Appendix B. Source and Reliability of the Estimates

SOURCE OF DATA

Most of the estimates in this report are based on data collected in November 1978 from the Current Population Survey of the Bureau of the Census. Some data were obtained from the November 1964, 1972, and 1976 CPS published reports. Also included in text table I are counts of official votes cast during the November elections of these years.

Current Population Survey (CPS). 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, supplemental questions are asked about voting and voter registration during the month of November in election years.

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). The monthly CPS sample is spread over 461 areas with coverage in each of the 50 States and the District of Columbia. A supplementary sample of housing units in 24 States and the District of Columbia was incorporated with the monthly CPS sample to produce the November 1978 voting and voter registration data. The expanded CPS sample is located in 614 areas comprising 1,113 counties, independent cities, and divisions in the nation. The 614 sample areas used in November 1978 include 461 areas from the monthly CPS and 153 supplementary areas.

Samples for previous sample designs were selected from files from the most recently completed census. 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.

RELIABILITY OF THE ESTIMATES

Since the CPS estimates in this report 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 questionnaires, 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 errors. 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

Description of the Current Population Survey

Time period	Number of sample areas[1]	Households eligible		Housing units visited not eligible[2]
		Interviewed	Not Interviewed	
November 1978	461	45,000	2,000	8,000
Supplemental sample	153	8,500	500	1,500
November 1976	461	45,000	2,000	8,000
November 1972	461	45,000	2,000	8,000
November 1964	357	33,500	1,500	6,000

[1]These areas were chosen to provide coverage in each State and the District of Columbia with the exception of the supplemental sample used for November 1978 data.

[2]These are housing units which were visited but were found to be vacant or otherwise not eligible for interview.

a relatively small number of cases or on small differences between estimates.

Nonsampling 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, 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 units with the sample (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 bias 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 similar undercoverage differentials by age, sex, and race as in CPS.

The approximate magnitude of the two sources of undercoverage of housing units is known. Of the 83,000,000 housing units in the U.S., about 600,000 new construction housing units other than mobile homes are not represented in the CPS sample because they were assigned building permits prior to January 1970, but building was not completed by the time of the census (i.e., April 1970). Almost all conventional new construction, for which building permits were issued after 1969, 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. 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. 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 entire population was surveyed. The sample estimate and its estimated error enable one to construct confidence intervals, ranges that would 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 essentially the same general conditions and using the same sample design, 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 estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the average estimate derived from all possible samples is included in the confidence 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.

Comparability with other data. Caution should be used in comparing metropolitan and nonmetropolitan estimates from 1978, when the expanded sample was used, to those from 1976 and earlier years. Some relatively large differences in estimates of population in metropolitan and nonmetropolitan areas have been observed between the 461 and the 614 area samples. Thus these differences do not provide reliable measures of actual changes in this population.

Note when using small estimates. Summary measures (such as averages and percent distributions) are shown in the report only when the base of the measure 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 corresponding percentages. These smaller estimates are provided primarily to permit such combinations of the categories as serve each user's need.

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.

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

(68 chances out of 100. Numbers in thousands)

Estimate	Standard error	Estimate	Standard error
25.....	8	5,000.....	110
50.....	11	7,500.....	133
75.....	14	10,000.....	152
100.....	16	25,000.....	223
250.....	25	50,000.....	271
500.....	35	75,000.....	266
750.....	43	100,000....	204
1,000.....	50	110,000....	151
2,500.....	79		

NOTE: For a particular characteristic see table B-5 or B-6 for the appropriate factor to apply to the above standard errors.

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

(68 chances out of 100. Numbers in thousands)

Estimate	Standard error	Estimate	Standard error
25.....	10	750.....	51
50.....	14	1,000.....	58
75.....	17	2,500.....	86
100.....	19	5,000.....	106
250.....	30	7,500.....	107
500.....	42	10,000.....	89

NOTE: For a particular characteristic see table B-5 for the appropriate factor to apply to the above standard errors.

The figures presented in tables B-1 and B-3 provide approximations to generalized standard errors of estimated numbers and percentages for total or White persons; tables B-2 and B-4 provide approximations to generalized standard errors of estimated numbers and percentages for Black persons. Standard errors for intermediate values not shown in the generalized tables of standard errors may be approximated by linear interpolation. Estimated standard errors for specific characteristics cannot be obtained from tables B-1 through B-4 without the use of factors in tables B-5 or B-6. 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. The standard error tables with which each factor should be used are indicated in tables B-5 and B-6.

Two parameters (denoted "a" and "b") are used to calculate standard errors for each type of characteristic; they are presented in tables B-5 and B-6. These parameters were used to calculate the standard errors in tables B-1 through B-4, and to calculate the factors in tables B-5 and B-6. They also may 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 in two ways. It may be obtained by use of the formula

$$\sigma_x = f\sigma \qquad (1)$$

where f is the appropriate factor from table B-5 or B-6, and σ is the standard error on the estimate obtained by interpolation from table B-1 or B-2. Alternatively, standard errors

Table B-3. Standard Errors of Estimated Percentages Total or White

(68 chances out of 100)

Base of estimated percentage (thousands)	Estimated percentage					
	2 or 98	5 or 95	10 or 90	20 or 80	25 or 75	50
50.....	3.1	4.9	6.7	9.0	9.7	11.2
100.....	2.2	3.5	4.8	6.3	6.9	7.9
250.....	1.4	2.2	3.0	4.0	4.3	5.0
500.....	1.0	1.5	2.1	2.8	3.1	3.5
1,000.....	0.7	1.1	1.5	2.0	2.2	2.5
5,000.....	0.3	0.5	0.7	0.9	1.0	1.1
10,000.....	0.2	0.3	0.5	0.6	0.7	0.8
25,000.....	0.14	0.2	0.3	0.4	0.4	0.5
50,000.....	0.10	0.15	0.2	0.3	0.3	0.4
100,000.....	0.07	0.11	0.15	0.2	0.2	0.3
150,000.....	0.06	0.09	0.12	0.2	0.2	0.2

NOTE: For a particular characteristic see table B-5 or B-6 for the appropriate factor to apply to the above standard errors.

Table B-4. Standard Errors of Estimated Percentages Black and Other Races

(68 chances out of 100)

Base of estimated percentage (thousands)	Estimated percentage					
	2 or 98	5 or 95	10 or 90	20 or 80	25 or 75	50
50	3.8	5.9	8.1	10.9	11.8	13.6
100	2.7	4.2	5.8	7.7	8.3	9.6
250	1.7	2.6	3.6	4.9	5.3	6.1
500	1.2	1.9	2.6	3.4	3.7	4.3
750	1.0	1.5	2.1	2.8	3.0	3.5
1,000	0.8	1.3	1.8	2.4	2.6	3.0
2,500	0.5	0.8	1.2	1.5	1.7	1.9
5,000	0.4	0.6	0.8	1.1	1.2	1.4
10,000	0.3	0.4	0.6	0.8	0.8	1.0
25,000	0.2	0.3	0.4	0.5	0.5	0.6

NOTE: For a particular characteristic see table B-5 for the appropriate factor to apply to the above standard errors.

may be approximated by formula (2) from which the standard errors were calculated in tables B-1 and B-2. Use of this formula will provide more accurate results than the use of formula (1) above.

$$\sigma_x = \sqrt{ax^2 + bx} \qquad (2)$$

Here x is the size of the estimate and a and b are the parameters in tables B-5 or B-6 associated with the particular type of characteristic.

Standard errors of estimated percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the 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. When the numerator and denominator of the percentage are in different categories, use the factor or parameters from table B-5 or B-6 indicated by the numerator. The approximate standard error, $\sigma_{(x,p)}$, of an estimated percentage can be obtained by use of the formula.

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

In this formula f is the appropriate factor from table B-5 or B-6, and σ is the standard error on the estimate from table B-2 or B-4. Alternatively, standard errors may be approximated by formula (4), from which standard errors in tables B-2 and B-4 were calculated; direct computation will give more accurate results than use of the standard error tables and the factors.

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

Here x is the size of the subclass of persons, families or households, which is the base of the percentage, p is the percentage ($0 \leq p \leq 100$), and b is the parameter in table B-5 or B-6 associated with the particular type of characteristic in the numerator of the percentage.

Illustration of the use of standard error tables. Table 8 of this report shows that 13,851,000 never-married persons reported registering for the November 1978 election. Table B-1 shows the standard error on an estimate of this size to be approximately 170,000. Applying the appropriate factor from table B-5 and using formula (1), the approximate standard error is 1.0 X 170,000 = 170,000.[1] The 68 percent confidence interval as shown by the data is from 13,681,000 to 14,021,000. 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 that the average estimate derived from all possible samples lies within the interval from 13,511,000 to 14,191,000 (using twice the standard error) with 95 percent confidence.

Of these 13,851,000 never married persons who reported being registered, 6,702,000 or 48.4 percent were female. From table B-5 the appropriate b parameter for computing standard errors is 2518. Using formula (4), the standard error on an estimate of 48.4 percent is

$$\sqrt{\frac{2518}{13{,}851{,}000}\,(48.4)\,(51.6)} \doteq 0.7 \text{ percent}^2$$

Consequently, the 68 percent confidence interval as shown by these data is from 47.7 to 49.1 percent, and the 95 percent confidence interval is 47.0 to 49.8 percent.

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

$$\sigma_{(x-y)} = \sqrt{\frac{\sigma^2}{x} + \frac{\sigma^2}{y}} \qquad (5)$$

[1] Formula (2) for this example gives a standard error of 176,000.
[2] Table B-3 and formula (3) also give a standard error of 0.7 percent.

where σ_x and σ_y are the standard errors of the estimates x and y; the estimates can be of numbers, percents, ratios, etc. This will represent the actual standard errors 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 two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration of the computation of the standard error of a difference. Table 6 of this report shows that in the November 1978 election 33.5 percent of the Black voting age population in the South (8,279,000) voted as compared to 44.6 percent of the Black voting age population in the North Central Region (3,044,000). The estimated difference between the two regions is 11.1 percent.

Using formula (3), table B-4 and the factor 1.3 for regions from table B-5, the standard error on the 33.5 percent of Black voters in the South is 1.3 percent. The standard error on the 44.6 percent of Black voters in the North Central Region is 2.3 percent. Therefore, the standard error of the estimated difference of 11.1 percent is about

$$2.6 \doteq \sqrt{(1.3)^2 + (2.3)^2}$$

This means that the 68 percent confidence interval for the difference is from 8.5 to 13.7 percent. 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. The corresponding 95 percent confidence interval is 5.9 to 16.3 percent. Thus, we can conclude with 95 percent confidence that there was greater Black voter participation in the North Central Region than in the South.

Table B-5. Factors to be Applied to Generalized Standard Errors in Tables B-1 through B-4 and "a" and "b" Parameters for Various Characteristics

(For meaning of symbols, see text)

Characteristics	Total or White			Black and other races			Spanish origin		
	a	b	f[1]	a	b	f[2]	a	b	f[1]
Voting, registration, reasons for not voting or registering									
Total or U.S.:									
CPS[3]	-0.000021	2518	1.0	-0.000289	3686	1.0	-0.000043	7469	1.7
Official	-	-	-	-	-	-	-	-	-
Citizenship, household relationship, family heads by presence of own children, marital status	-0.000021	2518	1.0	-0.000289	3686	1.0	-0.000043	7469	1.7
Educational level, employment status, family income of persons and occupation group	-0.000021	2518	1.0	-0.000021	2518	0.8	-0.000025	3851	1.2
Residence:									
Region, metro/nonmetro, farm/nonfarm, duration of residence	-0.000052	6242	1.6	-0.000052	6242	1.3	-0.000075	15028	2.4
Marital status	-0.000017	3500	1.2	-0.000210	5020	1.2	-0.000043	7469	1.7
Education of persons	-0.000016	2064	0.9	-0.000186	2792	0.9	-0.000025	3851	1.2
Education of family head	-0.000010	1389	0.7	-0.000087	1255	0.6	-0.000033	2397	1.0
Employment, not in labor force, occupation	-0.000016	2078	0.9	-0.000133	2078	0.8	-0.000030	4394	1.3
Unemployment	-0.000015	1971	0.9	-0.000139	2265	0.8	-0.000013	1864	0.9
Persons by family income	-0.000020	3770	1.2	-0.000178	4310	1.1	-0.000067	10112	2.0
Household relationship:									
Head, wife of head	-0.000010	1389	0.7	-0.000087	1255	0.6	-0.000033	2397	1.0
Nonrelative or other relative of head	-0.000017	3500	1.2	-0.000210	5020	1.2	-0.000043	7469	1.7

[1]Factors in this column should be applied to tables B-1 and B-3.
[2]Factors in this column should be applied to tables B-2 and B-4.
[3]For 1964 data, multiply parameters by 1.5 and factors by 1.22.

Table B-6. Factors and Parameters to Obtain Standard Errors for Voting and Registration Characteristics for Individual SMSA's, States, and Census Divisions

Type of residence	Parameters		Factor[1]
	a	b	
Individual SMSA	-0.000052	6242	1.57
States:			
Alabama	-0.000023	2770	1.05
California	-0.000022	2619	1.02
Connecticut	-0.000021	2568	1.01
Florida	-0.000023	2719	1.04
Georgia	-0.000023	2770	1.05
Illinois	-0.000023	2719	1.04
Indiana	-0.000022	2669	1.03
Iowa	-0.000019	2266	0.95
Kentucky	-0.000023	2770	1.05
Louisiana	-0.000022	2669	1.03
Maryland	-0.000026	3097	1.11
Massachusetts	-0.000021	2568	1.01
Michigan	-0.000023	2719	1.04
Minnesota	-0.000022	2669	1.03
Missouri	-0.000023	2719	1.04
New Jersey	-0.000022	2619	1.02
New York	-0.000022	2669	1.03
North Carolina	-0.000025	2946	1.08
Ohio	-0.000022	2669	1.03
Pennsylvania	-0.000022	2669	1.03
Tennessee	-0.000023	2770	1.05
Texas	-0.000024	2820	1.06
Virginia	-0.000026	3148	1.12
Washington	-0.000024	2820	1.06
Wisconsin	-0.000024	2871	1.07
Census divisions:			
New England	-0.000017	2065	0.91
Middle Atlantic	-0.000022	2669	1.03
East North Central	-0.000021	2568	1.01
West North Central	-0.000018	2165	0.93
South Atlantic	-0.000022	2669	1.03
East South Central	-0.000022	2619	1.02
West South Central	-0.000022	2669	1.03
Mountain	-0.000014	1737	0.83
Pacific	-0.000021	2568	1.01

[1]Apply these factors to the generalized standard error tables B-1 and B-3 only to obtain standard errors for individual SMSA's, States and Census Divisions or use the parameters and formulas provided above.