Appendix B. Source and Reliability of Estimates

SOURCE OF DATA

Most of the estimates in this report are based on data collected in March 1981 and 1980 from the Current Population Survey (CPS) of the Bureau of the Census. Some estimates are based on data obtained from the CPS in earlier years and from the 1960, 1950 and 1940 decennial censuses. 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, supplementary questions are asked each March about educational attainment. In order to obtain more reliable data for the Spanish-origin population, the March CPS sample was enlarged to include all households from the previous November sample which contained at least one person of Spanish origin. This results in almost doubling the number of sample persons of Spanish origin. For this report, persons in the Armed Forces living off post or with their families on post are included.

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. The current CPS sample is located in 629 areas comprising 1,148 counties, independent cities, and minor civil divisions in the Nation. In this sample, approximately 68,500 occupied households were eligible for interview. Of this number, about 3,000 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.

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 March Current Population Survey

Time period	Number of sample areas	Housing units eligible: Interviewed	Housing units eligible: Not interviewed
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
1957 to 1959.......	330	33,500	1,500
1954 to 1956.......	230	21,000	500-1,000
1947 to 1953.......	68	21,000	500-1,000

The estimation procedure used in this survey involved 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 were based on statistics from decennial censuses, statistics on births, deaths, immigration and emigration; and statistics on the strength of the Armed Forces. The independent population estimates used in this report to obtain data for 1980 and 1981 are based on the 1980 decennial census. Some of the data in this report for 1980 were also obtained using independent population estimates based on the 1970 decennial census. In earlier reports in this series (P-20), data for 1972 through 1979 were obtained using independent population estimates based on the 1970 decennial census. For more details on this change, see the section in appendix A entitled, "Introduction of 1980 Census Population Controls." The estimation procedure for the data in the report also involved a further adjustment so that husband and wife of a household received the same weight.

RELIABILITY OF ESTIMATES

Since the CPS 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 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 in 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 nonsampling error is unknown. Consequently, particular care should be exercised in the interpretation of figures based on a relatively small number of cases or on small differences between estimates.

Nonsampling variability. As in any survey work, the results are subject to errors of response and nonreporting in addition to sampling variability. Nonsampling errors can be attributed to many sources, e.g., inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, inability or unwillingness of respondents to provide correct information, inability to recall information, errors made in collection such as in recording or coding 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 with 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 combined 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 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 decennial census.

Comparability with other data. In using metropolitan and nonmetropolitan data, caution should be used in comparing estimates for 1977 and 1978 to each other or to any other years. Methodological and sample design changes occurred in these years resulting in relatively large differences in the metropolitan and nonmetropolitan area estimates. However, estimates for 1979 and later are comparable as are estimates for 1976 and earlier.

A number of changes were made in data collection and estimation procedures beginning with the March 1980 CPS. One major change was the use of the "householder" concept instead of the traditional "head" concept. The other major change occurred in the estimation procedure. Due to these and other changes caution should also be used when comparing estimates for 1980 and later, which reflect 1980 census-based population controls, to those for 1971 through 1979, which reflect 1970 census-based population controls. This change in population controls had relatively little impact on summary measures such as means and medians, but did have a significant impact on levels. For example, use of 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 1980 and later will differ from those for 1979 and earlier 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.

Decennial Census of Population. The 1940, 1950 and 1960 decennial census data shown in this report are not strictly comparable to the CPS data. This is due in a large part to differences in interviewer training and experience and in different survey processes. This is an additional component of error not reflected in the standard error tables. Therefore, caution should be used in comparing results between these different sources.

Sampling variability. The standard errors given in tables B-1 through B-4 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 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 being 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:

1. Approximately 68 percent of the intervals from one standard errors 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.

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 are 1) the population parameters are identical, or 2) they are different. An example of this would be comparing the percent of adults who were high school graduates in 1981 to those in 1971. 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 will be qualified in some way; e.g., by use of the phrase "some evidence."

Note when using small estimates. Summary measures (such as means, medians, and percent distributions) are shown 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 the corresponding percentages.

These smaller estimates are provided primarily to permit such combinations of the categories as serve each 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, 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.

Standard errors for data based on the CPS sample. The figures presented in tables B-1 through B-4 are approximations to standard errors of various educational attainment estimates for persons in the United States. Estimated standard errors for specific characteristics cannot be obtained from tables B-1 through B-4 without the use of the factors in table B-5. The factors in table B-5 must be applied to the generalized standard errors in order to adjust for the combined effect of sample design and estimating procedure on the value of the characteristic. Further adjustments must be made to account for State, SMSA and regional variation using factors given in table B-6. Standard errors for intermediate values not shown in the generalized tables of standard errors may be approximated by linear interpolation.

Two parameters (denoted ''a'' and ''b'') are used to calculate standard errors for each type of characteristics; they are also presented in table B-5. These parameters were used to calculate the factors in table B-5. 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. The standard errors in tables B-1 through B-4 were calculated using the ''b'' parameters in table B-5; however, the ''a'' parameters used were revised to reflect the total persons in the age group.

Data based on the 1940, 1950 and 1960 censuses. Sampling errors of all sample data from 1940, 1950 and 1960 decennial censuses in this report are small enough to be disregarded. However, these standard errors may be found in the appropriate volumes.

Standard errors of estimated numbers. The approximate standard error σ_x, of an estimated number can be obtained in two ways. It may be obtained by use of the formula

$$\sigma_x = f_1 \cdot f_2 \cdot \sigma \qquad (1)$$

Table B-1. Generalized Standard Errors for Estimated Numbers of Persons—Total or White

(Numbers in thousands)

Estimated number of persons	Total persons in age group[1]									
	100	250	500	1,000	2,500	5,000	10,000	25,000	50,000	100,000
10	4.3	4.4	4.4	4.5	4.5	4.5	4.5	4.5	4.5	4.5
20	5.7	6.1	6.2	6.3	6.3	6.3	6.3	6.3	6.3	6.3
30	6.5	7.3	7.5	7.7	7.7	7.7	7.8	7.8	7.8	7.8
40	7.0	8.2	8.6	8.8	8.9	8.9	9.0	9.0	9.0	9.0
50	7.1	9.0	9.5	9.8	9.9	10.0	10.0	10.0	10.0	10.0
75	6.1	10.3	11.3	11.8	12.1	12.2	12.2	12.3	12.3	12.3
100	-	11.0	12.7	13.5	13.9	14.0	14.1	14.2	14.2	14.2
200	-	9.0	15.5	18.0	19.3	19.7	19.9	20.0	20.0	20.0
300	-	-	15.5	20.6	23.1	23.8	24.2	24.4	24.5	24.5
400	-	-	12.7	22.0	26.0	27.2	27.8	28.2	28.3	28.3
500	-	-	-	22.4	28.4	30.1	30.9	31.4	31.6	31.7
750	-	-	-	19.4	32.5	35.8	37.4	38.3	38.6	38.7
1,000	-	-	-	-	34.8	40.1	42.6	44.0	44.4	44.7
2,000	-	-	-	-	28.4	49.2	56.8	60.9	62.2	62.8
3,000	-	-	-	-	-	49.2	65.0	72.9	75.4	76.6
4,000	-	-	-	-	-	40.1	69.5	82.3	86.1	87.9
5,000	-	-	-	-	-	-	71.0	89.8	95.2	97.8
7,500	-	-	-	-	-	-	61.5	102.8	113.3	118.2
10,000	-	-	-	-	-	-	-	109.9	126.9	134.6
20,000	-	-	-	-	-	-	-	89.8	155.5	179.5
30,000	-	-	-	-	-	-	-	-	155.5	205.7
40,000	-	-	-	-	-	-	-	-	126.9	219.9
50,000	-	-	-	-	-	-	-	-	-	224.4
75,000	-	-	-	-	-	-	-	-	-	194.3
100,000	-	-	-	-	-	-	-	-	-	-

[1] These values must be multiplied by the appropriate "f" factor in tables B-5 and/or B-6 to obtain the standard error for a specific characteristic.

Note: To estimate the standard errors for the 1956-66 period, multiply these standard errors by 1.23. For years prior to 1956, multiply by 1.5.

Table B-2. Generalized Standard Errors for Estimated Numbers of Persons—Black and Other Races

(Numbers in thousands)

Estimated number of persons	Total persons in age group[1]						
	100	250	500	1,000	2,500	5,000	10,000
10	4.5	4.7	4.7	4.7	4.7	4.8	4.8
20	6.0	6.5	6.6	6.7	6.7	6.7	6.7
30	6.9	7.7	8.0	8.1	8.2	8.2	8.2
40	7.4	8.7	9.1	9.3	9.4	9.5	9.5
50	7.5	9.5	10.1	10.4	10.5	10.6	10.6
75	6.5	10.9	12.0	12.5	12.8	12.9	13.0
100	-	11.7	13.5	14.3	14.7	14.9	15.0
200	-	9.5	16.5	19.0	20.4	20.9	21.1
300	-	-	16.5	21.8	24.5	25.3	25.7
400	-	-	13.5	23.3	27.6	28.9	29.5
500	-	-	-	23.8	30.1	31.9	32.8
750	-	-	-	20.6	34.5	38.0	39.6
1,000	-	-	-	-	36.9	42.6	45.1
2,000	-	-	-	-	30.1	52.1	60.2
3,000	-	-	-	-	-	52.1	69.0
4,000	-	-	-	-	-	42.6	73.7
5,000	-	-	-	-	-	-	75.2
7,500	-	-	-	-	-	-	65.2
10,000	-	-	-	-	-	-	-

[1]These values must be multiplied by the appropriate "f" factor in tables B-5 and/or B-6 to obtain the standard error for a specific characteristic.

Note: To estimate the standard errors for the 1956-66 period, multiply these standard errors by 1.23. For years prior to 1956, multiply by 1.5.

where f_1 is the appropriate factor from table B-5, f_2 is the appropriate factor from table B-6 and σ is the standard error of the estimate obtained by interpolation from tables B-1 or B-2. Alternatively, standard errors may be approximated by using formula (2). The use of formula (1) will provide more accurate results when the number of persons in the age group is relatively small.

$$\sigma_x = f_2 \sqrt{ax^2 + bx} \qquad (2)$$

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 an estimate involves two different categories use the "a" and "b" parameters corresponding to the category with the larger "b" parameter.

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 this percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. The approximate standard error, $\sigma_{(x,p)}$ of an estimated percentage can be obtained by use of the formula

$$\sigma_{(x,p)} = f_1 \cdot f_2 \cdot \sigma \qquad (3)$$

In this formula, f_1 is the appropriate factor from table B-5, f_2 is the appropriate factor from table B-6 and σ is the standard error of the estimate from table B-3 or B-4. When the numerator and denominator of the percentage are in different categories, use the factor or parameters from table B-5 indicated by the numerator. Alternatively, standard errors may be approximated by using formula (4), from which the standard errors in tables B-3 and B-4 were calculated. Use of this formula will provide more accurate results than use of formula (3).

$$\sigma_{(x,p)} = f_2 \sqrt{\frac{b}{x} \cdot p(100-p)} \qquad (4)$$

Here x is the size of the subclass of persons or families and unrelated individuals 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 standard error computations. Table 8 of this report shows that in March 1981 there were 70,390,000 women 25 years and over. At that time, an estimated 28,896,000 had completed 4 years of high school. Using formula (2) and the appropriate "a" and "b" parameters from table B-5, the approximate standard error of this estimate[1] is

$$193{,}000 \doteq \sqrt{-.000025\,(28{,}896{,}000)^2 + 2014\,(28{,}896{,}000)}$$

The 95-percent confidence interval as derived from the data is from 28,510,000 to 29,282,000 (using twice the

[1]Using formula (1), interpolating from table B-1 and applying the appropriate factors from table B-5 and/or B-6 gives a standard error of approximately 175,000 = 1.0 x 1.0 x 175,000.

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

Base of percentage (thousands)	Estimated percentage[1]				
	2 or 98	5 or 95	10 or 90	25 or 75	50
100	2.0	3.1	4.3	6.1	7.1
250	1.3	2.0	2.7	3.9	4.5
500	0.9	1.4	1.9	2.7	3.2
1,000	0.6	1.0	1.3	1.9	2.2
2,500	0.4	0.6	0.9	1.2	1.4
5,000	0.3	0.4	0.6	0.9	1.0
10,000	0.2	0.3	0.4	0.6	0.7
25,000	0.1	0.2	0.3	0.4	0.4
50,000	0.1	0.1	0.2	0.3	0.3
100,000	0.1	0.1	0.1	0.2	0.2
150,000	0.1	0.1	0.1	0.2	0.2

[1]These values must be multiplied by the appropriate "f" factor in tables B-5 and/or B-6 to obtain the standard error for a specific characteristic.

NOTE: To estimate the standard errors for the 1956-66 period, multiply these standard errors by 1.23. For years prior to 1956, multiply by 1.5.

standard error). A conclusion that the average estimate of women 25 years old and over having completed 4 years of high school derived from all possible samples and lying between these two values would be correct for roughly 95 percent of all possible samples.

Table 12 of this report shows that in March 1981, an estimated 21.1 percent of the 62,509,000 men 25 years old and over had completed college. Using formula (4) and the appropriate "b" parameter from table B-5, the approximate standard error of this estimate[2] is

$$0.2 \doteq \sqrt{\frac{2014}{62{,}509{,}000} \times 21.1 \times (100 - 21.1)}$$

The 95-percent confidence interval for this estimate is from 20.7 to 21.5

[2]Using formula (3), interpolating and applying the appropriate factors from table B-5 and /or B-6 gives a standard error of approximately 0.2.

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

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

where σ_x and σ_y are the standard errors of the estimates x and y; the estimates can be numbers, percents, ratios, etc. This formula approximates the standard error quite accurately for the difference between two estimates of the same characteristics 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 calculation of the standard error of a difference. Table 12 shows that in March 1981, an estimated 21.1 percent of men 25 years old and over had completed

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

Base of percentage (thousands)	Estimated percentage[1]				
	2 or 98	5 or 95	10 or 90	25 or 75	50
75	2.4	3.8	5.2	7.5	8.7
100	2.1	3.3	4.5	6.5	7.5
250	1.3	2.1	2.9	4.1	4.8
500	0.9	1.5	2.0	2.9	3.4
1,000	0.7	1.0	1.4	2.1	2.4
2,500	0.4	0.7	0.9	1.3	1.5
5,000	0.3	0.5	0.6	0.9	1.1
10,000	0.21	0.3	0.5	0.7	0.8
15,000	0.17	0.27	0.4	0.5	0.6
20,000	0.15	0.23	0.3	0.5	0.5

[1]These values must be multiplied by the appropriate "f" factor in tables B-5 and/or B-6 to obtain the standard error for a specific characteristic.

NOTE: To estimate the standard errors for the 1956-66 period, multiply these standard errors by 1.23. For years prior to 1956, multiply by 1.5.

Table B-5. "a" and "b" Parameters and "f" Factors for Calculating Approximate Standard Errors of Estimated Numbers and Percentages

Type of characteristic	Parameter[1] a	Parameter[1] b	f_1 factor[4]
Educational attainment of persons 14 and over:			
Total or White	-.000025	2,014	1.0
Black	-.000179	2,265	1.0
Spanish origin (levels)	[2]+.000901	[2]1,101	(X)
Spanish origin (percentages only)	(X)	[3]2,002	1.0
Marital status:			
Total or White	-.000017	3,500	1.3
Black	-.000210	5,020	1.5
Spanish origin	-.000026	4,432	1.5
Household relationship:			
Head, wife, or primary individual:			
Total or White	-.000010	1,389	0.8
Black	-.000087	1,255	0.7
Spanish origin	-.000020	1,422	0.8
Child or other relative in primary family, secondary family member secondary individual, or persons living in group quarters:			
Total or White	-.000017	3,500	1.3
Black	-.000210	5,020	1.5
Spanish origin	-.000026	4,432	1.5

[1]Multiply parameters by 1.5 for CPS data collected from 1956 to 1966 and by 2.25 for CPS data collected before 1956.
[2]These "a" and "b" parameters are to be used to calculate standard errors of levels only for the March supplement. For the October enrollment data, Spanish origin, use a = 0.001519 and b = 1856. Use formulas (2) and (4) only for these standard error calculations.
[3]This "b" parameter is to be used to calculate the standard error of percentages only. For the October enrollment data use b = 3374.
[4]These factors are to be applied when standard error calculations are made using formulas (1) and (3) only.

college as compared to 13.4 percent of women. The apparent difference between the two estimates is 7.7 percent. The standard error of 21.1 percent as shown above is 0.2. Similarly, the standard error of 13.4 percent is computed to be 0.2 percent. Therefore, using formula (5), the standard error of the estimated difference, 7.7 percent, is about

$$.3 \doteq \sqrt{(.2)^2 + (.2)^2}$$

This means that the 95-percent confidence interval about the difference of 7.7 percent is from 7.1 to 8.3 percent. 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 95 percent of all possible samples. Since this confidence interval does not contain zero, we may conclude with 95 percent confidence that the percent of men 25 years old and over completing college is greater than the percent of women.

Confidence interval and 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 an estimated median is to determine a confidence interval about it. (See the section on sampling variability for a general discussion of confidence intervals.) The following procedure may be used to estimate the 68-percent confidence limits and hence the standard error of a median based on sample data.

1. Determine, using the standard error tables and factors or 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 characteristics, calculate the values from the distribution corresponding to the two points established in step (2). These values will be the limits for the 68-percent confidence interval.

4. Divide the difference between the two points determined in step (3) by two to obtain the standard error of the median.

A 95-percent confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step (1).

The formula used to implement step (3) is

$$X = \left(\frac{L - A_1}{A_2 - A_1}\right) X \left(T_2 - T_1\right) + T_1 \qquad (6)$$

where

X = the interpolated estimate of the median's upper or lower limit,

L = the upper or lower limit obtained from step (2),

A_1 and A_2 = the lower and upper values of the cumulative relative frequency for the limits of the interval which contains L,

T_1 and T_2 = the lower and upper values of the interval which contains L.

Illustration of the computation of a confidence interval for a median. Table 9 of this report shows that the median number of school years completed by Floridians 25 years old and over is 12.5.

Table B-6. Factors to be Applied to Standard Errors

Type of characteristic	f_2 factor
U.S. totals	1.0
States:	
California	1.1
Florida	1.1
Georgia	1.1
Illinois	1.1
Indiana	1.0
Massachusetts	0.9
Michigan	1.1
Missouri	1.1
New Jersey	1.0
New York	1.0
North Carolina	1.1
Ohio	1.1
Pennsylvania	1.1
Texas	1.2
Virginia	1.3
Regions:	
Northeast	1.0
Midwest[1]	1.0
South	1.0
West	0.9
SMSA's (except Washington, D.C.)	1.1
Washington, D.C. - MD. - VA.	1.0

Note: Apply these factors to standard errors obtained using either tables B-1 through B-4 or the "a" and "b" parameters in table B-5.

[1]Formerly the North Central Region.

1. There was a total of 6,484,000 Floridians 25 years and older. Using formula (4) and tables B-5 and B-6, the standard error of a 50 percent characteristic is found to be approximately 1.0 x 1.1 x .9 = 1.0.

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 49.0 and 51.0.

3. To use formula (6) as described in step 3 of this previous section, the cumulative relative frequency distribution of number of school years completed by Floridians must be calculated from the data in table 9.

   a. Construct a cumulative relative frequency table by number of school years completed as shown in example 1.

   b. Apply formula (6), using the cumulative relative frequency in example 1, to the upper and lower limits, **51.0** and **49.0**, calculated in step (2) above. From the example, 31.4 percent of Floridians have completed less than 12 years of high school. Thus, in formula (6), $T_1 = 12$, $T_2 = 13$, $A_1 = 31.4$, $A_2 = 69.1$ and $L = 49.0$.

Example 1. Cumulative Relative Frequency of Number of School Years Completed by Floridians

Number of school years completed	0-12.0	12.1-13.0	13.1 and over
Observed relative frequency	31.4	37.7	30.9
Cumulative relative frequency	31.4	69.1	100.0

Thus, the estimated lower limit of the **68** percent confidence interval for the median would be

$$12.47 = \left(\frac{49.0 - 31.4}{69.1 - 31.4}\right) X \left(13-12\right) + 12$$

Similarly, the upper limit may be found using formula (6) with $T_1 = 12$, $T_2 = 13$, $A_1 = 31.4$, $A_2 = 69.1$ and $L = 51.0$.

$$12.52 \doteq \left(\frac{51.0 - 31.4}{69.1 - 31.4}\right) X \left(13-12\right) + 12$$

4. The standard error of the median is,therefore, (12.52 – 12.47)/2, i.e., .03

*U.S. GOVERNMENT PRINTING OFFICE : 1984 0-420-988/216