Appendix B. Source and Reliability of Estimates

SOURCE OF DATA

Most of the estimates in this report are based on data collected in November 1982 from the Current Population Survey of the Bureau of the Census. Some data were obtained from published November reports from earlier years dating back to 1964. These reports are noted at the bottom of the text tables. Also included in text table G are counts of official votes cast during the November elections of the election 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 files with coverage in all 50 States and the District of Columbia. The sample is continually updated to reflect new construction where possible. The monthly CPS sample is located in 629 areas comprising 1,148 counties, independent cities, and minor civil divisions in the Nation. In this sample, approximately 61,500 occupied housing units were eligible for interview. Of this number, about 2,500 occupied units were visited but interviews were not obtained because the occupants were not found at home after repeated calls or were unavailable for some other reason.

Samples for previous 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.

Description of the Current Population Survey

Time period	Number of sample areas[1]	Housing units eligible	
		Interviewed	Not interviewed
November 1982	629	59,000	2,500
November 1980	629	64,000	2,500
November 1972-76	461	45,000	2,000
November 1968-70	449	48,000	2,000
November 1964-66	357	33,500	1,500

[1]These sample areas were chosen to provide coverage in each State and the District of Columbia.

The estimation procedure used in this survey involves the inflation of the weighted sample results to independent estimates of the total civilian noninstitutional population of the United States by age, race, and sex. These independent estimates are based on statistics from decennial censuses; statistics on the strength of the Armed Forces. The independent population estimates used in this report to obtain data for November 1982 are based on the 1980 decennial census. Data for 1970 to 1980 were obtained using independent population estimates based on the 1970 decennial census; likewise, data for 1964 to 1968 were obtained using independent population estimates based on the 1960 decennial census.

RELIABILITY OF ESTIMATES

Since the CPS estimates 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 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 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 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 1980 decennial census, is about 7 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 from interviewed persons in the same age-sex-race group. Further, the independent population controls used have not been adjusted for undercoverage in the 1980 census.

For additional information on nonsampling error including the possible impact on CPS data when known, refer to Statistical Policy Working Paper 3, *An Error Profile: Employment as Measured by the Current Population Survey,* Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1978, and Technical Paper 40, *The Current Population Survey: Design and Methodology,* Bureau of the Census, U.S. Department of Commerce.

Sampling variability. The standard errors given in the following tables are primarily measures of sampling variability, that is, of the variation that occurred by chance because a sample rather than the entire population was surveyed. The sample estimate and its estimated standard 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 if an estimate and its estimated standard error were calculated from each sample, then:

1. 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.

2. 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.

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common types of hypotheses appearing in this report are 1) the population parameters are identical or 2) they are different. An example of this would be comparing the voter participation rate of men versus that of women. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the parameters are different when, in fact, they are identical. All statements of comparison in the text have passed a hypothesis test at the 0.10 level of significance or better, and most have passed a hypothesis test at the 0.05 level of significance or better. This means that, for most differences cited in the text, the estimated difference between parameters is greater than twice the standard error of the difference. For the other differences mentioned, the estimated difference between parameters is between 1.6 and 2.0 times the standard error of the difference. When this is the case, the statement of comparison is qualified in some way, e.g., by the use of the phrase "some evidence."

Comparability of data. In using metropolitan and nonmetropolitan data, caution should be used in comparing estimates for 1978 to other years. Methodological and sample design changes occurred in 1978 resulting in relatively large differences in the metropolitan and nonmetropolitan area estimates. However, estimates for 1980 and later are comparable as are estimates for 1976 and earlier.

Caution should also be used when comparing estimates for 1982, which reflect 1980 census-based population controls, to those for 1972 through 1980, which reflect 1970 census-based population controls. This change in population controls had relatively little impact on summary measures such as means, medians, and percent distributions, but did have a significant impact on levels. For example, use of the 1980-based population controls resulted in about a 2-percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for 1982 will differ from those for earlier years more than what could be attributed to actual changes in the population and these differences could be disproportionately greater for certain subpopulation groups than for the total population.

Note when using small estimates. Summary measures (such as percent distributions) are shown in this 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 corresponding percentages. These smaller estimates are provided primarily to permit such combinations of the categories as serve each data 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 pro-

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

(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 for the appropriate factor to apply to the above standard errors.

vided 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 B-1 and B-3 provide approximations to standard errors of estimated numbers and estimated percentages for total or White persons, and persons of Spanish origin; tables B-2 and B-4 provide approximations to standard errors of estimated numbers and estimated 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 table B-5. 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 table B-5.

Two parameters (denoted "a" and "b") are used to calculate standard errors for each type of characteristic; they are presented in table B-5. These parameters were used to calculate the standard errors in tables B-1 through B-4 and to calculate the factors in table B-5. They also may be used to directly calculate the standard errors for estimated numbers and estimated 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 and σ is the standard error on the estimate obtained by interpolation from table B-1 or B-2. Alternatively, standard errors may be approximated by formula (2) below, 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 table B-5 associated with the particular type of characteristic. When calculating standard errors for numbers from cross-tabulations involving different characteristics, use the factor or set of parameters for the characteristic which will give the largest standard error.

Illustration of the computation of the standard error of an estimated number. Table C of this report shows that 12,227,000 persons 18 to 24 years old reported that they registered to vote in the November 1982 election. Using formula (2) with a =-0.000021 and b= 2,518 from table B-5, the approximate standard error is

$$(-0.000021)(12{,}227{,}000)^2 + (2{,}518)(12{,}227{,}000) \doteq 166{,}000\ ^1$$

[1] Using formula (1), table B-1, and the appropriate factor from table B-5, the approximate standard error is 1.0 x 163,000 = 163,000.

Table B-2. Standard Errors of Estimated Numbers: Black

(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.

This means that the 68-percent confidence interval for the number of 18-24-year-old persons that reported they registered to vote in the November 1982 election is from 12,061,000 to 12,393,000 and the 95-percent confidence interval is from 11,895,000 to 12,559,000.

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. When the numerator and denominator of the percentage are in different categories, use the factor or parameters 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 and σ is the standard error on the estimate from table B-3 or B-4.

Alternatively, the standard error may be approximated by formula (4) below, from which the standard errors in table B-3 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 or families which is the base of the percentage, p is the percentage ($0<p<100$), and b is the parameter in table B-5 associated with the particular type of characteristics in the numerator of the percentage.

Illustration of the computation of the standard error of a percentage. Table C also shows that of the 12,227,000 18-to-24-year-olds that reported they registered to vote in November 1982, 7,139,000 or 58.4 percent voted. Using formula (4) and the appropriate b-parameter from table B-5, 2,518, the standard error on an estimate of 58.4 percent is

$$\sqrt{\frac{2518}{12{,}227{,}000}\,(58.4)(100-58.4)} = 0.7 \text{ percent}^2$$

[2] Using formula (4), table B-3 and the appropriate factor from table B-5, the approximate standard error is 1.0 x 0.7 = 0.7.

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

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 for the appropriate factor to apply to the above standard errors.

Table B-4. Standard Errors of Estimated Percentages: Black

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 apporpriate factor to apply to the above standard errors.

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: 1972 to Present

Characteristics	Total or White			Black			Spanish origin		
	a	b	f[1]	a	b	f[2]	a	b	f[1]
Voting, registration, reasons for not voting or registering:									
CPS counts	-0.000021	2518	1.0	-0.000289	3686	1.0	-0.000043	7469	1.7
Official counts	-	-	-	(X)	(X)	(X)	(X)	(X)	(X)
Citizenship, household relationship, family heads, by presence of own children, marital status, duration of residence, tenure	-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	(X)	3374	1.2
Characteristics of all persons:									
Marital status	-0.000017	3500	1.2	-0.000210	5020	1.2	-0.000043	7469	1.7
Education of persons	-0.000025	2014	0.9	-0.000179	2265	0.8	(X)	3374	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	2327	1.0	-0.000144	2327	0.8	-0.000810	1847	([3])
Unemployment	-0.000015	2206	0.9	-0.000150	2536	0.8	0.001490	1600	([3])
Persons by family income	-0.000018	3770	1.2	-0.000154	4310	1.1	-0.000067	10112	2.0
Duration of residence, tenure	-0.000017	3500	1.2	-0.000210	5020	1.2	-0.000043	7469	1.7
Household relations:									
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

X Not applicable.

[1]Factors in this column are to be applied to tables B-1 and B-3.
[2]Factors in this column are to be applied to tables B-2 and B-4.
[3]To obtain standard errors for this characteristic, use formula (2).

Note: For 1964 data, multiply parameters by 1.5 and factors by 1.22.

This means that the 68-percent confidence interval for the percentage of 18-to-24-year-olds registered to vote that actually voted is from 57.7 to 59.1, and the 95-percent confidence interval is from 57.0 to 59.8 percent.

Standard error of a difference. The formula for the approximate standard error of the difference between two estimates, x and y, is given by

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

where σ_x and σ_y are the standard errors of the estimates x and y, respectively; 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 two 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. As stated earlier, table C shows that of the 12,227,000 18-to-24-year-olds that reported they registered to vote, 58.4 percent voted. Table C also shows that 73.8 percent of the 41,103,000 registered 25-to-44-year-olds actually voted. Thus, the apparent difference between the percent 18-to-24-year-olds and 25-to-44-year-olds that voted is 15.4 percent. The standard error (σ_x) of 58.4 percent is 0.7 percent as shown above. Using formula (4) the standard error (σ_y) on 73.8 percent is approximately 0.3 percent. Therefore, using formula (5), the standard error of the estimated difference of 15.4 percent is about

$$\sqrt{(0.7)^2 + (0.3)^2} \doteq 0.8 \text{ percent}$$

This means that the 68-percent confidence interval for the difference is 14.6 to 16.2 percent, and the 95-percent confidence interval is 13.8 to 17.0 percent. Since the 95-percent confidence interval does not contain zero, we may conclude with 95-percent confidence that a higher proportion of registered 25-to-44-years-olds voted than did registered 18-to-24-years-olds.