Appendix. Source and Reliability of Estimates

SOURCE OF DATA

The estimates of school enrollment in 1985 are based on data obtained in October 1985 from the Current Population Survey (CPS) conducted by the Bureau of the Census. 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 60,500 occupied households were 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.

Current Population Survey weighting procedure. The estimation procedure used for this survey involved the inflation of the weighted sample results to independent estimates of the civilian noninstitutional population of the United States by age, race and sex, and Hispanic/non-Hispanic categories. These independent estimates are based on statistics from decennial censuses of population; statistics on births, deaths, immigration, and emigration; and statistics on the strength of the Armed Forces.

The estimates in this report for 1985 are also based on a revised Spanish weighting procedure for persons of Spanish 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 Spanish 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, undocumented immigrants.

Data from 1972 to 1980 were obtained using independent population estimates based on the 1970 decennial census. Estimates for earlier years were based on earlier censuses. Data for 1981 and later years utilized independent estimates derived from the 1980 decennial census.

Two sets of estimates for 1981 are shown in some of the tables in this report; one set results from using independent population estimates based on the more up-to-date 1980 decennial census and the other set results from using 1970 Census based population estimates. The 1970 based estimates have been included to provide continuity in the time series with previous years. Comparing the 1980 based estimates with the 1970 based estimates provides a measure of the effect of changing to the 1980 based estimation procedure.

RELIABILITY OF 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.)

Sampling variability. The standard errors presented in tables A-1 through A-4 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 interval estimates that include the average results of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under identical conditions using the same design; and an estimate and its standard error were calculated from each sample, then:

1. Approximately 90 percent of the intervals from 1.6 standard errors below the estimates to 1.6 standard errors above the estimate would include the average results of all possible samples.
2. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average results of all possible samples.

Table A-1. Generalized Standard Errors for Estimated Numbers of Persons—Total or White

(Numbers in thousands)

Estimated number of persons	Total persons in age group									
	100	250	500	1,000	2,500	5,000	1,0000	25,000	50,000	10,0000
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	(X)	11.8	13.6	14.4	14.9	15.1	15.1	15.2	15.2	15.2
200	(X)	9.6	16.7	19.2	20.6	21.1	21.3	21.4	21.5	21.5
300	(X)	(X)	16.7	22.0	24.7	25.5	25.9	26.2	26.3	26.3
400	(X)	(X)	13.6	23.6	27.9	29.2	29.8	30.2	30.3	30.3
500	(X)	(X)	(X)	24.0	30.4	32.3	33.1	33.7	33.8	33.9
750	(X)	(X)	(X)	20.8	34.8	38.4	40.0	41.0	41.3	41.5
1,000	(X)	(X)	(X)	(X)	37.2	43.0	45.6	47.1	47.6	47.8
2,000	(X)	(X)	(X)	(X)	30.4	52.7	60.8	65.2	66.6	67.3
3,000	(X)	(X)	(X)	(X)	(X)	52.7	69.7	78.1	80.7	82.0
4,000	(X)	(X)	(X)	(X)	(X)	43.0	74.5	88.1	92.2	94.2
5,000	(X)	(X)	(X)	(X)	(X)	(X)	76.0	96.2	102.0	104.8
7,500	(X)	(X)	(X)	(X)	(X)	(X)	65.8	110.2	121.4	126.6
10,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	117.8	136.0	144.2
20,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	96.2	166.6	192.3
30,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	166.6	220.3
40,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	136.0	235.6
50,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	240.4
75,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	208.2
100,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)

(X) 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 1980 multiply by 0.93. (ii) The standard errors were calculated using the formula, $\sqrt{-(b/T)\,x^2 + b\,x}$, where b = 2312 (from table A—5) and T is the total number of persons in an age group.

Table A-2. Generalized Standard Errors for Estimated Numbers of Persons: Black and Hispanic

(Numbers in thousands)

Estimated number of persons	Total persons in age group						
	100	250	500	1000	2500	5000	10000
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	(X)	12.5	14.4	15.3	15.8	16.0	16.0
200	(X)	10.2	17.7	20.4	21.9	22.3	22.6
300	(X)	(X)	17.7	23.4	26.2	27.1	27.5
400	(X)	(X)	14.4	25.0	29.6	30.9	31.6
500	(X)	(X)	(X)	25.5	32.2	34.2	35.1
750	(X)	(X)	(X)	22.1	36.9	40.7	42.5
1,000	(X)	(X)	(X)	(X)	39.5	45.6	48.4
2,000	(X)	(X)	(X)	(X)	32.2	55.9	64.5
3,000	(X)	(X)	(X)	(X)	(X)	55.9	73.9
4,000	(X)	(X)	(X)	(X)	(X)	45.6	79.0
5,000	(X)	(X)	(X)	(X)	(X)	(X)	80.6
7,500	(X)	(X)	(X)	(X)	(X)	(X)	69.8
10,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)

(X) 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 1980 multiply by 0.93. (ii) The standard errors were calculated using the formula, $\sqrt{-(b/T)x^2 + bx}$, where b = 2600 (from table A-5) and T is the total number of persons in an age group.

Table A-3. Generalized Standard Errors of Estimated Percentages—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	0.13	0.2	0.3	0.4	0.5
50,000	0.09	0.15	0.2	0.3	0.3
100,000	0.07	0.10	0.05	0.2	0.2
150,000	0.05	0.12	0.12	0.2	0.2

[1]These values must be multiplied by the appropriate "f" factor in table A-5 to obtain the standard error for a specific 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 1980 multiply by 0.93.

(ii) The b parameter from table A-5 used to create this table is b = 2312.

Table A-4. Generalized Standard Errors of Estimated Percentages—Black and Hispanic

Base of percentage (thousands)	Estimated percentage				
	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
2,500	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" factor in table A-5 to obtain the standard error for a specific 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 1980 multiply by 0.93. (ii) The b parameter from table A-5 used to create this table is b = 2600.

The average result of all possible samples is or is not contained in any particular computed interval. However, for a particular sample one can say with a specified confidence that the average results of all possible samples is included within the constructed interval.

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 difference between two sample estimates is greater than twice the standard error of the difference. Statements of comparison qualified in some way (e.g., by use of the phrase, "some evidence") have a level of significance between 1.6 and 2.0 standard errors.

Note when using small estimates. Percent distributions are shown in this report only when the base of the percentage is greater than 75,000. Because of the large standard errors involved, there is little chance that percentages would reveal useful information when computed on a smaller base. Estimated numbers of persons are shown, however, even though the relative standard errors of these numbers are larger than those for the corresponding percentages. These smaller estimates are provided primarily to permit those combinations of the categories which serves each user's needs.

Standard error tables and their use. In order to derive standard errors that would be applicable to a larger 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 A-1 through A-4 are approximations to the standard errors of various school enrollment estimates for persons. Standard errors for intermediate values not shown in the generalized tables of standard errors may be approximated by linear interpolation.

Standard errors of estimated numbers and estimated percentages. The approximate standard errors of estimated numbers and percentages can be computed directly with formulas (1) and (2) below, respectively. The formulas are:

$$\sigma_x = \sqrt{ax^2 + bx} \qquad (1)$$

where "x" is the size of the estimate and "a" and "b" are the parameters associated with the characteristic; and

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

where "x" is the size of the subclass of the population which is the base of the percentage, "p" is the percentage ($0 < p < 100$), and "b" is the parameter associated with the characteristic.

Table A-5 provides the values of the "a" and "b" parameters that are used in formulas (1) and (2) to approximate standard errors of estimated numbers of persons and estimated percentages.

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

$$\sigma_{(x,y)} = \sqrt{\sigma_x^2 - \sigma_y^2} \qquad (3)$$

where σ_x and σ_y are the standard errors of the estimates x and y; the estimates can be numbers, percents, ratios, etc.

Table A-5. Parameters to be Used for School Enrollment Characteristics for Direct Computation of Standard Errors

(Use for 1981 and later years)

Characteristic	Parameters[1]	
	a	b
Persons enrolled in school		
3-34 years old:		
Total or White	-0.000019	2,312
Black	-0.000129	2,600
Spanish origin	-0.000177	2,600
14-34 years old:		
Total or White	-0.000028	2,312
Black	-0.000195	2,600
Spanish origin	-0.000209	2,600
3-13 years old:		
Total, White	-0.000073	2,698
Black	-0.000393	2,698
Spanish origin	-0.000550	2,698

[1]To obtain "a" and "b" parameters for the October 1967-80 supplements, multiply these parameters by 0.871. To obtain "a" and "b" parameters for 1965, multiply these parameters by 1.307.

This will represent the estimated standard error quite accurately for the difference between two estimates of the same characteristic in two different areas, or for the difference between two 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.

☆U.S. GOVERNMENT PRINTING OFFICE: 1986-491-060:40082