Appendix B. Source and Reliability of the Estimates

SOURCE OF DATA

The estimates in this report are based on data obtained from a supplement to the Current Population Surveys (CPS) of the Bureau of the Census in October of 1958, 1966, 1971 and 1976. 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 14 years old and older in each sample household. In the recent October supplements questions concerning educational characteristics, such as school enrollment, have been asked to acquire information about all levels of education.

The present CPS sample was initially selected from the 1970 census file and is updated continuously to reflect new construction where possible (see section, "Nonsampling Variability" below). Previous sample designs used files from the census most recently completed at the time and updated for new construction.

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 Current Population Survey

Time period	Number of sample areas[1]	Households eligible		Housing units visited, not eligible[2]
		Interviewed	Not interviewed	
October 1976	461	45,000	2,000	8,000
October 1971	449	45,000	2,000	8,000
October 1966	357	33,500	1,500	6,000
October 1958	330	33,500	1,500	6,000

[1]These areas were chosen to provide coverage in each State and the District of Columbia.
[2]These are housing units which were visited but were found to be vacant or otherwise not eligible for interview.

The estimation procedure used for both the CPS data and supplemental data involves the inflation of the weighted sample results to independent estimates of the civilian noninstitutional population of the United States by age, race and sex. These independent estimates were based on statistics from the 1970 Census of Population; statistics on births, deaths, immigration and emigration; and statistics on the strength of the Armed Forces.

Reliability of the Estimates

Since the estimates in this report are 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 schedules, instructions and enumerators. There are two types of errors possible in an estimate based on a sample survey: sampling and nonsampling. The standard errors provided for this report primarily indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. 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.

Nonsampling Variability

Nonsampling errors in surveys 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 of respondents to provide correct information, inability to recall information, errors made in 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 sample households and all persons within sample households (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 decennial census, is about 5 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 than for Whites. Ratio estimation to independent age-sex-race population controls, as described

previously, partially corrects for the biases 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 than interviewed persons in the same age-sex-race group. Further, the independent population controls used have not been adjusted for undercoverage in the 1970 census, which was estimated at 2.5 percent of the population, with differentials by age, sex, and race similar to those observed in CPS.

The approximate magnitude to two sources of undercoverage of housing units is known. Of the 83,000 housing units in the U.S. about 600,000 new construction housing units other than mobile homes are not represented in the CPS sample because they were assigned building permits prior to the 1970 census but building was not completed by the time of the census (i.e., April 1970). Most conventional new construction, for which building permits were issued after the census, is represented. About 290,000 occupied mobile homes are not represented in CPS; these units were either missed in the census or have been built or occupied since the census. These estimates of missed units are relevant to the present sample only and not to earlier designs where the extent of undercoverage was generally less. The extent of other sources of undercoverage of housing units is unknown but believed to be small.

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 whole of the population was surveyed. The sample estimate and its estimated standard error enable one to construct interval estimates that include the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under identical conditions and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples;
2. Approximately 90 percent of the interval 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;
3. 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.

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

All the statements of comparison appearing in the text are significant at a 1.6 standard error level or better, and most are significant at a level of more than 2.0 standard errors. This means that for most differences cited in the text, the estimated difference 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 the report only when the base of the percentage is 75,000 or greater. Because of the large standard errors involved, there is little chance that percentages would reveal useful information when computed on a smaller base. Estimated totals are shown, however, even though the relative standard errors of these totals are larger than those for corresponding percentages. These smaller estimates are provided primarily to permit such combinations of the categories as serve each user's needs.

Comparability With Other Data

Data from sources other than Census Bureau may be subject to different amounts of sampling and nonsampling variability. In addition, data obtained from the CPS are not entirely comparable with data obtained from other sources. This is due in a large part to differences in interviewer training and experience and in differing survey processes. This is an additional component of error not reflected in the standard error tables. Therefore, caution should be used in comparing results from these different sources.

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 after applying factors give an indication of the order of magnitude of the standard errors of an estimate rather than of the precise standard error.

The figures presented in tables B-1 through B-4 provide approximations to standard errors of various estimates for total, White, or Black persons in the total United States for education only. To obtain standard errors for other characteristics, factors from table B-5 must be applied to the standard errors given for education in order to adjust for the combined effect of sample design and estimating procedure on the value of the characteristic. Standard errors for intermediate values not shown in the tables may be approximated by interpolation.

Standard Errors of Estimated Numbers

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

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

In this formula f is the appropriate factor from table B-5 and σ is the standard error for total or White persons in table B-1 or the Standard error for Black and other races persons in table B-2.

Standard Errors of Estimated Percentages

The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends on 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. The approximate standard error, $\sigma_{(x,p)}$ of an estimated percentage can be obtained by use of the formula:

$$\sigma_{(x,p)} = f\sigma \qquad (2)$$

In this formula f is the appropriate factor from table B-5 and σ is the standard error for total or White persons in table B-3, or the standard error for Black and other races in table B-4. When the numerator and denominator of the percentage are in different categories, use the table and factor indicated by the numerator.

Illustration of the Use of Standard Error Tables

Table A of this report shows that in October 1976 the number of students aged 14 to 34 years maintaining their own household is 3,207,000. An estimated 9,950,000 civilian noninstitutional students 14 to 34 years of age were enrolled in college at the time. The factor in table B-5 for education, total or White is 1.0. Interpolation in table B-1 shows the standard error for an estimate of this size to be approximately 68,500[1]. Thus, using formula (1) an approximate standard error of 68,500 = (68,500 x 1) is obtained. The 68-percent confidence interval as shown by the data is from 3,138,500 to 3,275,500. Therefore, a conclusion that the average estimate derive from all possible samples lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude with 95-percent confidence that the average estimate derived from all possible samples lies within the intervals from 3,070,000 to 3,344,000, i.e., 3,207,000 $\pm$ (2 x 68, 500).

Table C shows that in October, 1976 24.5 = $(\frac{2435}{9950} \times 100)$ percent of the college students 14 to 34 years of age

[1] Rounded to the nearest hundred.

Table B-1. Generalized Standard Errors for Estimated Numbers of Persons—Total or Whites

(68 chances out of 100. Numbers in thousands)

Estimated number of persons	100	250	500	1,000	2,500	5,000	10,000	25,000	50,000	100,000
10	4.4	4.6	4.6	4.6	4.6	4.7	4.7	4.7	4.7	4.7
20	5.9	6.3	6.5	6.5	6.6	6.6	6.6	6.6	6.6	6.6
30	6.8	7.6	7.8	7.9	8.0	8.0	8.1	8.1	8.1	8.1
40	7.2	8.5	8.9	9.1	9.2	9.3	9.3	9.3	9.3	9.3
50	7.4	9.3	9.9	10.2	10.3	10.4	10.4	10.4	10.4	10.4
75	6.4	10.7	11.8	12.3	12.6	12.7	12.7	12.7	12.7	12.8
100	-	11.4	13.2	14.0	14.4	14.6	14.7	14.7	14.7	14.7
200	-	9.3	16.1	18.6	20.0	20.4	20.6	20.7	20.8	20.8
300	-	-	16.1	21.3	23.9	24.7	25.1	25.4	25.4	25.5
400	-	-	13.2	22.8	27.0	28.3	28.9	29.2	29.3	29.4
500	-	-	-	23.3	29.5	31.2	32.1	32.6	32.8	32.9
750	-	-	-	20.2	33.8	37.2	38.8	39.7	40.0	40.2
1,000	-	-	-	-	36.1	41.7	44.2	45.6	46.1	46.3
2,000	-	-	-	-	29.5	51.0	58.9	63.2	64.5	65.2
3,000	-	-	-	-	-	51.0	67.5	75.7	78.2	79.5
4,000	-	-	-	-	-	41.7	72.2	85.4	89.4	91.3
5,000	-	-	-	-	-	-	73.7	93.2	98.8	101.5
7,500	-	-	-	-	-	-	63.9	106.7	117.6	122.7
10,000	-	-	-	-	-	-	-	114.1	131.8	139.7
20,000	-	-	-	-	-	-	-	93.2	161.4	186.3
30,000	-	-	-	-	-	-	-	-	161.4	213.5
40,000	-	-	-	-	-	-	-	-	131.8	228.2
50,000	-	-	-	-	-	-	-	-	-	232.9
75,000	-	-	-	-	-	-	-	-	-	201.7
100,000	-	-	-	-	-	-	-	-	-	-

Note: To estimate the standard errors for the period 1956 to 1966, multiply these standard errors by 1.23. For years prior to 1956, multiply by 1.5.

were enrolled in 2-year colleges. An estimated 9,950,000 people 14 to 34 years of age were enrolled in colleges at the time. Table B-3 shows the standard error of 24.5 percent to be approximately .6 percent. The factor in table B-5 for education, total or White is 1.0. Hence, applying formula (2) the standard error of 24.5 percent for education is approximately .6 = (.6 x 1.0) percent. Consequently, the 68-percent confidence interval is from 23.9 to 25.1 percent and the 95-percent confidence interval is from 23.3 to 25.7 percent.

Standard error of a difference. For a difference between two sample estimates, the standard error is approximately equal to the square root of the sum of the squared standard errors of the estimates

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

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

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

(68 chances out of 100. Numbers in thousands)

Estimated number of persons	100	250	500	1,000	2,500	5,000	10,000
10	4.8	5.0	5.0	5.0	5.1	5.1	5.1
20	6.4	6.9	7.0	7.1	7.1	7.1	7.2
30	7.3	8.2	8.5	8.6	8.7	8.7	8.8
40	7.8	9.3	9.7	9.9	10.0	10.1	10.1
50	8.0	10.1	10.7	11.0	11.2	11.3	11.3
75	6.9	11.6	12.8	13.3	13.7	13.8	13.8
100	-	12.4	14.3	15.2	15.7	15.8	15.9
200	-	10.1	17.5	20.2	21.7	22.2	22.4
300	-	-	17.5	23.2	26.0	26.9	27.3
400	-	-	14.3	24.8	29.3	30.7	31.4
500	-	-	-	25.3	32.0	34.0	34.9
750	-	-	-	21.9	36.7	40.4	42.2
1,000	-	-	-	-	39.2	45.3	48.0
2,000	-	-	-	-	32.0	55.4	64.0
3,000	-	-	-	-	-	55.4	73.3
4,000	-	-	-	-	-	45.3	78.4
5,000	-	-	-	-	-	-	80.0
7,500	-	-	-	-	-	-	69.3
10,000	-	-	-	-	-	-	-

Note: To estimate the standard errors for the period 1956 to 1966, multiply these standard errors by 1.23. For years prior to 1956, multiply by 1.5.

Table B-3. Generalized Standard Errors of Estimated Percentages—Total or Whites

(68 chances out of 100)

Base of percentage (thousands)	Estimated percentages				
	2 or 98	5 or 95	10 or 90	25 or 75	50
100	2.0	3.1	4.3	6.2	7.2
250	1.3	2.0	2.8	4.0	4.5
500	0.9	1.4	1.9	2.8	3.2
1,000	0.6	1.0	1.4	2.0	2.3
2,500	0.4	0.6	0.9	1.2	1.4
5,000	0.3	0.4	0.6	0.9	1.0
10,000	0.2	0.3	0.4	0.6	0.7
25,000	0.13	0.2	0.3	0.4	0.5
50,000	0.09	0.14	0.2	0.3	0.3
100,000	0.06	0.10	0.14	0.2	0.2

Note: To estimate the standard errors for the period 1956 to 1966, multiply these standard errors by 1.23. For years prior to 1956, multiply by 1.5.

etc. This will represent the actual standard error quite accurately for the difference between two estimates of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same areas. If, however, there is a high positive (negative) correlation between the estimates of the two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration of the Computation of the Standard Error of a Difference Between Percentages. Table C shows that in October 1966, 17.4 = $(\frac{1046}{5999} \times 100)$ percent of 5,999,000 college students 14 to 34 years of age were enrolled in 2-year colleges. The apparent difference between the 24.5 percent of college student 14 to 34 years of age enrolled in 2-year colleges in October 1966 and the 17.4 percent described above is 7.1 percent. The standard error, σ_x, of 24.5 percent is .6 percent as shown previously. The factor from table B-5 appropriate for the 17.4 percent is again 1.0. Interpolation in table B-3 shows the standard error of 17.4 percent on a base of 5,999,000 to be .86 percent. Thus, the standard error, σ_y, of 17.4 percent is .86 = (.70 x 1.23 x 1.0) percent. Using formula 3, the standard error of difference, 7.1 percent, is $1.05 \doteq \sqrt{(.86)^2 + (.6)^2}$. This means that the 95-percent confidence interval is 5.0 to 9.2 percent. Thus, we can conclude with 95-percent confidence that there was a significant difference between the percentage of college students 14 to 34 years of age enrolled in 2-year colleges in October 1976 and the percentage of college students 14 to 34 years of age enrolled in 2-year colleges in October 1966.

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

(68 chances out of 100)

Base of percentage (thousands)	Estimated percentage				
	2 or 98	5 or 95	10 or 90	25 or 75	50
50	3.3	5.2	7.1	10.2	11.8
100	2.3	3.6	5.0	7.2	8.4
250	1.5	2.3	3.2	4.6	5.3
500	1.0	1.6	2.2	3.2	3.7
1,000	0.7	1.2	1.6	2.3	2.6
2,500	0.5	0.7	1.0	1.4	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

Note: To estimate the standard errors for the period 1956 to 1966, multiply these standard errors by 1.23. For years prior to 1956, multiply by 1.5.

Table B-5. "f" Factors to be Applied to Generalized Standard Errors in Tables B-1 through B-4

Type of characteristic[1]	Total or White (B-1 or B-3)		Black and other races (B-2 or B-4)	
	Persons	Families	Persons	Families
Marital status and household and family	1.3	0.8	1.3	0.7
Income	[2]1.2	0.7	[2]1.0	0.6
Residence	1.4	(X)	1.6	(X)
Kindergarten and nursery school enrollment	0.9	(X)	0.9	(X)
Educational attainment and school enrollment	1.0	(X)	1.0	(X)

[1]For metropolitan-nonmetropolitan data cross-tabulated with other data, multiply the above factor by 1.4.

[2]Persons tabulated by family income.

☆U.S. GOVERNMENT PRINTING OFFICE: 1979—311-040:288