Appendix B. Source and Accuracy of Estimates

SOURCE OF DATA

Most estimates in this report come from data obtained from the Current Population Survey (CPS) conducted in March of years 1960 through 1993. The Bureau of the Census conducts the survey every month, although this report uses mostly March data for its estimates. The March survey uses two sets of questions, the basic CPS and the supplements.

Basic CPS. The basic CPS collects primarily labor force data about the civilian noninstitutional population. Interviewers ask questions concerning labor force participation about each member 15 years old and over in every sample household.

The present CPS sample was selected from the 1980 Decennial Census files with coverage in all 50 states and the District of Columbia. The sample is continually updated to account for new residential construction. It's located in 729 areas and includes 1,973 counties, independent cities, and minor civil divisions. About 60,000 occupied housing units are eligible for interview every month. Interviewers are unable to obtain interviews at about 2,600 of these units because the occupants are not found at home after repeated calls or are unavailable for some other reason.

Since the introduction of the CPS, the Bureau of the Census has redesigned the CPS sample several times. These redesigns have improved the quality and reliability of the data and have satisfied changing data needs. The most recent changes were completely implemented in July 1985.

Table B-1 summarizes changes in the CPS designs for the years for which data appear in this report.

Table B-1. **Description of the Current Population Survey**

Time Period	Number of Sample Areas	Housing Units Eligible[1]	
		Interviewed	Not Interviewed
1990 to 1993	729	57,400	2,600
1989	729	53,600	2,500
1986 to 1988	729	57,000	2,500
1985	[2]629/729	57,000	2,500
1982 to 1984	629	59,000	2,500
1980 to 1981	629	65,500	3,000
1977 to 1979	614	55,000	3,000
1973 to 1976	461	46,500	2,500
1972	449	45,000	2,000
1967 to 1971	449	48,000	2,000
1963 to 1966	357	33,500	1,500
1960 to 1962	333	33,500	1,500

[1]Excludes about 2,500 Hispanic households added from the previous November sample. (See *"March Supplement"*.)
[2]The CPS was redesigned following the 1980 Decennial Census of Population and Housing. During phase-in of the new design, housing units from the new and old designs were in the sample.

March Supplement. In addition to the basic CPS questions, interviewers asked supplementary questions in March about place of residence one year ago.

To obtain more reliable data for the Hispanic population, the March CPS sample was increased by about 2,500 eligible housing units. These housing units were interviewed the previous November and contained at least one sample person of Hispanic origin. In addition, the sample included persons in the Armed Forces living off post or with their families on post.

Estimation Procedure. This survey's estimation procedure inflates weighted sample results to independent estimates of the civilian noninstitutional population of the United States by age, sex, race and Hispanic/non-Hispanic categories. The independent estimates were based on statistics from decennial censuses of population; statistics on births, deaths, immigration and emigration; and statistics on the size of the Armed Forces.

The independent population estimates used for 1981 to present were based on updates to controls established by the 1980 Decennial Census. Before 1981 independent population estimates from the most recent decennial census were used. For more details on the change in independent estimates, see the section entitled "Introduction of 1980 Census Population Controls" in an earlier report (Series P-60, No. 133). The estimation procedure for the March supplement included a further adjustment so husband and wife of a household received the same weight.

The estimates in this report for 1985 and later also employ a revised survey weighting procedure for persons of Hispanic origin. In previous years, weighted sample results were inflated to independent estimates of the noninstitutional population by age, sex, and race. There was no specific control of the survey estimates for the Hispanic population. Since then, the Bureau of the Census developed independent population controls for the Hispanic population by sex and detailed age groups. Revised weighting procedures incorporate these new controls. The independent population estimates include some, but not all, undocumented immigrants.

ACCURACY OF THE ESTIMATES

Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. A sample survey estimate has two possible types

of error: nonsampling and sampling. The accuracy of an estimate depends on both types of error, but the full extent of the nonsampling error is unknown. Consequently, one should be particularly careful when interpreting results based on a relatively small number of cases or on small differences between estimates. The standard errors for CPS estimates primarily indicate the magnitude of sampling error. They also partially measure the effect of some nonsampling errors in responses and enumeration, but do not measure systematic biases in the data. (Bias is the average over all possible samples of the differences between the sample estimates and the desired value.)

Nonsampling Variability. There are several sources of nonsampling errors including the following:

- Inability to get information about all sample cases.
- Definitional difficulties.
- Differences in interpretation of questions.
- Respondents' inability or unwillingness to provide correct information.
- Respondents' inability to recall information.
- Errors made in data collection, such as recording and coding data.
- Errors made in processing the data.
- Errors made in estimating values for missing data.
- Failure to represent all units with the sample (undercoverage).

CPS undercoverage results from missed housing units and missed persons within sample households. Compared to the level of the 1980 Decennial Census, overall CPS undercoverage is about 7 percent. CPS undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks and other races combined than for Whites. As described previously, ratio estimation to independent age-sex-raceHispanic population controls partially corrects for the bias due to 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 those of interviewed persons in the same age-sex-race-Hispanic group. Furthermore, the independent population controls have not been adjusted for undercoverage in the 1980 census.

A common measure of survey coverage is the coverage ratio, the estimated population before ratio adjustment divided by the independent population control. Table B-2 shows CPS coverage ratios for age-sex-race groups for a recent month. The CPS coverage ratios can exhibit some variability from month to month, but these are a typical set of coverage ratios. Other Census Bureau household surveys experience similar coverage.

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.

Comparability of Data. Data obtained from the CPS and other sources are not entirely comparable. This results from differences in interviewer training and experience and in differing survey processes. This is an example of nonsampling variability not reflected in the standard errors. Use caution when comparing results from different sources.

Caution should also be used when comparing estimates in this report (which reflect 1980 census-based population controls) with estimates for 1980 and earlier years (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. It did have a significant impact on levels. For example, use of 1980 based population controls results in about a 2-percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 1981 and later years will differ from those for earlier years by more than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain subpopulation groups than for the total population.

Since no independent population control totals for persons of Hispanic origin were used before 1985, compare Hispanic estimates over time cautiously.

Note When Using Small Estimates. Summary measures (such as medians and percentage distributions) are shown only when the base is 75,000 or greater. Because of the large standard errors involved, summary measures would probably not reveal useful information when computed on a smaller base. However, estimated numbers are shown even though the relative standard errors of these numbers are larger than those for corresponding percentages. These smaller estimates permit combinations of the categories to suit data users' needs. These estimates may not be reliable for the interpretation of small differences. For instance, even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Table B- 2. **CPS Coverage Ratios**

Age	non-Black		Black		All Persons		
	M	F	M	F	M	F	Total
0-14	0.948	0.960	0.913	0.930	0.943	0.955	0.949
15	0.953	0.986	0.975	1.025	0.956	0.993	0.974
16	0.877	0.997	0.886	0.963	0.879	0.991	0.934
17	0.958	0.956	0.860	0.932	0.942	0.952	0.947
18	0.950	0.958	0.931	0.692	0.947	0.916	0.931
19	0.882	0.953	0.773	0.740	0.866	0.920	0.893
20-24	0.889	0.918	0.645	0.820	0.856	0.904	0.881
25-26	0.867	0.964	0.687	0.820	0.844	0.943	0.894
27-29	0.919	0.941	0.700	0.834	0.892	0.926	0.909
30-34	0.884	0.947	0.667	0.865	0.859	0.936	0.898
35-39	0.892	0.936	0.693	0.928	0.871	0.935	0.903
40-44	0.895	0.933	0.781	0.889	0.884	0.928	0.906
45-49	0.933	0.955	0.842	0.938	0.925	0.953	0.939
50-54	0.953	0.958	0.845	0.869	0.942	0.948	0.945
55-59	0.918	0.905	0.797	0.906	0.906	0.905	0.905
60-62	0.926	0.874	0.702	0.779	0.904	0.864	0.883
63-64	0.851	0.960	0.814	0.944	0.848	0.959	0.906
65-67	0.891	0.945	0.785	0.991	0.881	0.950	0.918
68-69	0.876	0.986	0.741	0.810	0.864	0.970	0.922
70-74	0.955	1.020	0.866	0.949	0.948	1.014	0.985
75-99	0.983	1.019	0.713	0.861	0.962	1.006	0.990
15+	0.911	0.951	0.752	0.877	0.893	0.942	0.919
0+	0.919	0.953	0.802	0.891	0.905	0.945	0.926

Sampling Variability. Sampling variability is variation that occurred by chance because a sample was surveyed rather than the entire population. Standard errors, as calculated by methods described next are primarily measures of sampling variability, although they may include some nonsampling error.

Standard Errors and Their Use. A number of approximations are required to derive, at a moderate cost, standard errors applicable to all the estimates in this report. Instead of providing an individual standard error for each estimate, generalized sets of standard errors are provided for various types of characteristics. Thus, the tables show levels of magnitude of standard errors rather than the precise standard errors.

Table B-3 provides standard errors of estimated numbers. Table B-4 provides standard errors of estimated percentages. Table B-5 has standard error parameters for persons, families, households, householders, and unrelated individuals. Table B-5 also provides factors to apply to the standard errors in tables B-3 and B-4. Table B-6 contains factors to be applied to the parameters in table B-5 to adjust for earlier years.

The sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a range that would include the average result of all possible samples with a known probability. For example, if all possible samples were surveyed under essentially the same general conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

A particular confidence interval may or may not contain the average estimate derived from all possible samples. However, one can say with specified confidence that the interval includes the average estimate calculated from all possible samples.

Some statements in the report may contain estimates followed by a number in parentheses. This number can be added to and subtracted from the estimate to calculate upper and lower bounds of the 90-percent confidence interval. For example, if a statement contains the phrase "grew by 1.7 percent (±1.0)," the 90-percent confidence interval for the estimate, 1.7 percent, is 0.7 percent to 2.7 percent.

Standard errors may be used to perform hypothesis testing. This is a procedure for distinguishing between population parameters using sample estimates. The most common type of hypothesis appearing in this report is that the population parameters are different. An example of this would be comparing the local mobility rates for 1992 and 1993.

Tests may be performed at various levels of significance. The significance level of a test is the probability of concluding that the characteristics are different when,

Table B-3. **Standard Errors of Estimated Numbers of Movers: March 1993**

(Numbers in Thousands)

Size of Estimate	Standard errors	
	Total, White,and Black	Hispanic
10	8	8
25	13	13
50	19	19
75	23	23
100	27	27
250	42	42
500	60	58
750	73	71
1,000	84	81
2,500	133	119
5,000	187	145
7,500	228	143
10,000	262	111
25,000	403	(x)
50,000	542	(x)
75,000	628	(x)
100,000	680	(x)
150,000	712	(x)
200,000	653	(x)
250,000	469	(x)

x = not applicable.

NOTE: Multiply these standard errors by 1.2 for nonmetropolitan data.

Table B-4. **Standard Errors of Estimated Percentages of Movers: March 1993**

(Total. White, Black, and Hispanic)

Base (000)	Estimated percentages						
	1 - 99	2 - 98	5 - 95	10 - 90	25 - 75	35 - 65	50
10	8.4	11.8	18.4	25.3	36.6	40.3	42.2
25	5.3	7.5	11.6	16.0	23.1	25.5	26.7
50	3.8	5.3	8.2	11.3	16.4	18.0	18.9
75	3.1	4.3	6.7	9.2	13.4	14.7	15.4
100	2.7	3.7	5.8	8.0	11.6	12.7	13.4
250	1.7	2.4	3.6	5.1	7.3	8.0	8.4
500	1.2	1.7	2.6	3.6	5.2	5.7	6.0
750	1.0	1.4	2.1	2.9	4.2	4.6	4.9
1,000	0.8	1.2	1.8	2.5	3.7	4.0	4.2
2,500	0.5	0.8	1.2	1.6	2.3	2.6	2.7
5,000	0.4	0.5	0.8	1.1	1.6	1.8	1.9
7,500	0.3	0.4	0.7	0.9	1.3	1.5	1.5
10,000	0.3	0.4	0.6	0.8	1.2	1.3	1.3
25,000	0.2	0.2	0.4	0.5	0.7	0.8	0.8
50,000	0.12	0.2	0.3	0.4	0.5	0.6	0.6
75,000	0.10	0.14	0.2	0.3	0.4	0.5	0.5
100,000	0.08	0.12	0.2	0.2	0.4	0.4	0.4
150,000	0.07	0.10	0.2	0.2	0.3	0.3	0.3
200,000	0.06	0.08	0.13	0.2	0.3	0.3	0.3
250,000	0.05	0.07	0.12	0.2	0.2	0.2	0.3

NOTE: Multiply these standard errors by 1.2 for nonmetropolitan data.

in fact, they are the same. All statements of comparison in the text have passed a hypothesis test at the 0.10 level of significance or better. This means that the absolute value of the estimated difference between characteristics is greater than or equal to 1.645 times the standard error of the difference.

Standard Errors of Estimated Numbers. There are two ways to compute the approximate standard error, s_x, of an estimated number shown in this report. The first uses the formula

$$s_x = fs \qquad (1)$$

where f is a factor from table B-5 and s is the standard error of the estimate obtained by interpolation from table B-3. The second method uses formula (2), from which the standard errors in table B-3 were calculated. This formula will provide more accurate results than formula (1).

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

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 that will give the largest standard error.

Illustration. Suppose that 42,693,000 persons moved between March 1991 and March 1992. Use the appropriate parameters from table B-5 and formula (2) to get

Number, x	42,693,000
a parameter	−0.000025
b parameter	7,130
Standard error	509,000
90% conf. int.	41,856,000 to 43,530,000

The standard error is calculated as

$$s_x = \sqrt{-0.000025x42{,}693{,}000^2 + 7{,}130x42{,}693{,}000} = 509{,}000$$

The 90-percent confidence interval is calculated as 42,693,000 ± 1.645x509,000.

A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

The alternate calculation of the standard error, using formula (1) with f = 1.0 from table B-5 and s = 501,000 by interpolation from table B-3, is

$$s_x = 1.0x501{,}000 = 501{,}000$$

Standard Errors of Estimated Percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on the size of the percentage and its base. 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 parameter from table B-5 indicated by the numerator.

The approximate standard error, $s_{x,p}$, of an estimated percentage can be obtained by use of the formula

$$s_{x,p} = fs \qquad (3)$$

In this formula, f is the appropriate factor from table B-5, and s is the standard error of the estimate obtained by interpolation from table B-4.

Alternatively, formula (4) will provide more accurate results:

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

Here x is the total number of persons, families, households, or unrelated individuals in the base of the percentage, p is the percentage ($0 \leq p \leq 100$), and b is the parameter in table B-5 associated with the characteristic in the numerator of the percentage.

Illustration. Suppose that of a total of 232,998,000 persons 1 year old and over, 1,200,000 or 0.5 percent moved from abroad. Use the appropriate parameter from table B-5 and formula (4) to get

Percentage, p	0.5
Base, x	232,998,000
b parameter	7,130
Standard error	0.04
90% conf. int.	0.4 to 0.6

The standard error is calculated as

$$s_{x,p} = \sqrt{\frac{7{,}130}{232{,}998{,}000} x0.5x(100.0-0.5)} = 0.04$$

The 90-percent confidence interval for the percentage of persons who moved from abroad is calculated as 0.5 ± 1.645x0.04.

The alternate calculation of the standard error, using formula (3) with f = 1.0 from table B-5 and s = 0.05 by interpolation from table B-4, is

$$s_{x,p} = 1.0x0.05 = 0.05$$

Standard Error of a Difference. The standard error of the difference between two sample estimates is approximately equal to

$$s_{x-y} = \sqrt{s_x^2 + s_y^2} \qquad (5)$$

where s_x and s_y are the standard errors of the estimates, x and y. The estimates can be numbers, percentages, ratios, etc. This will represent the actual standard error quite accurately for the difference between estimates of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. However, if there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration. Suppose 42,693,000 persons moved this year and 43,237,000 persons moved last year. Use the appropriate parameters from table B-5 and formulas (2) and (5) to get

	x	y	difference
Number	42,693,000	43,237,000	544,000
a parameter	−0.000025	−0.000028	-
b parameter	7,130	7,925	-
Standard error	509,000	539,000	741,000
90% conf. int.	41,856,000 to 43,530,000	42,350,000 to 44,124,000	-675,000 to 1,763,000

The standard error of the difference is calculated as

$$s_{x\text{-}y} = \sqrt{509{,}000^2 + 539{,}000^2} = 741{,}000$$

The 90-percent confidence interval around the difference is calculated as 544,000 ± 1.645x741,000. Since this interval contains zero we cannot conclude, at the 10-percent significance level, that the number of movers this year is less than the number of movers last year.

Standard Error of a Mean for Grouped Data. The formula used to estimate the standard error of a mean for grouped data is

$$s_{\bar{x}} = \sqrt{(b/y)S^2} \qquad (6)$$

In this formula, y is the size of the base of the distribution and b is a parameter from table B-3. The variance, S^2, is given by the following formula:

$$S^2 = \sum_{i=1}^{c} p_i\bar{x}_i^2 - \bar{x}^2 \qquad (7)$$

where $\bar{x}$, the mean of the distribution, is estimated by

$$\bar{x} = \sum_{i=1}^{c} p_i\bar{x}_i \qquad (8)$$

c is the number of groups; i indicates a specific group, thus taking on values 1 through c.

p_i is the estimated proportion of households, families or persons whose values, for the characteristic (x-values) being considered, fall in group i.

$\bar{x}$ is $(Z_{i-1} + Z_i)/2$ where Z_{i-1} and Z_i are the lower and upper interval boundaries, respectively, for group i. $\bar{x}_i$ is assumed to be the most representative value for the characteristic for households, families, and unrelated individuals or persons in group i. Group c is open-ended, i.e., no upper interval boundary exists. For this group the approximate average value is

$$\bar{x}_c = \frac{3}{2} z_{c-1} \qquad (9)$$

Standard Error of a Ratio. Certain estimates may be calculated as the ratio of two numbers. The standard error of a ratio, x/y, may be computed using

$$s_{x/y} = \frac{x}{y}\sqrt{\left[\frac{s_x}{x}\right]^2 + \left[\frac{s_y}{y}\right]^2 - 2r\frac{s_x}{x}\frac{s_y}{y}} \qquad (10)$$

The standard error of the numerator, s_x, and that of the denominator, s_y, may be calculated using formulas described earlier. In formula (10), r represents the correlation between the numerator and the denominator of the estimate.

For one type of ratio, the denominator is a count of families or households and the numerator is a count of persons in those families or households with a certain characteristic. If there is at least one person with the characteristic in every family or household, use 0.7 as an estimate of r. An example of this type is the mean number of children per family with children.

For all other types of ratios, r is assumed to be zero. If r is actually positive (negative), then this procedure will provide an overestimate (underestimate) of the standard error of the ratio. Examples of this type are the mean number of children per family and the poverty rate. NOTE: For estimates expressed as the ratio of x per 100 y or x per 1,000 y, multiply formula (10) by 100 or 1,000, respectively, to obtain the standard error.

Illustration. Suppose the ratio of male movers from abroad, x, to female movers from abroad, y, is 1.28. Use the appropriate parameters from table B-5. The standard error of this ratio is calculated as follows:

	x	y	ratio
Estimate	641,000	501,000	1.28
a parameter	−0.000025	−0.000025	-
b parameter	7,130	7,130	-
Standard error	68,000	60,000	0.20
90% conf. int.	529,000 to 753,000	402,000 to 600,000	0.95 to 1.61

Using formula (10) with r = 0, the estimate of the standard error is

$$s_{x/y} = \frac{641{,}000}{501{,}000}\sqrt{\left[\frac{68{,}000}{641{,}000}\right]^2 + \left[\frac{60{,}000}{501{,}000}\right]^2} = 0.20$$

The 90-percent confidence interval is calculated as 1.28 ± 1.645x0.20.

Standard Error of a Median. The sampling variability of an estimated median depends on the form of the distribution and the size of the base. One can approximate the reliability of an estimated median by determining a confidence interval about it. (See the section on standard errors and their use for a general discussion of confidence intervals.)

Estimate the 68-percent confidence limits of a median based on sample data using the following procedure.

Table B-5. **a and b Parameters and Factors (f) for Data collected March 1993**

Type of Characteristic	Parameters		
	a	b	f
MOVERS			
U.S., county, state or region			
Both sexes:			
Total or White	−0.000025	7,130	1.0
Black	−0.000025	7,130	1.0
Hispanic	−0.000589	7,130	1.0
Male only or female only:			
Total or White	−0.000011	3,530	0.7
Black	−0.000011	3,530	0.7
Hispanic	−0.000292	3,530	0.7
Educational attainment, labor force, marital status, household, family, and income:			
Total or White	−0.000035	2,626	0.6
Black	−0.000035	2,626	0.6
Hispanic	0.000217	2,626	0.6
Persons below poverty level:			
Total or White	−0.000037	9,501	1.2
Black	−0.000037	9,501	1.2
Hispanic	−0.000037	9,501	1.2
TOTAL OR NONMOVERS			
Residence:			
Total or White	−0.000025	7,130	1.0
Black	−0.000025	7,130	1.0
Hispanic	−0.000589	7,130	1.0
Family income:			
Total or White	−0.000012	2,058	0.5
Black	−0.000106	2,243	0.6
Hispanic	−0.000185	2,243	0.6
Persons below the poverty level:			
Total or White	−0.000037	9,501	1.2
Black	−0.000323	9,501	1.2
Hispanic	−0.000785	9,501	1.2
Years of school completed:			
Total or White	−0.000019	2,468	0.6
Black	−0.000222	3,339	0.7
Hispanic	−0.000318	3,339	0.7
Employment status and occupation:			
Both sexes	−0.000019	2,485	0.6
Male only	−0.000030	2,150	0.6
Female only	−0.000023	1,843	0.5
Unemployment:			
Total or White	−0.000018	2,357	0.6
Black	−0.000166	2,708	0.6
Hispanic	−0.000224	2,708	0.6
Marital status, household & family characteristics			
Total or White:			
Families, householders or primary individuals	−0.000011	1,899	0.5
Some household members	−0.000026	4,785	0.8
All household members	−0.000033	5,815	0.9
Black:			
Families, householders or primary individuals	−0.000071	1,716	0.5
Some household members	−0.000283	6,864	1.0
All household members	−0.000417	10,121	1.2
Hispanic:			
Families, householders or primary individuals	−0.000142	1,716	0.5
Some household members	−0.000567	6,864	1.0
All household members	−0.000836	10,121	1.2

Note: Multiply the above parameters by 1.5 for nonmetropolitian data.

Table B-6. **Factors to Obtain Parameters for Previous Years to be Applied to the Parameters for March 1993**

Total	Total, White or Black	Hispanic
March 1951 and March 1956	1.88	2.64
March 1961 and March 1966	1.25	1.76
March 1970 through March 1972	0.81	1.14
March 1973 through March 1976	0.82	1.14
March 1977 through March 1981	0.84	1.17
March 1982 through March 1984	0.94	1.31
March 1985 through March 1988	0.94	0.94
March 1989	1.11	1.16
March 1990 through March 1992	1.00	1.00

Note: Multiply the above parameters by 1.5 for nonmetropolitan data

1. Determine, using formula (4), the standard error of the 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, determine upper and lower limits of the 68-percent confidence interval by calculating values corresponding to the two points established in step 2.
   Use the following formula to calculate the upper and lower limits.

$$x_{pN} = \frac{pN - N_1}{N_2 - N_1}(A_2 - A_1) + A_1 \qquad (11)$$

   where

   X_{pN} = estimated upper and lower bounds for the confidence interval ($0 \leq p \leq 1$). For purposes of calculating the confidence interval, p takes on the values determined in step 2. Note that X_{pN} estimates the median when $p = 0.50$.

   N = *for distribution of numbers:* the total number of units (persons, households, etc.) for the characteristic in the distribution.

   = *for distribution of percentages:* the value 1.0.

   p = the values obtained in step 2.

   A_1, A_2 = the lower and upper bounds, respectively, of the interval containing X_{pN}.

   N_1, N_2 = *for distribution of numbers:* the estimated number of units (persons, households, etc.) with values of the characteristic greater than or equal to A_1 and A_2, respectively.

   = *for distribution of percentages:* the estimated percentage of units (persons, households, etc.) having values of the characteristic greater than or equal to A_1 and A_2, respectively.
4. Divide the difference between the two points determined in step 3 by two to obtain the standard error of the median.

Illustration

A recent report by the Bureau of the Census shows the following distribution and median income for families in 1989.

Income levels	Families
Total	66,090
Under $5,000	2,398
$5,000 to $9,999	4,141
$10,000 to $14,999	5,354
$15,000 to $19,999	5,565
$20,000 to $24,999	5,461
$25,000 to $29,999	5,576
$30,000 to $34,999	5,294
$35,000 to $39,999	4,959
$40,000 to $44,999	4,464
$45,000 to $49,999	3,689
$50,000 to $54,999	3,545
$55,000 to $59,999	2,595
$60,000 to $64,999	2,278
$65,000 to $69,999	1,839
$70,000 to $74,999	1,463
$75,000 to $79,999	1,251
$80,000 to $84,999	1,036
$85,000 to $89,999	774
$90,000 to $94,999	695
$95,000 to $99,999	518
$100,000 and over	3,197
Median income (dollars)	$34,213

[1]U.S. Bureau of the Census, Current Population Reports, Series P-60, No. 168, *Money Income and Poverty Status in the United States: 1989 (Advance Data from the March 1990 Current Population Survey)* U.S. Government Printing Office, Washington, D.C., 1990.

1. Using formula (4) with b = 2,058, the standard error of 50 percent on a base of 66,090,000 is about 0.3 percent.
2. To obtain a 68-percent confidence interval on an estimated median, add to and subtract from 50 percent the standard error found in step 1. This yields percent limits of 49.7 and 50.3.
3. The lower and upper limits for the interval in which the median falls are $30,000 and $35,000, respectively.
   Then, by addition, the estimated numbers of families with an income greater than or equal to $30,000 and $35,000 are 37,597,000 and 32,303,000, respectively.

Using formula (11), the upper limit for the confidence interval of the median is found to be about

$$\frac{0.497 \times 66{,}090{,}000 - 37{,}597{,}000}{32{,}303{,}000 - 37{,}597{,}000} \times (35{,}000 - 30{,}000) + 30{,}000 = 34{,}500$$

Similarly, the lower limit is found to be about

$$\frac{0.503 \times 66{,}090{,}000 - 37{,}597{,}000}{32{,}303{,}000 - 37{,}597{,}000} \times (35{,}000 - 30{,}000) + 30{,}000 = 34{,}100$$

Thus, a 68-percent confidence interval for the median income for families is from $34,100 to $34,500.

4. The standard error of the median is, therefore,

$$\frac{34{,}500 - 34{,}100}{2} = 200$$

Standard Error of the Average Number of Moves. For this report, the standard error of the average number of moves for a given age interval is equal to

$$S\left(\frac{TM_j}{l_j}\right) = \frac{\sqrt{\sum_{i=j}^{18}\left(\frac{L_i^2 R_i (1 - R_i) b}{y_i}\right)}}{l_j}$$

where L_i is the stationary population for age interval i; R_i is the average mobility rate for age interval i; l_j is the population still living per 100,000 born for age interval j; y_i is the total population for age interval i; b is the geographical mobility parameter. There are 18 age intervals, where age interval 1 includes 1-4 year olds, age interval 2 includes 5-9 year olds, etc.

Note that this formula assumes L_i and l_j are constants and therefore have zero variance. Please contact the Demographic Statistical Methods Division if a more precise standard error estimate is needed.