Appendix C. Source and Reliability of the Estimates

SOURCE OF DATA

The estimates in this report are based on data collected in October of the years 1967 through 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," below). Previous sample designs used, as a basis, 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 period.

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 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 types of characteristics. 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 C-2 through C-7 provide approximations to standard errors of various estimates shown in this report. Estimated standard errors cannot be obtained from tables C-2 through C-4, however, without the use of table 1 of this report. The numbers in table 1 or combinations of them, correspond to the total-persons-in-age-group for October 1978 and should be used in conjunction with the column headings in tables C-2 through C-4. The total-persons-in-age-group data for previous years may be obtained in the appropriate Current Population Reports on school enrollment.

Table C-8 provides factors which must be used to calculate standard errors for each characteristic. 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. For example, to produce approximate standard errors for the marital status of Spanish-origin persons, multiply the appropriate figures in tables C-4 or C-7 by the factor 1.4 from table C-8. The determination of the proper factor for a percentage depends upon the subject matter of the numerator of the percentage, not the denominator. For example, if a percent referred to the percentage of males enrolled in college whose families had income of less than $10,000 per year in October 1978, then the factor 1.2 for income of total persons is used.

Table C-1. CPS Sample Sizes From October 1967 to October 1978

Time period	Number of sample areas[1]	Households eligible		Housing units visited, not eligible[2]
		Interviewed	Not interviewed	
October 1978[3]	614	54,000	2,500	10,000
October 1977[4]	614	53,500	2,500	9,500
October 1972	461	45,000	2,000	8,000
October 1971	449	45,000	2,000	8,000
October 1967	449	48,000	2,000	8,500

[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.

[4]A supplementary sample of housing units in 24 States and the District of Columbia was incorporated with the monthly CPS to produce October 1977 data.

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 C-8 and σ is the standard error for total or White persons in table C-2, the standard error for Black and other races persons in table C-3, or the standard error for Spanish-origin persons in table C-4.

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 C-8 and σ is the standard error for total or White persons in table C-5, the standard error for Black and other races in table C-6, or the standard error for Spanish-origin persons in table C-7. 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 tables of standard errors. Table B of this report shows that in October 1978 there were 1,507,000 women 18 and 19 years old enrolled in college. The total number of women 18 and 19 years old from table B is 4,178,000. The factor in table C-8 for education, total or White, is 1.0. Using both the 4,178,000 estimated women in the age group and the estimated 1,507,000 women in the age group in college with table C-2 and formula (1), an approximate standard error of 40,900 = (40,900 x 1.0) is obtained. The chances are 68 out of 100 that the estimate would have been a figure differing from the average of all possible samples by less than 40,900. The chances are 95 out of 100 that the estimate would have been a figure differing from the average of all possible samples by less than 81,800 (twice the standard error).

Table B shows that 30.7 percent of the 3,911,000 20- and 21-year-old White males were enrolled in college in October of 1978. The factor in table C-8 for education, total or White is 1.0. Interpolation in table C-5 shows the standard error of 30.7 percent to be 1.1 percent. Thus, the standard error of 30.7 percent for education is 1.1 = (1.1 x 1.0). Consequently, the chances are 68 out of 100 that the estimated percent would be within 1.1 percentage points of the average of all possible samples. Chances are 95 out of 100 that the estimate would be within 2.2 percentage points of the average of all possible samples, i.e., the 95-percent confidence interval would be from 28.5 to 32.9.

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 (understimate) the true standard error.

Table H shows that 8.9 percent of the 3,621,000 18- to 24-year-old male college students were married and living with spouse in October of 1978. The factor in table C-8 for marital status, total or White is 1.3. Interpolation in table C-5 shows the standard error of 8.9 percent to be 0.7 percentage points. Thus, the standard error of 8.9 percent for marital status is 0.9 = (0.7 x 1.3).

Table H also shows that 28.2 percent of the 9,764,000 18- to 24-year-old males not enrolled in college were married and living with spouse in October of 1978. Similar to the above, the standard error of 28.2 percent for marital status is 0.8 = (0.6 x 1.3) percentage points.

The apparent difference between the 8.9 percent of male college students who are married and the 28.2 percent of males not enrolled in college who are married is 19.3 percentage points.

To get the standard error of the estimated difference, $\sigma_{(x-y)}$, use the following formula:

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

Therefore the standard error of the difference of 19.3 percent is

$$1.2 \doteq \sqrt{(.9)^2 + (.8)^2}$$

This means the chances are 68 out of 100 that the estimated difference based on the sample estimates would vary from the difference derived from the average of all possible samples by less than 1.2 percentage points. The 68-percent confidence interval about the 19.3 percentage point difference is from 18.1 to 20.5, i.e., 19.3 $\pm$ 1.2. A conclusion that the average estimate of the difference derived from all possible samples of the same size and design lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. The 95-percent confidence interval is 16.9 to 21.7. Since this interval does not contain zero, we can conclude with 95-percent confidence that in 1978 there was a significant difference between the percentage of 18- to 24-year-old males enrolled in college who were married and the percentage of 18- to 24-year-old males not enrolled in college who were married.

Table C-2. Standard Errors for Estimated Numbers of Persons: Total or White

(Numbers in thousands. For meaning of symbols, see text)

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	4.3	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5
20	5.7	6.2	6.3	6.4	6.4	6.4	6.4	6.4	6.4	6.4
30	6.6	7.4	7.6	7.8	7.8	7.8	7.9	7.9	7.9	7.9
40	7.0	8.3	8.7	8.9	9.0	9.0	9.1	9.1	9.1	9.1
50	7.2	9.1	9.6	9.9	10.1	10.1	10.1	10.1	10.2	10.2
75	6.2	10.4	11.5	12.0	12.3	12.3	12.4	12.4	12.4	12.4
100	-	11.1	12.9	13.6	14.1	14.2	14.3	14.3	14.4	14.4
200	-	9.1	15.7	18.2	19.5	19.9	20.1	20.2	20.3	20.3
300	-	-	15.7	20.8	23.3	24.1	24.5	24.7	24.8	24.8
400	-	-	12.9	22.3	26.3	27.6	28.2	28.5	28.6	28.7
500	-	-	-	22.7	28.7	30.5	31.3	31.8	32.0	32.0
750	-	-	-	19.7	32.9	36.3	37.8	38.8	39.1	39.2
1,000	-	-	-	-	35.2	40.6	43.1	44.5	45.0	45.2
2,000	-	-	-	-	28.7	49.8	57.5	61.6	63.0	63.6
3,000	-	-	-	-	-	49.8	65.8	73.8	76.3	77.5
4,000	-	-	-	-	-	40.6	70.4	83.3	87.2	89.0
5,000	-	-	-	-	-	-	71.8	90.9	96.4	99.0
7,500	-	-	-	-	-	-	62.2	104.1	114.7	119.7
10,000	-	-	-	-	-	-	-	111.3	128.5	136.3
20,000	-	-	-	-	-	-	-	90.9	157.4	181.7
30,000	-	-	-	-	-	-	-	-	157.4	208.2
40,000	-	-	-	-	-	-	-	-	128.5	222.6
50,000	-	-	-	-	-	-	-	-	-	227.2
75,000	-	-	-	-	-	-	-	-	-	196.7
100,000	-	-	-	-	-	-	-	-	-	-

Table C-3. Standard Errors for Estimated Numbers of Persons: Black and Other Races

(Numbers in thousands. For meaning of symbols, see text)

Estimated number of persons	Total persons in age group							
	100	250	500	1,000	2,500	5,000	10,000	15,000
10	5.0	5.2	5.2	5.3	5.3	5.3	5.3	5.3
20	6.7	7.2	7.3	7.4	7.4	7.5	7.5	7.5
30	7.7	8.6	8.9	9.0	9.1	9.1	9.1	9.1
40	8.2	9.7	10.1	10.4	10.5	10.5	10.5	10.6
50	8.4	10.6	11.2	11.5	11.7	11.8	11.8	11.8
75	7.2	12.1	13.3	13.9	14.3	14.4	14.4	14.4
100	-	12.9	14.9	15.9	16.4	16.5	16.6	16.7
200	-	10.6	18.3	21.1	22.7	23.2	23.4	23.5
300	-	-	18.3	24.2	27.1	28.1	28.5	28.7
400	-	-	14.9	25.9	30.6	32.1	32.7	33.0
500	-	-	-	26.4	33.4	35.4	36.4	36.7
750	-	-	-	22.9	38.3	42.2	44.0	44.6
1,000	-	-	-	-	40.9	47.3	50.1	51.0
2,000	-	-	-	-	33.4	57.9	66.8	69.6
3,000	-	-	-	-	-	57.9	76.6	81.9
4,000	-	-	-	-	-	47.3	81.9	90.5
5,000	-	-	-	-	-	-	83.5	96.5
7,500	-	-	-	-	-	-	72.4	102.3
10,000	-	-	-	-	-	-	-	96.5
15,000	-	-	-	-	-	-	-	-

Table C-4. Standard Errors for Estimated Numbers of Persons: Spanish Origin

(Numbers in thousands. For meaning of symbols, see text)

Estimated number of persons	Total persons in age group						
	100	250	500	1,000	2,500	5,000	10,000
10	5.9	6.1	6.1	6.2	6.2	6.2	6.2
20	7.8	8.4	8.6	8.7	8.7	8.8	8.8
30	9.0	10.1	10.4	10.6	10.7	10.7	10.7
40	9.6	11.4	11.9	12.2	12.3	12.4	12.4
50	9.8	12.4	13.2	13.5	13.7	13.8	13.8
75	8.5	14.2	15.7	16.3	16.7	16.9	16.9
100	-	15.2	17.6	18.6	19.2	19.4	19.5
200	-	12.4	21.5	24.8	26.6	27.2	27.5
300	-	-	21.5	28.4	31.9	33.0	33.5
400	-	-	17.6	30.4	36.0	37.6	38.5
500	-	-	-	31.0	39.2	41.6	42.8
750	-	-	-	26.9	45.0	49.5	51.7
1,000	-	-	-	-	39.3	55.5	58.9
2,000	-	-	-	-	-	68.0	78.5
3,000	-	-	-	-	-	68.0	89.9
4,000	-	-	-	-	-	55.5	96.1
5,000	-	-	-	-	-	-	98.1
7,500	-	-	-	-	-	-	85.0
10,000	-	-	-	-	-	-	-

Table C-5. 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
50,000	0.09	0.14	0.2	0.3	0.3
100,000	0.06	0.10	0.14	0.2	0.2

Table C-6. Standard Errors of Estimated Percentages: Black and Other Races

Base of percentage (thousands)	Estimated percentage				
	2 or 98	5 or 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 C-7. Standard Errors of Estimated Percentages: Persons of Spanish Origin

Base of percentage (thousands)	Estimated percentages				
	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 C-8. "f" Factors to be Applied to Tables C-2 Through C-7 to Approximate Standard Errors

(For meaning of symbols, see text)

Type of characteristic[1]	Total or White (C-2 or C-5)		Black and other (C-3 or C-6)		Spanish origin (C-4 or C-7)	
	Persons	Families	Persons	Families	Persons	Families
Marital status and household and family..	1.3	0.8	1.3	0.7	1.4	0.8
Income	[2]1.2	0.7	[2]1.0	0.6	[2]1.2	0.8
Residence	1.4	...	1.6	...	...	...
Kindergarten and nursery school enrollment	0.9	...	0.9	...	0.8	...
Educational attainment and school enrollment	1.0	...	1.0	...	1.0	...

[1]For metropolitan-nonmetropolitan data cross-tabulated with other data, also apply the factor 2.0 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 characteristics of interest.

☆U.S. GOVERNMENT PRINTING OFFICE: 1979—311-040 208