Appendix C. Source and Reliability of the Estimates

SOURCE OF DATA

The estimates in this report are based on data collected in the Current Population Survey (CPS), and the Nationwide Personal Transportation Survey.

The source of data in a table other than the Current Population Survey is identified at the bottom of that table. Brief description of the sources from and the procedures by which data of the Bureau of the Census were obtained are presented below.

Current Population Survey (CPS). The CPS estimates in this report are based on data obtained in the October 1978 survey. 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 one-way distance traveled to school, time spent traveling to school, and mode of transportation for each member in each sample household.

The present CPS sample was initially selected from the 1970 census file and is updated continuously to reflect new construction where possible. (See the 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.

Nationwide Personal Transportation Survey. Data for the Nationwide Personal Transporation Survey were collected in 1969-70 by the Bureau of the Census for the Federal Highway Administration of the Department of Transportation. This survey was designed to obtain up-to-date information on national patterns of travel and the data was collected in the same manner as the 1978 CPS data. The survey was based on a multi-stage probability sample of housing units located in 235 sample areas, comprising 485 counties and independent cities, representing every State and the District of Columbia. The 235 sample areas were selected by grouping all the national counties and independent cities into about 1,900 primary sample units (PSU's) and further forming 235 strata containing one or more PSU's that were relatively homogeneous according to socio-economic characteristics. Within each of the strata, a single PSU was selected to represent the stratum. Within each PSU, a probability sample of housing units was selected to represent the civilian noninstitutional population.

The households in the Nationwide Transportation Survey comprised two outgoing panels in the Quarterly Housing Survey (QHS) conducted by the Bureau of the Census. One

Description of the Current Population Survey

Time Period	Number of sample areas[1]	Households eligible: Interviewed	Households eligible: Not interviewed	Housing units visited, not eligible[2]
October 1978[4]	614	54,000	2,500	10,000
October 1977[3]	614	53,500	2,500	9,500
October 1972	461	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]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.

[4]A coverage improvement sample was incorporated beginning in October 1978 in order to provide better representation of mobile homes and new construction housing units.

panel was interviewed in April, July, and October 1969 and January 1970; the second panel was interviewed only once in August 1969.

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.

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 test 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 whenthe 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 both higher 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 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 C-1 and C-2 are approximations to generalized standard errors of estimated numbers and estimated percentages. Estimated standard errors for specific characteristics cannot be obtained from tables C-1 and C-2 without the use of the factors in table C-3. 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 C-1 and C-2 may be approximated by interpolation.

Two parameters— 'a' and 'b'—that are used to calculate standard errors for each type of characteristics are presented in table C-4. These parameters were used to calculate the standard errors in table C-1 and C-2 and to calculate factors in table C-3. They may also be used to directly calculate the standard errors for estimated numbers and percentages. Methods for direct computation are given in the following sections.

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

$$\sigma_x = f\sigma \tag{1}$$

In this formula f is the appropriate factor from table C-3 and σ is the standard error of the estimate obtained from table C-1. Alternatively, standard errors may be approximated by the following formula (2), use of which will provide more accurate results than the use of formula (1) above:

$$\sigma_x = \sqrt{a x^2 + b x} \tag{2}$$

Here x is the size of the estimate and a and b are the parameters in table C-4 associated with the particular type of characteristic.

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 percantage, p, can be obtained by use of the formula:

$$\sigma_{(x,p)} = f\sigma \tag{3}$$

In this formula f is the appropriate factor from table C-3 and σ is the generalized standard error for the percentage in table

Table C-1. Generalized Standard Errors of Estimated Numbers

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
25	10	2,500	103
50	15	5,000	144
75	18	7,500	175
100	21	10,000	201
250	33	25,000	306
500	46	50,000	403
750	56	75,000	454
1,000	65	100,000	475

Table C-2. Generalized Standard Errors of Estimated Percentages

Base of estimated percentage (in thousands)	Estimated percentage					
	1 or 99	2 or 98	5 or 95	10 or 90	25 or 75	50
75	2.4	3.3	5.2	7.1	10.3	11.9
100	2.1	2.9	4.5	6.2	8.9	10.3
250	1.3	1.8	2.8	3.9	5.6	6.5
500	0.9	1.3	2.0	2.8	4.0	4.6
1,000	0.6	0.9	1.4	2.0	2.8	3.3
2,500	0.4	0.6	0.9	1.2	1.8	2.1
5,000	0.3	0.4	0.6	0.9	1.3	1.5
10,000	0.2	0.3	0.4	0.6	0.9	1.0
25,000	0.13	0.18	0.3	0.4	0.6	0.7
50,000	0.09	0.13	0.2	0.3	0.4	0.5
100,000	0.06	0.09	0.14	0.2	0.3	0.3

C-2. When the numerator and denominator of the percentage are in different categories, use the factor indicated by the numerator. Alternately, the standard errors may be approximated by the following formula (4) from which the standard errors in table C-2 were calculated; direct computation will give more accurate results than use of the generalized standard error table and the factors.

$$\sigma_{(x,p)} = \sqrt{\frac{b}{x}\,(p)\,(100-p)} \qquad (4)$$

Here x is the base of the percentage, p is the percentage ($0 < p < 100$) and b is the parameter in table C-4 associated with the particular type of characteristic in the numerator of the percentage.

Illustration of use of standard error tables. Table 4 of this report shows that 171,000 Blacks living in metropolitan areas, enrolled in high school, were either driven or drove to their schools. Interpolation in table C-1 shows the standard error (σ_x) of an estimate of this size to be approximately 26,300. The factor in table C-3 for Blacks is 1.32; thus using formula (1) the standard error is approximately $1.32 \times 26300 \doteq 35{,}000$[1]. The 68 percent confidence interval as shown by the data is from 136,000 to 206,000. 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 101,000 to 241,000, i.e., 171,000 ± (2 x 35,000).

[1] Formula (2) gives a standard error of 35,000.

Table C-3. 'f' Factors to be Applied to Generalized Standard Errors in Tables C-1 and C-2

Type of characteristics	'f' factors
MODE, TIME AND DISTANCE OF TRANSPORTATION TO SCHOOL	
Total, metropolitan-non-metropolitan:	
Total or White..................	1.00
Black and other races...........	1.32
Spanish origin..................	1.33
Students living at home:	
Total or White..................	1.00
Black and other races...........	1.32
Spanish origin..................	1.33
School enrollment[1]:	
Total or White..................	0.70
Black and other races...........	0.81
NATIONWIDE PERSONAL TRAVEL SURVEY MODE, TIME OR DISTANCE TRAVELED TO SCHOOL:	
All races.......................	3.30

[1] For school enrollment cross-tabulated by metropolitan-nonmetropolitan residence, multiply the above factor by 1.41.

Table C-4. "a" and "b" Parameters for Estimated Numbers and Percentages of Persons

Type of characteristic	Parameters	
	a	b
MODE, TIME AND DISTANCE OF TRANSPORTATION TO SCHOOL		
Total, metropolitan-non-metropolitan:		
Total or White.............	-0.000020	4253
Black and other races......	-0.000308	7402
Spanish origin.............	-0.000043	7469
Students living at home:		
Total or White.............	-0.000020	4253
Black and other races......	-0.000308	7402
Spanish origin.............	-0.000043	7469
School enrollment[1]:		
Total or White.............	-0.000016	2064
Black and other races......	-0.000186	2792
NATIONWIDE PERSONAL TRAVEL SURVEY		
Mode, time or distance traveled to school:		
All races..................	-0.000543	46195

[1] For school enrollment cross-tabulated by metropolitan-nonmetropolitan residence, multiply the above parameters by 2.0.

Table 4 also shows that out of 1,660,000 Blacks, residing in metropolitan areas and enrolled in high school, 171,000 or 10.3 percent were either driven or drove to their schools. interpolation in table C-2 shows the standard error of 10.3 percent to be 2.2 percent. Consequently, the 68 percent confidence interval is from 8.1 to 12.5 percent and the 95 percent confidence interval is from 5.9 to 14.7 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

$$\sigma_{(x\text{-}y)} = \sqrt{\sigma_x^{\,2} + \sigma_y^{\,2}} \qquad (5)$$

where σ_x and σ_y are the standard errors of the estimate x and y; 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 the computation of the standard error of a difference between percentages. Table 4 of this report shows that in metropolitan areas, 10.3 percent of all Black high school students (1,660,000) living in their homes traveled to their schools by cars whereas 30.1 percent of these Black students walked to their schools. Thus, the apparent difference in percents between these two groups of Black students is 19.8 percent. Using formula (5), the standard error of the estimated difference of 19.8 percent is about

$$\sqrt{(2.2)^2 + (3.2)^2} = 3.9 \text{ percent}$$

This means that the 90-percent confidence interval around the difference is from 13.6 to 26.0 percent and the 95-percent confidence interval is from 12.0 to 27.6 percent. Thus, we can conclude with 95-percent confidence that in 1978 there was a significant difference between the percentage of Black students who walked to school and the percentage of those Black students who were either driven or drove to school by themselves.

Standard error of a median. The sampling variability of an estimated median depends upon the form of the distribution as well as 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 confidence limits of a median based on sample data:

1. Determine, using the standard error table and an appropriate factor or formula (4), the standard error of an estimate of 50 percent from the distribution.

2. Add to and subtract from 50 percent the standard error determined in step 1.

3. Using the distribution of the characteristic, calculate the 68-percent confidence interval by finding the values 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.

Note: When combining two or more distributions, the medians of the distribution must be computed by the user. The median is the estimate for the person at the center of the distribution and may be approximated by linear interpolation within the group which containes this person.

Illustration of the computation of a confidence interval for a median. Table 1 of the report shows that the median time spent by White college students 14 to 34 years old is 22.7 minutes. Table 1 also indicates the base of the distribution from which this median was determined is 5,672,000.

1. Interpolation in table C-2 shows the estimated standard error of 50 percent on a base of 5,672,000 is about 1.4 percent.

2. To obtain a 95-percent confidence interval on an estimate of a median, add to and subtract from 50 percent twice the standard error found in step 1. This yields percent limits of 47.2 and 52.8 percent.

3. From table 1 of this report, 31.0 percent of White college students spent less than 15 minutes and 37.1 percent spent 15-29 minutes traveling to college. By linear interpolation the lower limit (of the 95-percent confidence interval) on the estimate is found to be about

$$15 + (30 - 15) \left(\frac{47.2 - 31.0}{37.1}\right) \doteq 21.5$$

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

$$15 + (30 - 15) \left(\frac{52.8 - 31.0}{37.1}\right) \doteq 23.8$$

Thus, the 95-percent confidence interval on the estimated median is from 21.5 to 23.8 minutes.

☆U.S. GOVERNMENT PRINTING OFFICE: 1979— 281-042:191