

Appendix

DEFINITIONS AND EXPLANATIONS

Population coverage. The data presented here are for the civilian noninstitutional population 14 to 34 years old.

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 contains 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 metropolitan population in this report is based on SMSA's as defined in the 1970 census and does not include any subsequent additions or changes.

Central cities. Each SMSA must include at least one central city, and the complete title of an SMSA identifies the central city or cities. If only one central city is designated, then it must have 50,000 inhabitants or more. The area title may include, in addition to the largest city, up to two city names on the basis and in the order of the following criteria: (1) The additional city has at least 250,000 inhabitants or (2) the additional city has a population of one-third or more of that of the largest city and a minimum population of 25,000. An exception occurs where two cities have contiguous boundaries and constitute, for economic and social purposes, a single community of at least 50,000, the smaller of which must have a population of at least 15,000.

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

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

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

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

Age. The age classification is based on the age of the person at his or her 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.

Persons of Spanish origin are persons who reported themselves as Mexican-American, Chicano, Mexican, Mexicano, Puerto Rican, Cuban, Central or South American, or other Spanish origin. However, all persons who reported themselves as Mexican-American, Chicano, Mexican, or Mexicano were combined into the one category: Mexican. Persons of Spanish origin may be of any race.

Family. The term "family," as used here, refers to a group of two persons or more related by blood, marriage, or adoption and residing together; all such persons are considered as members of one family.

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

Head of family. One person in each family residing together was designated as the head. The head of a family is usually the person regarded as the head by members of the family. Women are not classified as heads if their husbands are resident members of the family at the time of the survey.

High school seniors. Persons were classified as high school seniors who were enrolled in the fourth year of a "regular" high school in October 1975. As defined in the survey, a "regular" high school is one which may advance a person toward a high school diploma. Examples of schools which are not regarded as "regular" schools are private business and trade schools, such as television repair schools, beautician schools, and secretarial schools.

College plans. Information on college plans was derived from responses of high school seniors in October 1975 to questions as to whether they planned to attend college, and if so the type of college they planned to attend (two-year, four-year, or both). If the students did not plan to attend college, they were asked whether they planned to attend any other type of school (see facsimile of questions below).

<p>46. Does . . . plan to attend a two-year community or junior college?</p> <p>Yes.. <input type="radio"/></p> <p>Maybe <input type="radio"/> } (Ask 47) <input checked="" type="checkbox"/></p> <p>No... <input type="radio"/></p>
<p>47. Does . . . plan to attend a four-year college or university?</p> <p>Yes.. <input type="radio"/></p> <p>Maybe <input type="radio"/> } (Fill 48)</p> <p>No... <input type="radio"/></p>
<p>48. INTERVIEWER CHECK ITEM:</p> <p><input checked="" type="checkbox"/> Entry of "Yes" or "Maybe" in item 46 or 47 <input type="radio"/> (End questions)</p> <p><input checked="" type="checkbox"/> Entry of "No" in items 46 and 47 <input type="radio"/> (Ask 49)</p>
<p>49. Does . . . plan to attend any other school, such as a business college, barber college, technical or trade school, or hospital school of nursing?</p> <p>Yes.. <input type="radio"/></p> <p>Maybe <input type="radio"/> } (End questions)</p> <p>No... <input type="radio"/></p> <p><input checked="" type="checkbox"/> <input checked="" type="checkbox"/></p>

Control of school. In this report, a public school is defined as any educational institution operated by publicly elected or appointed school officials and supported by public funds. Private schools included educational institutions established and operated by religious bodies, as well as those which are under other private control. In cases where enrollment was in a school or college which was both publicly and privately controlled or supported, enrollment was counted according to whether it was primarily public or private.

Occupation. Data on occupation are shown for the employed and relate to the job held during the survey week. Persons employed at two or more jobs were reported in the job at which they worked the greatest number of hours during the week. The major groups used here are generally the major groups used in the 1970 Census of Population. The composition of these groups is shown in 1970 Census of Population, Vol. I, **Characteristics of the Population, U.S. Summary**, chapter C.

Family income. Income as defined in this report represents the combined total money income of the family before deductions for personal taxes, Social Security, bonds, etc. It is the algebraic sum of money wages and salaries, net income from self-employment, and income other than earnings received by all family members during the 12 months prior to the surveys. It should be noted that although the family income statistics refer to receipts during the previous 12 months, the characteristics of the person, such as age, marital status, etc., and the composition of families refer to the date of the survey.

The detailed income tables include in the lowest income group (under \$5,000) those who were classified as having no income in the previous 12 months and those reporting a loss in net income from farm and nonfarm self-employment or in rental income.

Rounding of estimates. Individual figures are rounded to the nearest thousand without being adjusted to group totals, which are independently rounded. With few exceptions, percentages are based on the unrounded absolute numbers.

SOURCE AND RELIABILITY OF THE ESTIMATES

Source of data. Most of the estimates contained in these tables are based on data obtained from a supplement to the Current Population Survey (CPS) in October 1975. Also, some of the estimates are based on data obtained from similar supplements to CPS in October 1972, 1973, and 1974. The remainder of the data is from the National Longitudinal Survey of the High School Class of 1972 of the National Center for Education Statistics.

The CPS sample was initially selected from the 1970 census files and is updated continuously to reflect new construction where possible (see section "Nonsampling Variability" below). This sample is spread over 461 areas comprising 923 counties and independent cities. These areas are chosen to provide coverage in each State and the District of Columbia. Approximately 47,000 occupied households

are eligible for interview each month. Of this number, 2,000 occupied units, on the average, are visited but interviews are not obtained because the occupants are not found at home after repeated calls or are unavailable for some other reason. In addition to the 47,000, there are also about 8,000 sample units in an average month which are visited but are found to be vacant or otherwise not to be interviewed.

The CPS deals mainly with labor force data. Questions relating to labor force participation are asked about each member 14 years old or older in the 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 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.

The National Longitudinal Survey of the High School Class of 1972 is based on a probability sample of 21,600 seniors from 1,200 randomly selected public and nonpublic (Catholic and non-Catholic) schools. These sample schools provide coverage in all 50 States and the District of Columbia. A more detailed description of the design of this survey can be found in the **National Longitudinal Study of the High School Class of 1972 - Tabular Summary of Student Questionnaire Data, Volume I**. This volume also contains the survey questions, tables of data from the survey, and data relating to the reliability of estimates from the survey.

Reliability of the estimates. Since the estimates in these tables 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 schedules, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey—sampling and non-sampling. For estimates in this report, indications of the magnitude of sampling error are provided, but the 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. 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). The approximate magnitude of two sources of undercoverage in CPS is known and is described next.

Approximately 600,000 conventional new construction units were issued 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. Conventional new construction, for which building permits were issued after the census, is represented. In addition to undercoverage of conventional new construction, CPS misses approximately 1/4 of all new mobile homes (i.e., 400,000 units). These are missed because there is no systematic sampling procedure to provide representation of mobile homes constructed since the 1970 census.

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 was surveyed. As calculated, the standard error also partially measures the effect of certain response and enumeration errors, but it does not measure any systematic biases in the data. 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 chances are 95 out of 100 that the difference would be less than twice the standard error.

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.

Note on comparisons with data from other surveys. Data obtained from the Current Population Survey and other surveys and sources are not entirely comparable, due in large part to differences in interviewer training and experience and in the 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 between these different sources.

Standard error tables and their use. Instead of providing individual standard error tables for each characteristic of interest, generalized standard error tables for estimated numbers and estimated percentages, by race, are provided to conserve space. In all the standard error tables, standard errors for intermediate values not shown may be approximated by interpolation. In addition, where two or more

items have nearly equal standard errors, such as total population and White population, one table is used to represent them. As a result, the tables of standard errors (along with the factors) provide an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

The figures presented in tables A-1 through A-4 provide approximations to standard errors of various estimates shown in this report. Table A-5 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 Spanish education, multiply the appropriate figures in tables A-1 or A-3 by the factor 1.37 from table A-5. 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 number of high school seniors who plan to attend college and who live in the metropolitan areas, relative to all high school seniors who plan to attend college, then the factor for type of residence would be used.

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

$$\sigma_x = f\sigma. \quad (1)$$

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

Table A-1. Standard Errors of Estimated Numbers of Persons: Total, White, or Spanish Population

(68 chances out of 100.
Numbers in thousands)

Size of estimate	Standard error
50.....	10.2
100.....	14.4
250.....	22.7
500.....	32.1
1,000.....	45.3
2,000.....	63.8
3,000.....	77.8
4,000.....	89.5

Note: For a particular characteristic, see table A-5 for the appropriate factor to apply to the above standard errors.

Table A-2. Standard Errors of Estimated Numbers of Persons: Black and Other Races

(68 chances out of 100.
Numbers in thousands)

Size of estimate	Standard error
10.....	5.3
20.....	7.5
30.....	9.1
40.....	10.6
50.....	11.8
75.....	14.4
100.....	16.7
200.....	23.5
300.....	28.6
400.....	33.0
500.....	36.7

Note: For a particular characteristic, see table A-5 for the appropriate factor to apply to the above standard errors.

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:

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

In this formula f is the appropriate factor from table A-5 and σ is the standard error for total or White persons in table A-3 or the standard error for Black and other races in table A-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 tables of standard errors. Table C of this report shows that in October 1975 there were 2,322,000 high school seniors who reported on their college plans and lived in metropolitan areas. The factor in table A-5 for Type of Residence, total or White, is 1.44. Thus, formula (1) and table A-1 show the standard error of an estimate of this size to be approximately 98,000=(68,300x1.44). The chances are 68 out of 100 that the estimate would have been a figure differing from a complete census figure by less than 98,000. The chances are 95 out of 100 that the estimate would have been a figure differing from a complete census figure by less than 196,000 (twice the standard error).

Tables 1 and C also show that of the 2,322,000 high school seniors mentioned above, 1,202,000 or 51.8 percent had definite plans to attend college. The factor in table A-5 for Type of Residence, total or White, is, again, 1.44. Interpolation in table A-3 shows the standard error of 51.8 percent to be 1.5 percent. Thus, the standard error of this estimate is approximately 2.2 = (1.44x1.5). Consequently, the chances are 68 out of 100 that the estimated 51.8 percent will be within 2.2 percentage points of a complete census figure. Chances are 95 out of 100 that the estimate would be within 4.4 percentage points of a complete census figure, i.e., the 95 percent confidence interval would be from 47.4 to 56.2.

Table A-3. Standard Errors for Estimated Percentages of Persons: Total, White, or Spanish Population

(68 chances out of 100)

Base of percentage (thousands)	Estimated percentage							
	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	35 or 65	50
75.....	1.7	2.3	3.6	5.0	5.9	7.2	7.9	8.3
100.....	1.4	2.0	3.1	4.3	5.1	6.2	6.9	7.2
250.....	0.9	1.3	2.0	2.7	3.2	3.9	4.3	4.5
500.....	0.6	0.9	1.4	1.9	2.3	2.8	3.1	3.2
1,000.....	0.5	0.6	1.0	1.4	1.6	2.0	2.2	2.3
2,000.....	0.3	0.4	0.7	1.0	1.1	1.4	1.5	1.6
3,000.....	0.3	0.4	0.6	0.8	0.9	1.1	1.3	1.3
4,000.....	0.2	0.3	0.5	0.7	0.8	1.0	1.1	1.1

Note: For a particular characteristic, see table A-5 for the appropriate factor to apply to the above standard errors.

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 correlation between the two characteristics, the formula will overestimate the true standard error.

Illustration of the computation of the standard error of a difference between estimated percentages. Tables 1 and C show that of the 984,000 high school seniors reporting on college plans and living in nonmetropolitan areas, 416,000 or 42.3 percent planned to attend college. The apparent difference between 42.3 percent for nonmetropolitan high school seniors and 51.8 percent for metropolitan high school seniors is 9.5 percent. The standard error, σ_x , of the 51.8 percent is 2.2, as shown above. Table A-5 shows the factor for Type of Residence to be 1.44. Table A-3 shows the standard error of 42.3 percent on a base of 984,000 to be 2.3. Thus, the standard error, σ_y , of the estimate is 3.3 (1.44x2.3).

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 9.5 percent is about

$$4.0 = \sqrt{(2.2)^2 + (3.3)^2}$$

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 4.0 percent. The 68 percent confidence interval about the 9.5 percent difference is from 5.5 to 13.5, i.e., 9.5 ± 4.0 . 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. The 95 percent confidence interval is 1.5 to 17.5. Thus, we can conclude with 95 percent confidence that there is a significant difference between the percentage for metropolitan and nonmetropolitan high school seniors regarding their plans to attend college.

Table A-4. Standard Errors for Estimated Percentages of Persons: Black and Other Races

(68 chances out of 100)

Base of percentage (thousands)	Estimated percentage							
	1 or 99	2 or 98	5 or 95	10 or 90	15 or 85	25 or 75	35 or 65	50
75.....	1.9	2.7	4.2	5.8	6.9	8.4	9.2	9.6
100.....	1.7	2.3	3.6	5.0	6.0	7.2	8.0	8.4
125.....	1.5	2.1	3.3	4.5	5.3	6.5	7.1	7.5
150.....	1.4	1.9	3.0	4.1	4.9	5.9	6.5	6.8
175.....	1.3	1.8	2.8	3.8	4.5	5.5	6.0	6.3
200.....	1.2	1.7	2.6	3.5	4.2	5.1	5.6	5.9
250.....	1.1	1.5	2.3	3.2	3.8	4.6	5.0	5.3
300.....	1.0	1.4	2.1	2.9	3.4	4.2	4.6	4.8
400.....	0.8	1.2	1.8	2.5	3.0	3.6	4.0	4.2
500.....	0.7	1.0	1.6	2.2	2.7	3.2	3.6	3.7

Note: For a particular characteristic, see table A-5 for the appropriate factor to apply to the above standard errors.

Table A-5. "f" Factors to be Applied to Tables A-1 through A-4 to Approximate Standard Errors

Type of characteristic	Values of f for--	
	Total, White or Spanish (Table A-1 or A-3)	Black and Other (Table A-2 or A-4)
Region or type of residence (Tables C and 1).....	1.44	(X)
Family income of students (Tables D and 2).....	1.22	(X)
Education of Spanish (Tables B and 1).....	1.37	(X)
Education (for all other CPS numbers in these tables).....	1.00	1.00

X Not applicable.