Appendix B. Source and Reliability of the Estimates

SOURCE OF DATA

Most of the estimates in this report are based on data collected in March 1986 and 1987 from the Current Population Survey (CPS) of the Bureau of the Census. Some estimates are based on data obtained from the CPS in earlier years and from earlier decennial censuses. 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, questions are asked each March about educational attainment. In order to obtain more reliable data for the Hispanic population, the March CPS sample was enlarged to include all households from the previous November sample which contained at least one person of Hispanic origin. For this report, persons in the Armed Forces living off post or with their families on post are included.

Current Population Survey (CPS). The present CPS sample was selected from the 1980 census files with coverage in all 50 States and the District of Columbia. The sample is continually updated to reflect new construction. The current CPS sample is located in 729 areas comprising 1,973 counties, independent cities, and minor civil divisions in the Nation. In this sample, approximately 61,500 occupied households were eligible for interview. Of this number, about 3,500 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 table below provides a description of some aspects of the CPS sample designs in use during the referenced data collection periods.

CPS Estimation Procedure. The estimation procedure used in this survey involved the inflation of the weighted sample results to independent estimates of the total civilian noninstitutional population of the United States by age, race, sex, and Hispanic/non-Hispanic categories. These independent estimates are based on statistics from the decennial censuses of population; statistics in births, deaths, immigration and emigration; and statistics on the strength of the Armed Forces. The independent population estimates used to obtain data for 1980 and later are based on the 1980 decennial census. In earlier reports in this series, data for 1972 through 1979 were obtained using independent population estimates based on the 1970 decennial census. The estimation procedure for the data from the March supplement involved a further adjustment so that husband and wife of a household received the same weight.

Description of the March Current Population Survey

Time period	Number of sample areas	Housing units eligible[1]	
		Interviewed	Not interviewed
1986 to 1987	729	59,000	3,000
1985	[2]629/729	57,000	2,500
1982 to 1984	629	58,000	2,500
1980 to 1981	629	65,500	3,000
1977 to 1979	614	55,000	3,000
1973 to 1976	461	46,500	2,500
1972	449	45,000	2,000
1967 to 1971	449	48,000	2,000
1963 to 1966	357	33,500	1,500
1960 to 1962	[3]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]Does not include supplemental Hispanic households.
[2]Three rotation groups were located in 629 areas and five rotation groups in 729 areas.
[3]Three sample areas were added in 1960 to represent Alaska and Hawaii after statehood.

The estimates in this report for the survey years 1985 to 1987 are also based on revised survey weighting procedures for persons of Hispanic origin. In previous years, the estimation procedures used in this survey involved the inflation of weighted sample results to independent estimates of the noninstitutional population by age, sex, and race. There was, therefore, no specific control of the survey estimates for the Hispanic origin population. During the last several years, the Bureau of the Census has developed independent population controls for the Hispanic population by sex and detailed age groups and has adopted revised weighting procedures to incorporate these new controls. It should be noted that the independent population estimates include some, but not all, illegal immigrants.

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 sample estimates 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 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-Hispanic 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 per sons in the same age-sex-race-Hispanic 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 the following tables are primarily measures of sampling variability, that is, of the variations 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 a confidence interval, a range that would include the average results 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 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 proportion of young male college graduates to young female college graduates. 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.

All statements of comparison in the text have passed a hypothesis test at the 0.10 level of significance or better. This means that, for all differences cited in the text, the estimated difference between characteristics is greater than 1.6 times the standard error of the difference.

Comparability of Data. Caution should be used when comparing estimates for 1980 and later, which reflect 1980 census-based population controls, to those for 1972 through 1979, which reflect 1970 census-based population controls. 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 1980 and later will differ from those for earlier years 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.

Care must also be taken when comparing Hispanic estimates over time due to the recent change in weighting of the Hispanic population beginning in 1985. Before 1985, there were no independent population control totals for persons of Hispanic origin. See the section entitled "CPS Estimation Procedure."

Also, 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. Data on metropolitan and nonmetropolitan residence are not available for 1985.

Decennial Census of Population. The decennial censuses data shown in this report are not strictly comparable to the CPS data. This is due in large part to differences in inter viewer training and experience and in different survey processes. This is an additional component of error not reflected in the standard error tables. Therefore, caution should be used in comparing results between these different sources.

Note When Using Small Estimates. Summary measures (such as medians and percent distributions) are shown 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 which 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 B-1 through B-4 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 tables B-1 through B-4 must be multiplied by the factor for that characteristic given in table B-5. 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.

The standard errors in tables B-1 through B-4 and the factors in table B-5 were calculated using the b parameters in table B-5. The 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, S_x, of an estimated number shown in this report can be obtained in two ways. It may be obtained by use of the formula

$$S_x = f_1 f_2 s \quad (1)$$

where f_1 is the appropriate factor from table B-5, f_2 is the appropriate factor from table B-6, and s is the standard error on the estimate obtained by interpolation from table B-1 or B-2.

Alternatively, standard errors may be approximated by formula (2) from which standard errors in tables B-1 and B-2 were calculated. Use of this formula will provide more accurate results than use of formula (1) above.

$$S_x = f_2 \sqrt{-\frac{b}{T} x^2 + bx} \quad (2)$$

Table B-1. Generalized Standard Errors for Estimated Numbers of Persons: Total or White

(Numbers in thousands)

Estimated number of persons	Total persons in age group[1]									
	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	-	-	-	-	-	-	-	-	-	-

[1]These values must be multiplied by the appropriate factor in tables B-5 and/or B-6 to obtain the standard error for a specific characteristic.

NOTE: (i) To estimate the standard errors for years prior to 1956, multiply the above standard errors by 1.4; for years 1956-66, multiply by 1.14; and for years 1967-79, multiply by 0.93.

(ii) The standard errors were calculated using the formula $\sqrt{-(b/T)\,x^2 + bx}$, where b = 2312 (from table B-5) and T is the total number of persons in an age group.

Here x is the size of the estimate, T is the total number of persons in a specific age group, b is the parameter in table B-5 associated with the particular characteristic, and f_2 is the appropriate factor from table B-6. If T is not known, for Total or white use 100,000,000; for Blacks and Hispanic use 10,000,000.

Illustration of the Computation of the Standard Error of an Estimated Number. Table 1 of this report shows that in 1987 there were 4,768,000 young adults (ages 25 to 29 years) who were college graduates and 21,636,000 total persons in that age group. Using formula (1) with f_1 = 1.0 from table B-5, f_2 = 1.0 from table B-6, and s = 90,000 from table B-1, the standard error of 4,768,000 is (1.0)(1.0)(90,000) = 90,000. The value of s (= 90,000) was obtained by linear interpolation in two directions in table B-1. The first interpolation was between 10,000,000 and 25,000,000 total persons for both 4,000,000 and 5,000,000 estimated number of persons. The value for 4,000,000 estimated persons was 85.0 and for 5,000,000 estimated persons was 91.7. The second interpolation was between these two values to get the value corresponding to 4,768,000 persons. Alternatively, using formula (2), with the appropriate b parameter of 2312 from table B-5 and f_2 factor of 1.0 from table B-6, the approximate standard error is.

$$93{,}000 = (1.0)\sqrt{-\frac{2312}{21{,}636{,}000}(4{,}768{,}000)^2 + (2312)(4{,}768{,}000)}$$

The 90-percent confidence interval for this estimate is from 4,619,000 to 4,917,000 (using 1.6 times 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 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 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 per centage, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories,

Table B-2. Generalized Standard Errors for Estimated Numbers of Persons: Black and Other Races

(Numbers in thousands)

Estimated number of persons	Total persons in age group[1]						
	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	-	-	-	-	-	-	-

[1]These values must be multiplied by the appropriate factor in tables B-5 and/or B-6 to obtain the standard error for a specific characteristic.

NOTE: (i) To estimate the standard errors for years prior to 1956, multiply the above standard errors by 1.4; for years 1956-66,multiply by 1.14; and for years 1967-79, multiply by 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{-(b/T)\, x2 + bx}$, where b = 2600 (from table B-5) and T is the total number of persons in an age group.

Table B-3. Generalized Standard Errors of Estimated Percentages: Total or White

Base of percentage (thousands)	Estimated percentage[1]					
	1 or 99	2 or 98	5 or 95	10 or 90	25 or 75	50
75	1.7	2.5	3.8	5.3	7.6	8.8
100	1.5	2.1	3.3	4.6	6.6	7.6
250	1.0	1.3	2.1	2.9	4.2	4.8
500	0.7	1.0	1.5	2.0	2.9	3.4
750	0.6	0.8	1.2	1.7	2.4	2.8
1,000	0.5	0.7	1.0	1.4	2.1	2.4
2,500	0.3	0.4	0.7	0.9	1.3	1.5
5,000	0.2	0.3	0.5	0.6	0.9	1.1
7,500	0.2	0.2	0.4	0.5	0.8	0.9
10,000	0.2	0.2	0.3	0.5	0.7	0.8
15,000	0.12	0.2	0.3	0.4	0.5	0.6
25,000	0.10	0.13	0.2	0.3	0.4	0.5
50,000	0.07	0.10	0.15	0.2	0.3	0.3
100,000	0.05	0.07	0.10	0.14	0.2	0.2

[1]These values must be multiplied by the appropriate factor in tables B-5 and/or B-6 to obtain the standard error for a specific characteristic.

NOTE: (i) To estimate the standard errors for years prior to 1956, multiply theabove standard errors by 1.4; for years 1956-66, multiply by 1.14; and for years 1967-79, multiply by 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{(b/x)\, p(100 - p)}$, where b = 2312 from table B-5.

Table B-4. Generalized Standard Errors of Estimated Percentages: Black and Other Races

Base of percentage (thousands)	Estimated percentage[1]					
	1 or 99	2 or 98	5 or 95	10 or 90	25 or 75	50
25	3.2	4.5	7.0	9.7	14.0	16.1
50	2.3	3.2	5.0	6.8	9.9	11.4
75	1.9	2.6	4.1	5.6	8.1	9.3
100	1.6	2.3	3.5	4.8	7.0	8.1
250	1.0	1.4	2.2	3.1	4.4	5.1
500	0.7	1.0	1.6	2.2	3.1	3.6
750	0.6	0.8	1.3	1.8	2.5	2.9
1,000	0.5	0.7	1.1	1.5	2.2	2.5
2,500	0.3	0.5	0.7	1.0	1.4	1.6
5,000	0.2	0.3	0.5	0.7	1.0	1.1
7,500	0.2	0.3	0.4	0.6	0.8	0.9
10,000	0.2	0.2	0.4	0.5	0.7	0.8
15,000	0.13	0.2	0.3	0.4	0.6	0.7
20,000	0.11	0.2	0.2	0.3	0.5	0.6

[1]These values must be multiplied by the appropriate factor in tables B-5 and/or B-6 to obtain the standard error for a specific characteristic.

NOTE: (i) To estimate the standard errors for years prior to 1956, multiply the above standard errors by 1.4; for years 1956-66, multiply by 1.14; and for years 1967-79, multiply by 0.93.

(ii) The standard errors were calculated using the formula, $\sqrt{(b/x)\,p\,(100-p)}$, where b = 2600 from table B-5.

use the factor or parameter from table B-5 indicated by the numerator. The approximate standard error, $S_{(x,p)}$, of an estimated percentage can be obtained by use of the formula

$$S_{(x,p)} = f_1\, f_2\, s. \qquad (3)$$

In this formula, f_1 is the appropriate factor from table B-5, f_2 is the appropriate factor from table B-6, and s is the standard error on the estimate from table B-3 or B-4. Alternatively, the standard error may be approximated by the following formula from which the standard errors in tables B-3 and B-4 were calculated. Use of this formula will give more accurate results than use of formula (3) above.

$$S_{(x,p)} = f_2 \sqrt{\frac{b}{x} \cdot p\,(100\text{-}p)} \qquad (4)$$

Here x is the size of the subclass of persons or households which is the base of the percentage, p is the percentage ($0 < p < 100$), and b is the parameter in table B-5 associated with the particular characteristic in the numerator of the percentage.

Illustration of the Computation of the Standard Error of an Estimated Percentage. Table 2 shows that an estimated 83.3 percent of the 2,683,000 Black persons aged 25 to 29 years were high school graduates in 1987. Using formula (3) with $f_1 = 1.0$ and $f_2 = 1.0$ from tables B-5 and B-6, respectively, and s = 1.2 from table B-4, the standard error of 83.3 percent is (1.0)(1.0)(1.2) = 1.2. Alternatively, using formula (4) with the appropriate b parameter of 2,600 from table B-5, the standard error of 83.3 percent is given by

$$1.2 = (1.0)\sqrt{\frac{2,600}{2,683,000}(83.3)\,(16.7)}\,.$$

Thus, a 90-percent confidence interval of this estimate, using the standard error found by formula (4), is from 81.4 to 85.2.

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

$$S_{(x,y)} = \sqrt{S_x^2 + S_y^2} \qquad (5)$$

where S_x and S_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 2 of this report shows that in 1987 an esti mated 87.0 percent of 9,075,000 White women 25 to 29 years old were high school graduates as compared to 82.1 percent of 1,454,000 Black women

25 to 29 years old. Using formula (4) with the appropriate b parameter of 2312 from table B-5, the approximate standard error of 87.0 percent is 0.5; with the appropriate b parameter of 2,600 from table B-5 the approximate standard error of 82.1 percent is 1.6. The apparent difference between these two estimates is 4.9 percent, and the standard error associated with the difference is:

$$1.7 = \sqrt{(0.5)^2 + (1.6)^2}$$

The 90-percent confidence interval on the difference of 4.9 percent is from 2.2 to 7.6 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 90 percent of the time. Since this interval does not contain zero, we can conclude with 90-percent confi dence that White females ages 25 to 29 have a greater percentage of high school graduates than Black females of the same age group.

Standard Error of a Median. The sampling variability of an estimated median depends upon the form of the distribution as well as the size of its base. An approximate method for measuring the reliability of an estimated median is to determine a confidence interval about it. (See the section on sampling variability for a general discussion of confidence intervals.) The following procedure may be used to estimate the 68-percent confidence limits and hence the standard error of a median based on sample data.

1. Determine, using the standard error tables and factors or formula (4), 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);
3. Using the distribution of the characteristic, calculate the 68-percent confidence interval by calculating the values from the distribution correspond ing to the two points established in step (2);
3. Once the limits of the 68-percent confidence interval are computed, the standard error of a median can be computed by the formula

$$S_{median} = (U - L)/2$$

where U = upper limit of the 68-percent confidence interval,
L = lower limit of the 68-percent confidence interval.

For calculations of the confidence interval in step (3), use linear interpolation.

The formula used to implement step (3) for linear interpolation is:

$$x_{pN} = \frac{pN - N_1}{N_2 - N_1}(A_2 - A_1) + A_1 \qquad (6)$$

where N = for distribution of numbers: total number of households, families, or persons in the dis tribution.
= for distribution of percents: the value 1.0.

x_{pN} = estimated value (e.g., years of school completed) for which the number pN ($0 < p < 1$) households, families, or persons in the distribution have larger or equal values. For the purposes of calculating the confidence interval, p takes on the values in step (2). Note that xpN estimates the median when p = 0.50 is used in the formula.

A_1 and A_2 = the lower and upper bounds, respectively, on the interval in which x_{pN} falls.

N_1 and N_2 = for distribution of numbers: the estimated number of households, families, or persons with values at least A_1 and A_2, respectively.
= for distribution of percents: the estimated percent of units (persons, households, etc.) having values of the characteristic greater than or equal to A_1 and A_2, respectively.

The procedure can also be used to estimate standard errors for quintiles or other percentiles by distributing the proper per centage value for p and following the steps outlined above. Note that when combining distributions the resulting median or percentile may lie in an open-ended interval. To calculate such standard errors the user must call Population Division of the Census Bureau to obtain the detailed distribution.

Illustration of the Computation of a Confidence Interval and Standard Error for a Median. Table 1 of this report shows that in 1987 the median years of school completed by all persons 25 to 29 years old was 12.8. Table 1 also indicates the base of the distribution from which this median was determined is 21,636,000 persons.

(1) Using formula (4) with the appropriate b parameter from table B-5, the standard error of 50 percent on a base of 21,636,000 is 0.5 percentage points.

(2) To obtain a 68-percent confidence interval on the estimated median, add to and subtract from 50 percent the standard error found in step (1). This yields percent limits of 49.5 and 50.5.

(3) From table 1, in 1987 13.7 percent of all persons aged 25 to 29 years had completed less than 12

years of school and 56.1 percent had completed less than 13 years of school. Using formula (6), the lower limit on the estimate is found to be about

$$\frac{.495 - .137}{.561 - .137} \times (13.0 - 12.0) + 12.0 = 12.84.$$

Table B-5. Parameters and Factors for Calculating Approximate Standard Errors of Estimated Numbers and Percentages

Type of characteristic	b parameter[1]	f_1 factor
EDUCATION CHARACTERISTIC		
Education Attainment of Persons 14+:		
Total or White	2312	1.0
Black	2600	1.0
Hispanic origin	2600	1.0
CHARACTERISTICS OTHER THAN EDUCATION		
Marital Status:		
Total or White	4480	1.0
Black	6426	1.0
Hispanic origin	5673	0.9
Household Relationship:		
Head, Wife, or Primary Individual:		
Total or White	1778	0.6
Black	1606	0.5
Hispanic origin	1606	0.5
Child or Other Relative in Primary Family, Secondary Family Member, Secondary Individual, or Persons Living in Group Quarters:		
Total or White	4480	1.0
Black	6426	1.0
Hispanic origin	5673	0.9
Occupation:		
Both Sexes:		
Total or White	2327	0.7
Black	2327	0.6
Hispanic origin	1247	0.4
Male:		
Total or White	2013	0.7
Black	2013	0.6
Hispanic origin	1241	0.4
Female:		
Total or White	1725	0.6
Black	1725	0.5
Hispanic origin	1241	0.4

[1]Multiply parameters by 1.96 for CPS data collected before 1956, by 1.3 for CPS data between 1956 and 1966, and by 0.87 for CPS data between1967 and 1979.

NOTE: (i) For nonmetropolitan data cross-tabulated with other data, multiply $f_1$by 1.2.

Similarly, the upper limit is found by linear interpolation to be about

$$\frac{.505 - .137}{.561 - 137} x (13.0 - 12.0) + 12.0 = 12.87.$$

Thus, an approximate 68-percent confidence interval for the median school years completed by all persons 25-29 years old is from 12.84 to 12.87.

(4) The standard error of the median is, therefore, (12.87 - 12.84)/2, or 0.015. (NOTE: Published medians and their standard errors are calculated by the same method as above. However, different medians and standard errors may be obtained because of rounding errors.)

Table B-6. Factors to be Applied to Standard Errors

Type of characteristic	f_2 factor[1]
U.S. totals	1.0
States:	
California	1.13
Florida	0.94
Georgia	1.25
Illinois	0.99
Indiana	1.09
Massachusetts	0.71
Michigan	0.88
Missouri	1.09
New Jersey	0.79
New York	0.91
North Carolina	0.73
Ohio	0.94
Pennsylvania	0.97
Texas	1.14
Virginia	1.24
Regions:	
Northeast	0.94
Midwest	0.95
South	0.94
West	0.90
MSA's	1.0

[1]Multiply standard errors obtained using tables B-1 through B-4 by these factors.

*U. S. GOVERNMENT PRINTING OFFICE:1988-201-478:80123