Appendix B. Source and Reliability of the Estimates

SOURCE OF DATA

The estimates in this report are based on data collected in October of the years 1966, 1972, 1974, and 1978 from the Current Population Survey (CPS) of the Bureau of the Census. 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 and in addition, questions are asked about school enrollment in October.

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.") Previous sample designs used were based on files from the most recently completed census updated for new construction.

The estimation procedure used for the monthly CPS data involves the inflation of weighted sample results to independent estimates of the civilian noninstitutional population of the United States by age, race, and sex. These independent estimates are based on statistics from decennial censuses; 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

Description of the Current Population Survey

Time	Number of sample areas[1]	Households eligible		Housing units visited, not eligible[2]
		Interviewed	Not interviewed	
October 1966	357	33,500	1,500	6,000
October 1972 and 1974	461	45,000	2,000	8,000
October 1978[3]	614	54,000	2,500	10,000

[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.

[3]Beginning in October 1978, a coverage improvement sample for mobile homes and new construction was included.

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.

A coverage improvement sample was included in computing the estimates beginning in October 1978 in order to provide coverage of mobile homes and new construction housing units that previously had no chance for selection in the CPS sample. This sample is composed of approximately 450 sample household units which represent 237,000 occupied mobile homes and 600,000 new construction units. These new construction units are composed of those units where building permits were issued prior to January 1970 and construction was not completed by the time of the 1970 Census (i.e., April 1970). 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 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;

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 interval.

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.

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 size of estimated numbers and percentages. As a result, the sets of standard errors (along with factors) 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-6 are approximations to generalized standard errors of estimated numbers and estimated percentages. Estimated standard errors for specific characteristics cannot be obtained from tables B-1 through B-6 without the use of the factors in table B-7. These factors must be applied to the generalized standard errors in order to adjust for the combined effect of sample design and estimation procedure on the value of the characteristic. Generalized standard errors for intermediate values of estimates not shown in tables B-1 through B-6 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-7 and σ is the standard error for total or White persons in table B-1, the standard error for Black or other races persons in table B-2, or the standard error for Spanish origin persons in table B-3.

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-7 and σ is the standard error for total or White persons in table B-4, the standard error for Black and other races in table B-5, or the standard error for Spanish origin persons in table B-6.

When a percentage involves two different categories, use the larger of the two factors corresponding to those categories. For example, consider the percentage of male students from families with income of less than $10,000 per year whose major field of study was business in October 1978. The correct factor from table B-7 is 1.2 for income of total persons since it is larger than 1.0, the factor for school enrollment and major field of study of total persons.

Illustration of the use of tables of standard errors. Table B of this report shows that in October 1978 there were 4,714,000 women 14 to 34 years old enrolled in college. The factor in table B-7 for school enrollment, total or White, in 1978, is 1.0. Multiplying this factor, 1.0, with the value obtained by interpolating in table B-1 provides an approximate standard error of 96,700. The 68-percent confidence interval as shown by the data is from 4,617,300 to 4,810,700. 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 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 interval from 4,520,600 to 4,907,400, i.e. 4,714,000 ± (2 x 96,700).

Table C shows that business was the major field of study for 14.5 percent of the 282,000 Blacks 14 to 34 years old who were enrolled in college in October of 1966. The factor in table B-7 for major field of study, Black and other races, in 1966 is 1.2. Multiplying this factor, 1.2, with the value obtained by interpolating in table B-5 provides an approximate standard error of 4.2 = 3.5 x 1.2 percentage points. Consequently, the 68-percent confidence interval is from 10.3 percent to 18.7 percent and the 95-percent confidence interval is from 6.1 percent to 22.9 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; the estimates can be of numbers, percents, ratios, 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 area. 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 calculation of standard error of a difference. Table C shows that business was the major field of study for 21.6 percent of the 1,021,000 Blacks 14 to 34 years old who were enrolled in college in October of 1978. The factor in table B-7 for major field of study, Black and other races in 1978 is 1.0. Multiplying this factor, 1.0, with the value obtained by interpolating in table B-5 provides an approximate standard error of 2.1 percentage points.

The apparent difference between the 14.5 percent of Blacks who majored in business in 1966 and the 21.6 who majored in business in 1978 is 7.1 percentage points. To get the standard error of the difference, $\sigma_{(x-y)}$, use the following formula:

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

Therefore, the standard error of the difference of 7.1 percentage points is

$$4.7 = \sqrt{(4.2)^2 + (2.1)^2}$$

This means that the 68-percent confidence interval around the difference is from 2.4 to 11.8 percent, and the 95-percent confidence interval is from −2.3 percent to 16.5 percent. Since this interval contains zero, we may not conclude with 95-percent confidence that there was a higher percentage of Black students majoring in business in 1978 than in 1966.

Table B-1. Standard Errors of Estimated Numbers of Persons—Total or White Population

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
10	4.5	500	32.1
20	6.4	750	39.2
30	7.9	1,000	45.3
40	9.1	2,000	63.7
50	10.2	3,000	77.8
75	12.4	4,000	89.4
100	14.4	5,000	99.6
200	20.3	7,500	120.7
300	24.9	10,000	138.0
400	28.7	20,000	186.8

Table B-2. Standard Errors of Estimated Numbers of Persons—Black and Other Races

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
10	5.3	200	23.5
20	7.5	300	28.7
30	9.1	400	33.0
40	10.6	500	36.7
50	11.8	750	44.6
75	14.4	1,000	51.0
100	16.7	2,000	69.6

Table B-3. Standard Errors of Estimated Numbers of Persons—Spanish Origin

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
10	6.2	100	19.6
20	8.8	200	27.7
30	10.7	300	34.0
40	12.4	400	39.2
50	13.9	500	43.8
75	17.0		

Table B-4. Standard Errors of Estimated Percentages—Total or White Population

Base of percentage (thousands)	Estimated percentage				
	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	2.7	3.9	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

Table B-5. Standard Errors of Estimated Percentages—Black and Other Races

Base of percentage (thousands)	Estimated percentage				
	2 or 98	5 to 95	10 or 90	25 or 75	50
75	2.7	4.2	5.8	8.4	9.6
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
15,000	0.2	0.3	0.4	0.6	0.7

Table B-6. Standard Errors of Estimated Percentages—Spanish Origin

Base of percentage (thousands)	Estimated percentage				
	2 or 98	5 or 95	10 or 90	25 or 75	50
50	3.9	6.0	8.3	12.0	13.9
100	2.7	4.3	5.9	8.5	9.8
250	1.7	2.7	3.7	5.4	6.2
500	1.2	1.9	2.6	3.8	4.4
1,000	0.9	1.4	1.9	2.7	3.1
2,500	0.5	0.9	1.2	1.7	2.0
5,000	0.4	0.6	0.8	1.2	1.4
10,000	0.3	0.4	0.6	0.8	1.0

Table B-7. "f" Factors to be Applied to Tables B-1 Through B-6 to Approximate Standard Errors

Type of characteristic[1]	Total or White (B-1 or B-4)		Black and other (B-2 or B-5)		Spanish origin (B-3 or B-6)	
	1966	1967-78	1966	1967-78	1966	1967-78
Income[2]	1.5	1.2	1.2	1.0	1.8	1.5
School enrollment and major field of study	1.2	1.0	1.2	1.0	1.2	1.0
Marital status	1.6	1.3	1.6	1.3	1.7	1.4
Occupation	1.2	1.0	1.1	0.9	1.0	0.9

[1]For metropolitan-nonmetropolitan data cross-tabulated with other data, also apply the factor 1.4 as well as the factor indicated in the table.

[2]Persons tabulated by family income.

Note: Apply these factors to the standard error tables to obtain appropriate standard errors for the characteristic of interest.