Appendix

DEFINITIONS AND EXPLANATIONS

Population coverage. The figures in this report for March 1977 and March 1976 are sample survey data and relate to the noninstitutional population of the 50 States and the District of Columbia. Members of the Armed Forces living off post are included, but all other members of the Armed Forces are excluded. The March 1977 survey included 954,000 members of the Armed Forces and the March 1976 survey includes 914,000 members of the Armed Forces in the United States living off post or with their families on post.

Population figures shown for States and SMSA's are based on weighted results of the Current Population Survey and have not been adjusted to independent Census Bureau estimates.

Age. The age classification is based on the age of the person at his last birthday.

Race. The population is divided into three groups on the basis of race: White, Black, and "other races." The last category includes Indians, Japanese, Chinese, and any other race except White and Black. In this report, "other races" are usually shown in combination with the Black population.

Origin or descent. Data on ethnic origin or descent were derived from responses to the question, "What is (this person's) origin or descent?" No attempt was made to define the terms but a flashcard was shown or a list was read, illustrative of types of ethnic origin. Some respondents having a diverse ethnic background or having several generations of residence in the United States may have reported the ethnic association they felt most strongly. For more detail, see **Current Population Reports,** Series P-20, No. 310, "Persons of Spanish Origin in the United States: March 1976."

Years of school completed. Data on years of school completed in this report were derived from the combination of answers to two questions: (a) "What is the highest grade of school he has ever attended?" and (b) "Did he finish this grade?"

The questions on educational attainment apply only to progress in "regular" schools. Such schools include graded public, private, and parochial elementary and high schools (both junior and senior high), colleges, universities, and professional schools, whether day schools or night schools. Thus, regular schooling is that which may advance a person toward an elementary school certificate or high school diploma, or a college, university, or professional school degree. Schooling in other than regular schools was counted only if the credits obtained were regarded as transferable to a school in the regular school system.

The median years of school completed is defined as the value which divides the population into two equal parts—one half having completed more schooling and one-half having completed less schooling than the median. This median was computed after the statistics on years of school completed had been coverted to a continuous series of numbers (e.g., completion of the first year of high school was treated as completion of the 9th year and the completion of the first year of college as completion of the 13th year). The persons completing a given school year were assumed to be distributed evenly within the interval from .0 to .9 of the year (for example, persons completing the 12th year were assumed to be distributed evenly between 12.0 and 12.9). In fact, at the time of the March survey, most of the enrolled persons had completed about three-fourths of a school year beyond the highest grade completed, whereas a large majority of persons who were not enrolled had not attended any part of a grade beyond the highest one completed. The effect of the assumption is to place the median for younger persons slightly below, and for older persons' slightly above, the true median. Because of the inexact assumption as to the distribution within an interval, this median is more appropriately used for comparing groups and the same group at different dates than as an absolute measure of educational attainment.

Assignment of educational attainment for those not reporting. When information on either the highest grade attended or completion of the grade was not reported in the March survey, entries for the items were assigned using an edit in the computer. The general procedure was to assign an entry for a person that was consistent with entries for other persons with similar characteristics. The specific technique used in the March survey was as follows:

1. The computer stored reported data on highest grade attended by race (White and all other) and age, and on completion of the grade by age and highest grade attended, for persons 14 years old and over in the population.

2. Each stored value was retained in the computer only until a succeeding person having the same characteristics (e.g., same race and age, in the case of assignments for highest grade attended) and having the item reported, was processed through the computer. Then the reported data for the succeeding person were stored in place of the one previously stored.

3. When one or both of the education items for a person 14 years old and over was not reported, the entry assigned to this person was that stored for the last person who had the same characteristics.

Geographic regions. The four major regions of the United States for which data are presented in this report represent groups of States as follows:

Northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont.

North Central: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin.

South: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia.

West: Arizona, Colorado, California, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming, Alaska, and Hawaii.

Metropolitan-nonmetropolitan residence. The population residing in standard metropolitan statistical areas (SMSA's) constitutes the metropolitan population. Except in New England, an SMSA is a county or group of contiguous counties which contain at least one city of 50,000 inhabitants or more, or "twin cities" with a combined population of at least 50,000. In addition to the county or counties containing such a city or cities, contiguous counties are included in an SMSA if, according to certain criteria, they are essentially metropolitan in character and are socially and economically integrated with the central city. In New England, SMSA's consist of towns and cities, rather than counties.

The figures shown in this report for standard metropolitan statistical areas (SMSA's) are based on the SMSA's as defined on the basis of the 1970 census.

Family status. Within households, persons who are family members are distinguished from those who are not family members. A family is defined as a group of two or more persons residing together who are related by blood, marriage, or adoption; all such persons are considered as members of one family even though they may include a "subfamily," that is, a married couple or a parent-child group sharing the living quarters of the family head. A primary family includes among its members the head of a household, and a "secondary family" includes no member related to the household head. Members of secondary families may include persons such as lodgers, guests, or resident employees and their relatives living in a household.

Persons living with relatives in group quarters were formerly considered as members of secondary families. However, the number of such families became so small (37,000 in 1967) that beginning with data for 1968 (and beginning with the census data for 1960) the Bureau of the Census includes persons in these families in the count of secondary individuals.

In this report, persons who are not family members (that is, are not recognized as living with any relatives) are "unrelated individuals." Unrelated individuals are of two types--primary and secondary. A "primary individual" is a household head with no relatives in the household. A "secondary individual" is a lodger, partner, guest, or resident employee with no relatives in the household or group quarters.

Primary family. A primary family is a family that includes among its members the head of a household.

Secondary family. A secondary family is a family that does not include among its members the head of a household. Members of secondary families may include persons such as guests, lodgers, or resident employees and their relatives living in a household.

Primary individual. A primary individual is a household head living alone or with nonrelatives only.

Secondary individual. A secondary individual is a person in a household or group quarters such as a guest, lodger, or resident employee (excluding primary individuals and inmates of institutions) who is not related to any other person in the household or group quarters. (See section above on "Secondary family" for slight change in coverage of secondary individuals in 1968.)

Group quarters. Group quarters are defined as noninstitutional living arrangements for groups containing five or more persons unrelated to the person in charge. Inmates of institutions (starting in 1972) are not included in the Current Population Survey.

Marital status. The marital status classification identifies four major categories: single, married, widowed, and divorced. These terms refer to the marital status at the time of the enumeration.

The category "married" is further divided into "married, spouse present," "separated," and "other married, spouse absent." A person was classified as "married, spouse present" if the husband or wife was reported as a member of the household, even though he or she may have been temporarily absent on business or on vacation, visiting, in a hospital, etc., at the time of the enumeration. Persons reported as separated included those with legal separations, those living apart with

intentions of obtaining a divorce, and other persons permanently or temporarily separated because of marital discord. The group "other married, spouse absent" includes married persons living apart because either the husband or wife was employed, and living at a considerable distance from home, was serving away from home in the Armed Forces, had moved to another area, or had a different place of residence for any other reason except separation as defined above.

Symbols. A dash "—" represents zero or rounds to zero, and the symbol "B" means that the base is less than 75,000. Three dots ". . ." mean not applicable.

Rounding of estimates. Individual figures are rounded to the nearest thousands without being adjusted to group totals, which are independently rounded. Percentages are based on the unrounded absolute numbers.

SOURCE AND RELIABILITY OF THE ESTIMATES

Source of data. The estimates in this report are based on data collected in March of 1970, 1976, and 1977 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. In addition, questions are asked about education each month and supplementary questions are asked each March about various population characteristics.

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.

The following table provides a description of some aspects of the CPS sample designs in use during the referenced data-collection periods.

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. For this report, persons in the Armed Forces living off post or with their families on post are also included. The estimation procedure for the data in the report also involves a further adjustment so that husband and wife of a household receive the same weight.

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 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. As in any survey work, the results are subject to errors of response and nonreporting in addition to sampling 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 to provide correct information on the part of respondents, inability to recall information, mistakes made in collection such as in recording or coding the data, mistakes made in processing the data, mistakes made in estimating values for missing data, and failure to represent all units with the sample (undercoverage).

Description of the Current Population Survey

Time period	Number of sample areas[1]	Households eligible		Households visited, not eligible[2]
		Interviewed	Not interviewed	
March 1977[3]	614	53,500	2,500	9,500
March 1976	461	45,000	2,000	8,000
March 1970	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 households which were visited but were found to be vacant or otherwise not eligible for interview.

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

The approximate magnitude of three sources of undercoverage in CPS is known. About 600,000 conventional new construction units (housing units other than mobile homes or group quarters) were assigned building permits prior to the 1970 census but building was not completed by the time of the census (i.e., April 1970); these units have no representation in the CPS sample. 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. In addition, about 70,000 units are not represented either because they have been converted from nonresidential units or are houses moved to different sites since the census. The extent of other sources of undercoverage is unknown. Note that 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.

Sampling variability. The standard errors given in the tables are primarily measures of sampling variability, that is, of the variations that occur by chance because a sample rather than the whole of the population is surveyed. The chances are about 68 out of 100 that an estimate from the survey differs from a complete census figure by less than the standard error. The chances are about 90 out of 100 that this difference would be less than 1.6 times the standard error and about 95 out of 100 that the difference would be less than twice the standard error.

All 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 the phrase, "some evidence") have a level of significance between 1.6 and 2.0 standard errors.

Note when using small estimates. Summary measures (such as medians and percent distributions) are shown in the report 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 the 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 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 provide approximations to standard errors of various estimates for total, White, or Black persons in the total U.S. for education only. To obtain standard errors for other characteristics, factors from table A-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.

Illustration of the use of standard error tables. Table 1 of this report shows that in March 1977 there were 75,239,000 White females aged 14 years old and over. An estimated 4,147,000 of them had, at that time, completed one year of college. Table A-1 shows the standard error of an estimate of this size to be approximately 90,000.[1] The chances are 68 out of 100 that the estimate would have been a figure differing from a complete census figure by less than 180,000 (twice the standard error).

Table 7 of this report shows that 66.9 percent of the 14,656,000 males aged 14 years and over in the West had completed four years of high school as of March 1977. Table A-3 shows the standard error of 66.9 percent on a base of 14,656,000 is approximately 0.6 percent. Applying the appropriate factor for the West from table A-5, the approximate standard error of the percentage is 1.1 x 0.6 = 0.7. Consequently, chances are 68 out of 100 that the estimated 66.9 percent would be within 0.7 percentage points of a complete census figure; chances are 95 out of 100 that the estimate would be within 1.4 percentage points of a complete census figure, i.e., the 95-percent confidence interval would be from 65.5 to 68.3.

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

$$\sigma_{(x-y)} \doteq \sqrt{\sigma_x^2 + \sigma_y^2} \qquad (1)$$

where σ_x and σ_y are the standard errors of the estimates x and y; the estimates can be of numbers, percents, medians, 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 two separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error.

[1] Rounded to the nearest thousand.

Table A-1. Standard Errors for Estimated Numbers of Persons

Total or White Population

(68 chances out of 100. 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	9.1	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 A-2. Standard Errors for Estimated Numbers of Persons

Black and Other Races

(68 chances out of 100. 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	5.0	5.2	5.2	5.3	5.3	5.3	5.3
20	6.7	7.2	7.3	7.4	7.4	7.5	7.5
30	7.7	8.6	8.9	9.0	9.1	9.1	9.1
40	8.2	9.7	10.1	10.4	10.5	10.5	10.5
50	8.4	10.6	11.2	11.5	11.7	11.8	11.8
75	7.2	12.1	13.3	13.9	14.3	14.4	14.4
100	-	12.9	14.9	15.9	16.4	16.5	16.6
200	-	10.6	18.3	21.1	22.7	23.2	23.4
300	-	-	18.3	24.2	27.1	28.1	28.5
400	-	-	14.9	25.9	30.6	32.1	32.7
500	-	-	-	26.4	33.4	35.4	36.4
750	-	-	-	22.9	38.3	42.2	44.0
1,000	-	-	-	-	40.9	47.3	50.1
2,000	-	-	-	-	33.4	57.9	66.8
3,000	-	-	-	-	-	57.9	76.6
4,000	-	-	-	-	-	47.3	81.9
5,000	-	-	-	-	-	-	83.5
7,500	-	-	-	-	-	-	72.4
10,000	-	-	-	-	-	-	-

Illustration of the computation of the standard error of a difference between estimated percentages. Table 7 of this report also shows that of the 15,347,000 females aged 14 years old and over living in the West, 67.5 percent had completed 4 years of high school. Thus the apparent difference between males and females living in the West and completing 4 years of high school is 0.6 percent. The standard error (σ_x) of 66.9 percent is 0.7, as shown above. Table A-3 and factor table A-5 show the standard error (σ_y) of 67.5 percent with a base of 15,347,000 to be approximately 0.7 percentage points. Therefore, using formula (1), the standard error of the difference of 0.6 percent is about

$$1.0 \doteq \sqrt{(0.7)^2 + (0.7)^2}$$

This means the chances are 68 out of 100 that the estimated difference based on the sample estimates would vary from the difference derived using complete census figures by less than 1.0 percent. The 68-percent confidence interval about the 0.6 percent difference is from −0.4 to 1.6 percent, i.e., 0.6 ± 1.0. 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. Since this interval includes zero, we cannot conclude with 68-percent confidence that there is a difference between the percentage of males and females living in the West who have completed 4 years of high school.

Standard error of a median. The sampling variability of an estimated median depends upon the form of the distribution as well as on the size of its base. An approximate method for measuring the reliability of a median is to determine an interval about the estimated median, such that there is a stated degree of confidence that the median based on a complete census lies within the interval. The following procedure may be used to estimate the 68-percent confidence limits of a median based on sample data.

1. Determine, using table A-3 or A-4 and the appropriate base, the standard error on a 50-percent characteristic;
2. add to and subtract from 50 percent the standard error determined in step (1);
3. using the distribution of the characteristic, calculate the confidence interval corresponding to the two points established in step (2).

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step (1).

Illustration of the computation of a confidence interval for a median. Table 2 of this report shows that the median number of school years completed by Blacks, aged 25 to 29 years, is 12.6.

1. There was a total of 1,899,000 persons in that age group. From table A-4 the standard error of a 50-percent characteristic is found to be approximately 2.1 percent.
2. To obtain a 95-percent confidence interval on an estimated median, add to and subtract from 50 percent twice the standard error found in step (1). This yields percent limits of 45.8 and 54.2.
3. From table 2 it can be seen that 823,000, or 43.3 percent, Blacks, aged 25 to 29 years, completed 12 years of school, and 486,000, or 25.6 percent, completed less than 12 years. By linear interpolation, the lower limit on the estimate is found to be about

$$12.0 + (13.0 - 12.0)\frac{45.8 - 25.6}{43.3} \doteq 12.5 \text{ years.}$$

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

$$12.0 + (13.0 - 12.0)\frac{54.2 - 25.6}{43.3} \doteq 12.7 \text{ years.}$$

Hence, the 95-percent confidence interval about the estimated median is 12.5 to 12.7 years.

Standard error of a three-year moving average. To calculate the standard error of a three-year moving average, first find the standard error for the estimate of the middle year using the appropriate table(s). Then divide this number by the square root of three (1.7) to obtain the approximate standard error of the average.

Table A-3. Standard Errors of Estimated Percentages

Total or White Population

(68 chances out of 100)

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
150,000	0.05	0.08	0.11	0.2	0.2

Table A-4. 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
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
20,000	0.2	0.3	0.4	0.5	0.6

Table A-5. Factors to be Applied to Standard Errors in Tables A-1 Through A-4

Type of characteristic	Total or White[1]	Black[2]	Spanish origin[1]
States:			
California	1.2	1.2	2.9
Florida	1.3	1.3	1.4
Georgia	1.3	1.3	2.6
Illinois	1.2	1.2	1.3
Indiana	1.2	1.2	2.4
Massachusetts	1.1	1.1	2.2
Michigan	1.3	1.3	2.6
Missouri	1.2	1.2	2.4
New Jersey	1.2	1.2	1.3
New York	1.2	1.2	1.3
North Carolina	1.5	1.5	2.9
Ohio	1.2	1.2	2.4
Pennsylvania	1.2	1.2	2.4
Texas	1.4	1.4	2.9
Virginia	1.7	1.7	2.9
Regions:			
Northeast	1.1	1.1	1.1
North Central	1.2	1.2	1.3
South	1.2	1.2	1.3
West	1.1	1.1	1.1
SMSA's	1.1	1.1	1.1
Marital status	1.3	1.3	1.5
Household relationship			
Head, wife, or primary individual	0.8	0.7	0.8
Child or other relative in primary family, secondary family member, secondary individual, or persons living in group quarters	1.3	1.3	1.5
Education	1.0	1.0	1.1

[1]Apply the factors in this column to tables A-1 and A-3 only.
[2]Apply the factors in this column to tables A-2 and A-4 only.