X)	(X)	(X)
Black	-0.000154	4310	1.0	(X)	(X)	(X)
Spanish origin[3]	-0.000067	10112	1.6	(X)	(X)	(X)
Residence:						
Total or White	-0.000031	5444	1.2	-0.000020	2778	0.9
Black	-0.000391	9475	1.5	-0.000228	3278	0.9
Spanish origin[3]	-0.000095	19235	2.3	-0.000018	3640	1.0

(X) Not applicable.

[1]Apply these factors to tables C-3, C-4, C-7, or C-8.
[2]Apply these factors to tables C-1, C-2, C-5, or C-6.
[3]Spanish origin: To compute standard errors for school enrollment levels ("persons"), use only formula (2). The formula, $\sqrt{-(b/T)x^2 + bx}$, cannot be used because Spanish was not ratio estimated to Hispanic population controls. To calculate any other Spanish standard errors use formulas (2) or (5), or use the generalized standard errors from table C-1, C-3, C-5, or C-7 in conjunction with formulas (1) or (4).

Note: (i) For nonmetropolitan data cross-tabulated with other data, also apply the factor 1.2 to the "f" factor and the factor 1.5 to the "a" and "b" parameters.
(ii) Multiply the "a" and "b" parameters by 0.87 for CPS data collected from 1967 to 1981, by 1.5 for CPS data collected from 1956 to 1966, and by 2.25 for CPS data collected before 1956.
(iii) For regional data, multiply the "a" and "b" parameters by 0.99 for the Northeast, 1.02 for the Midwest, 0.98 for the South, and 0.84 for the West.