Appendix C. Source and Reliability of Estimates

SOURCE OF DATA

The data were obtained in the fourth and fifth interview waves of the 1984 panel of the Survey of Income and Program Participation (SIPP). The SIPP universe is the noninstitutionalized resident population living in the United States. This population includes persons living in group quarters, such as dormitories, rooming houses, and religious-group dwellings. Crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates and nursing home residents, were not eligible to be in the survey. Similarly, United States citizens residing abroad were not eligible to be in the survey. With these qualifications, persons who were at least 15 years of age at the time of interview were eligible to be in the survey.

The 1984 SIPP sample is located in 174 areas comprising 450 counties (including one partial county) and independent cities. Within these areas, the bulk of the sample consisted of clusters of two to four living quarters (LQ's), systematically selected from lists of addresses prepared for the 1970 decennial census. The sample was updated to reflect new construction.

Approximately 26,000 living quarters were designated for the sample. For wave 1, interviews were obtained from the occupants of about 19,900 of the designated living quarters. Most of the remaining 6,100 living quarters were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. However, approximately 1,000 of the 6,100 living quarters were not interviewed because the occupants refused to be interviewed, could not be found at home, were temporarily absent, or were otherwise unavailable. Thus, occupants of about 95 percent of all eligible living quarters participated in wave 1 of the survey.

For the subsequent waves, only original sample persons (those interviewed in the first wave) and persons living with them were eligible to be interviewed. With certain restrictions, original sample persons were to be followed even if they moved to a new address. All noninterviewed households from wave 1 were automatically designated as noninterviews for all subsequent waves. When original sample persons moved without leaving a forwarding address or moved to extremely remote parts of the country, additional noninterviews resulted.

Tabulations in this report were drawn from interviews conducted from November 1984 through April 1985. Table C-1 summarizes information on nonresponse for the interview months used to produce this report. Note that a sample cut of 17.8 percent was implemented in March 1985 due to funding difficulties.

The estimation procedure used to derive SIPP person weights involved several stages of weight adjustments. In the first wave, each person received a base weight equal to the inverse of his/her probability of selection. For each subsequent interview, each person received a base weight that accounted for following movers.

A noninterview adjustment factor was applied to the weight of every occupant of interviewed households to account for households which were eligible for the sample but were not interviewed. (Individual nonresponse within partially interviewed households was treated with imputation. No special adjustment was made for noninterviews in group quarters.) A factor was applied to each interviewed person's weight to account for the SIPP sample areas not having the same population distribution as the strata from which they were selected.

An additional stage of adjustment to persons' weights was performed to bring the sample estimates into agreement with independent monthly estimates of the civilian (and some military) noninstitutional population of the United States by age, race, and sex. These independent estimates were based on statistics from the 1980 Decennial Census of Population; statistics on births, deaths, immigration, and emigration; and statistics on the strength of the Armed Forces. To increase accuracy, weights were further adjusted in such a manner that SIPP sample estimates would closely agree with special Current Population Survey (CPS) estimates by type of householder (married, single with relatives or single without relatives by sex and race) and relationship to householder (spouse or other).[1] The estimation procedure for the data in the report also involved an adjustment so that the husband and wife of a household received the same weight.

RELIABILITY OF ESTIMATES

SIPP 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 questionnaire, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: nonsampling and sampling. We are able to provide estimates of the magnitude of SIPP sampling error, but this is not true of nonsampling error. Descriptions of sources of

[1]These special CPS estimates are slightly different from the published monthly CPS estimates. The differences arise from forcing counts of husbands to agree with counts of wives.

SIPP nonsampling error, along with a discussion of sampling error, its estimation, and its use in data analyses follow.

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 on the part of the respondents to provide correct information, inability to recall information, errors made in collection such as in recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, biases resulting from the differing recall periods caused by the rotation pattern, and failure to represent all units within the sample (undercoverage). Quality control and edit procedures were used to minimize errors made by respondents and interviewers.

Table C-1. Sample Size, by Month and Interview Status

Month	Household units eligible			
	Total	Interviewed	Not interviewed	Nonresponse rate
November 1984 . .	5,600	4,700	900	[1]15
December 1984 . .	5,600	4,700	900	17
January 1985. . . .	5,600	4,700	900	16
February 1985 . . .	5,600	4,700	1,000	17
March 1985	4,600	3,800	800	18
April 1985	4,700	3,800	900	18

[1]Due to rounding of all numbers at 100, there are some inconsistencies. The percentage was calculated using unrounded numbers.

Undercoverage in SIPP results from missed living quarters and missed persons within sample households. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for Blacks than for non-Blacks. Ratio estimation to independent age-race-sex population controls partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that persons in missed households or missed persons in interviewed households have different characteristics than interviewed persons in the same age-race-sex group. Further, the independent population controls used have not been adjusted for undercoverage in the decennial census.

The Bureau has used complex techniques to adjust the weights for nonresponse, but the success of these techniques in avoiding bias is unknown.

Comparability with other statistics. Caution should be exercised when comparing data from this report with data from earlier SIPP publications or with data from other surveys. The comparability problems are caused by the seasonal patterns to which many characteristics are subject and by different nonsampling errors.

Sampling variability. Standard errors 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 standard errors for the most part measure the variations 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 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 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.

Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common types of hypotheses tested are 1) the population parameters are identical or 2) they are different. 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 report 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 report, the estimated absolute difference between parameters is greater than twice the standard error of the difference. For the other differences mentioned, the estimated absolute 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 in the report only when the base is 200,000 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. Also, care must be taken in 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.

Standard error parameters and tables and their use. To derive standard errors that would be applicable to a wide variety of statistics and could be prepared at a moderate cost, a number of approximations were required. Most of the SIPP statistics have greater variance than those obtained through a simple random sample of the same size because clusters of living quarters are sampled for SIPP. Two parameters (denoted "a" and "b") were developed to calculate variances for each type of characteristic.

The "a" and "b" parameters vary by type of estimate and by subgroup to which the estimate applies. Table C-7 provides "a" and "b" parameters for various subgroups and types of estimates. The "a" and "b" parameters may be used to directly calculate the standard errors for estimated numbers and percentages. Because the actual variance behavior was not identical for all statistics within a group, the standard errors computed from either parameters of the tables provide an indication of the order of magnitude of the standard error rather than the precise standard error for any specific statistic.

For those users who wish further simplification, we have also provided general standard errors in tables C-3 through

Table C-2. Distribution of Monthly Household Income Among Persons 25 to 34 Years Old

Income level	Number (thousands)	Cumulative percent distribution
Total	39,851	(X)
Under $300	1,371	100.0
$300 to $599	1,651	96.6
$600 to $899	2,259	92.4
$900 to $1,199	2,734	86.7
$1,200 to $1,499	3,452	79.9
$1,500 to $1,999	6,278	71.2
$2,000 to $2,499	5,799	55.5
$2,500 to $2,999	4,730	40.9
$3,000 to $3,499	3,723	29.1
$3,500 to $3,999	2,519	19.7
$4,000 to $4,999	2,619	13.4
$5,000 to $5,999	1,223	6.8
$6,000 and over	1,493	3.7

X Not applicable.

Table C-3. Standard Errors of Estimated Numbers of Households

(Numbers in thousands)

Size of estimate	Standard error[1]	Size of estimate	Standard error[1]
200	37	10,000	245
300	45	15,000	290
500	58	25,000	348
750	71	30,000	366
1,000	82	40,000	385
2,000	115	50,000	384
3,000	140	60,000	362
5,000	179	70,000	315
7,500	216	80,000	229

[1]These values must be multiplied by the appropriate factor in table C-7 to obtain the correct standard error.

C-6. Note that these standard errors must be adjusted by an "f" factor from table C-7 which is derived from the "b" parameter for the type of estimate and subgroup. The general standard errors are easier to use because there is no need to compute square roots, but they are less accurate. Methods for using these parameters and tables for computation of standard errors are given in the following sections.

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

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

where f is the appropriate "f" factor from table C-7, and s is the standard error on the estimate obtained by interpolation from table C-3 or C-4. Alternatively, it may be approximated by the following formula, (2), from which the standard errors in tables C-3 and C-4 were calculated. Use of this formula will provide more accurate results than the use of formula (1) above.

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

Here x is the size of the estimate and "a" and "b" are the parameters associated with the particular type of characteristic being estimated.

Illustration of the computation of the standard error of an estimated number. Table 1 shows that there were 16,812,000 persons in nonfarm households with a mean monthly household cash income during the fourth quarter of 1984 of $4,000 to $4,999. The appropriate "a" and "b" parameters and "f" factor from table C-7 and the appropriate general standard error from table C-4 are

$$a = -.0000864, \; b = 19{,}911, \; f = 1.00, \; s = 557{,}000$$

Using formula (1), the approximate standard error is

$$S_x = 1.00 \times 557{,}000 = 557{,}000$$

Using formula (2), the approximate standard error is

$$\sqrt{(-.0000864)(16{,}812{,}000)^2 + (19{,}911)(16{,}812{,}000)} \doteq 557{,}000$$

The 68-percent confidence interval as shown by the data is from 16,255,000 to 17,369,000. 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 samples.

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. When the numerator and denominator of the percentage have different parameters, use the parameter (or an appropriate factor) of the numerator. The approximate standard error, S_p, of the estimated percentage P can be obtained by the formula

$$S_p = fs \qquad (3)$$

In this formula, f is the appropriate "f" factor from table C-7 and s is the standard error on the estimate from tables C-5 or C-6. Alternatively, it may be approximated by the following formula, (4), from which the standard errors in tables C-5 and C-6 were calculated. Use of this formula will give more accurate results than use of formula (3) above.

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

Table C-4. Standard Errors of Estimated Numbers of Persons

(Numbers in thousands)

Size of estimate	Standard error[1]	Size of estimate	Standard error[1]
200	63	30,000	721
300	77	50,000	883
600	109	80,000	1,020
1,000	141	100,000	1,062
2,000	199	130,000	1,062
5,000	312	135,000	1,055
8,000	392	150,000	1,021
11,000	457	160,000	987
13,000	494	180,000	886
15,000	528	200,000	725
17,000	560	210,000	609
22,000	629	220,000	446
26,000	678		

[1]These values must be multiplied by the appropriate factor in table C-7 to obtain the correct standard error.

Here x is the size of the subclass of households or persons in households which is the base of the percentage, p is the percentage ($0<p<100$), and b is the parameter associated with the charcteristics in the numerator.

Illustration of the computation of the standard error of an estimated percentage. Continuing the example from above, of the 16,812,000 persons in nonfarm households where the mean monthly household cash income of $4,000 to $4,999, 6.7 percent were Black. Using formula (3) with the "f" factor from table C-7 and the appropriate standard error from table C-6, the approximate standard error is

$$S_p = 0.61 \times 0.8 = 0.5$$

Using formula (4) with the "b" parameter from table C-7, the approximate standard error is

$$S_p = \sqrt{\frac{(7{,}366)}{(16{,}812{,}000)}(6.7)(100 - 6.7)} \doteq 0.5$$

Consequently, the 68-percent confidence interval as shown by these data is from 6.2 to 7.2 percent, and the 95-percent confidence interval is from 5.7 to 7.7 percent.

Standard error of a difference within a quarter. The standard error of a 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, percents, ratios, etc. The above formula assumes that the sample correlation coefficient between the two estimates is zero. If the correlation between the two estimates is really positive (negative), then this assumption will lead to overestimates (underestimates) of the true standard error.

Illustration of the computation of the standard error of a difference within a quarter. Table 1 shows that the number of persons age 35 to 44 years in nonfarm households with mean monthly household cash income of $4,000 to $4,999 during the fourth quarter of 1984 was 3,186,000 and the number of persons age 25 to 34 years in nonfarm households with mean monthly household cash income of $4,000 to $4,999 was 2,619,000. The standard errors of these numbers based on formula (2), are 250,000 and 227,000, respectively. Assuming that these two estimates are not correlated, the standard error of the estimated difference of 567,000 is

$$S_{(x-y)} = \sqrt{(250{,}000)^2 + (227{,}000)^2} \doteq 338{,}000$$

Suppose that it is desired to test at the 5-percent significance level whether the number of persons with mean monthly

Table C-5. Standard Errors of Estimated Percentages of Households

Base of estimated percentage (thousands)	Estimated percentage[1]					
	≤1 or ≥99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	1.8	2.6	4.0	5.5	8.0	9.2
300	1.5	2.1	3.3	4.5	6.5	7.5
500	1.2	1.6	2.5	3.5	5.0	5.8
750	1.0	1.3	2.1	2.8	4.1	4.7
1,000	0.8	1.2	1.8	2.5	3.6	4.1
2,000	0.6	0.8	1.3	1.7	2.5	2.9
3,000	0.5	0.7	1.0	1.4	2.1	2.4
5,000	0.4	0.5	0.8	1.1	1.6	1.8
7,500	0.3	0.4	0.7	0.9	1.3	1.5
10,000	0.26	0.4	0.6	0.8	1.1	1.3
15,000	0.21	0.3	0.5	0.6	0.9	1.1
25,000	0.16	0.2	0.4	0.5	0.7	0.8
30,000	0.15	0.2	0.3	0.5	0.7	0.8
40,000	0.13	0.2	0.3	0.4	0.6	0.7
50,000	0.12	0.16	0.3	0.3	0.5	0.6
60,000	0.11	0.15	0.2	0.3	0.5	0.5
80,000	0.09	0.13	0.2	0.3	0.4	0.5

[1]These values must be multiplied by the appropriate factor in table C-7 to obtain the correct standard error.

household cash income of $4,000 to $4,999 during the fourth quarter of 1984 was different for persons age 35 to 44 years in nonfarm households than for persons age 25 to 34 years in nonfarm households. The 95-percent confidence interval is from -109,000 to 1,243,000 (using two standard errors). Since this interval contains zero, the data do not show that there is a difference between the two age groups at the 5-percent significance level.

Standard error of a mean. A mean is defined here to be the average quantity of some item (other than persons, families, or households) per person, family, or household. For example, it could be the average monthly household income of females age 25 to 34. The standard error of a mean can be approximated by formula (6) below. Because of the approximations used in developing formula (6), an estimate of the standard error of the mean obtained from that formula will

Table C-6. Standard Errors of Estimated Percentages of Persons

Base of estimated percentage (thousands)	Estimated percentage[1]					
	≤1 or ≥99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	3.1	4.4	6.9	9.5	13.7	15.8
300	2.6	3.6	5.6	7.7	11.2	12.9
600	1.8	2.6	4.0	5.5	7.9	9.1
1,000	1.4	2.0	3.1	4.2	6.1	7.1
2,000	1.0	1.4	2.2	3.0	4.3	5.0
5,000	0.6	0.9	1.4	1.9	2.7	3.2
8,000	0.5	0.7	1.1	1.5	2.2	2.5
11,000	0.4	0.6	0.9	1.3	1.8	2.1
13,000	0.4	0.5	0.8	1.2	1.7	2.0
17,000	0.34	0.5	0.7	1.0	1.5	1.7
22,000	0.29	0.4	0.7	0.9	1.3	1.5
26,000	0.28	0.4	0.6	0.8	1.2	1.4
30,000	0.26	0.4	0.6	0.8	1.1	1.3
50,000	0.20	0.3	0.4	0.6	0.9	1.0
80,000	0.16	0.2	0.3	0.5	0.7	0.8
100,000	0.14	0.2	0.3	0.4	0.6	0.7
130,000	0.12	0.17	0.3	0.4	0.5	0.6
220,000	0.10	0.13	0.2	0.3	0.4	0.5

[1]These values must be mulitiplied by the appropriate factor in table C-7 to obtain the correct standard error.

generally underestimate the true standard error. The formula used to estimate the standard error of a mean x is

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

where y is the size of the base, s^2 is the estimated population variance of the item, and b is the parameter associated with the particular type of item.

The estimated population variance, s^2, is given by formula (7):

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

where each sample unit falls in one of c groups; p_i is the estimated proportion of group i; $x_i = (Z_{i-1} + Z_i)/2$ where Z_{i-1} and Z_i are the lower and upper interval boundaries, respectively, for group i. x_i is assumed to be the most representative value for the characteristic of interest in group i. If group c is open-ended, i.e., no upper interval boundary exists, then an approximate average value of x_c is

$$x_c = \frac{3}{2} Z_{c-1}$$

Illustration of the computation of the standard error of an estimated mean. The average monthly household income of persons age 25 to 34 are given by the table C-2. Using formula (7) and the mean monthly household cash income of $2,530, the approximate population variance, s^2, is

$$S^2 = \left(\frac{1,371}{39,851}\right)(150)^2 + \left(\frac{1,651}{39,851}\right)(450)^2 + \ldots.$$

$$\left(\frac{1,493}{39,851}\right)(9,000)^2 - (2,530)^2 = 3,159,887$$

using formula (6) the estimated standard error of a mean x

$$S_{\bar{x}} = \sqrt{\frac{19,911}{39,851,000}(3,159,887)} = \$40$$

Note that the standard error of the mean given in the tables may not agree with those computed using this formula since those in the tables were computed using the raw data and not grouped data.

Standard error of a median. The median quantity of some item such as income for a given group of persons, families, or households is that quantity such that at least half the group have as much or more and at least half the group have as much or less. The sampling variability of an estimated median depends upon the form of the distribution of the item as well as the size of the group. 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.

Table C-7. "a" and "b" Parameters for Direct Computation of Standard Errors of Estimated Numbers and Percentages of Households and Persons: Fourth Quarter 1984

Characteristic	Parameters a	Parameters b	f factor
HOUSEHOLDS			
All races or White	-0.0000764	6,766	1.00
Black	-0.0004661	4,675	0.83
PERSONS			
All Races or White, 16 Years and Over			
Program participation and benefits			
Both sexes	-0.0000943	16,059	0.90
Male	-0.0001984	16,059	0.90
Female	-0.0001796	16,059	0.90
Income and labor force			
Both sexes	-0.0000321	5,475	0.52
Male	-0.0000677	5,475	0.52
Female	-0.0000612	5,475	0.52
All Races or White, All Ages[1]			
Both sexes	-0.0000864	19,911	1.00
Male	-0.0001786	19,911	1.00
Female	-0.0001672	19,911	1.00
Black			
Both sexes	-0.0002670	7,366	0.61
Male	-0.0005737	7,366	0.61
Female	-0.0004993	7,366	0.61

[1]Also use these parameters when age or work disability status are cross-tabulated with person's household income.

1. Determine, using either formula (3) or formula (4), the standard error of an estimate of 50 percent of the group;

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

3. Using the distribution of the item within the group, calculate the quantity of the item such that the percent of the group owning more is equal to the smaller percentage found in step (2). This quantity will be the upper limit for the 68-percent confidence interval. In a similar fashion, calculate the quantity of the item such that the percent

of the group owning more is equal to the larger percentage found in step (2). This quantity will be the lower limit for the 68-percent confidence interval;

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

To perform step (3), it will be necessary to interpolate. Different methods of interpolation may be used. The most common are simple linear interpolation and Pareto interpolation. The appropriateness of the method depends on the form of the distribution around the median. For this report, we recommend Pareto interpolation for any point in a monthly income interval greater than $300, and linear interpolation otherwise.

Interpolation is used as follows. The quantity of the item such that "p" percent own more is

$$X_{pN} = A_1 \exp\left[\operatorname{Ln}\left(\frac{pN}{N_1}\right)\operatorname{Ln}\left(\frac{A_2}{A_1}\right)\Big/\operatorname{Ln}\left(\frac{N_2}{N_1}\right)\right] \tag{8}$$

if Pareto interpolation is indicated and

$$X_{pN} = \frac{N_1 - pN}{N_1 - N_2}(A_2 - A_1) + A_1 \tag{9}$$

if linear interpolation is indicated,

where N = size of the group,

A_1 and A_2 = the quantities of the item which can be easily seen to be the lower and upper bounds, respectively, of the interval in which X_{pN} falls,

N_1 and N_2 = the estimated number of group members owning more of the item than A_1 and A_2, respectively,

exp = refers to the exponential function, and

Ln = refers to the natural logarithm function.

It should be noted that a mathematically equivalent result is obtained by using common logarithms (base 10) and antilogarithms.

Illustration of the computation of a confidence interval and the standard error for a median. To illustrate the calculations for the sampling error on a median, we return to the same example used to illustrate the standard error of a mean. The median monthly income for this group is $2,158. The size of the group is 39,851,000.

1. Using formula (4), the standard error of 50 percent on a base of 39,851,000 is about 1.2 percentage points.

2. Following step (2), the two percentages of interest are 48.8 and 51.2.

3. By examining table C-2, we see that the percentage 48.8 falls in the income interval from $2,000 to $2,499. Thus, A_1 = $2,000, A_2 = $2,500, N_1 = 22,106,000, and N_2 = 16,307,000. Since the width of this interval is greater than $300, Pareto interpolation is used. So the upper bound of a 68-percent confidence interval for the median is

$$(\$2{,}000)\exp\left[\operatorname{Ln}\left(\frac{(.488)(39{,}851{,}000)}{22{,}106{,}000}\right)\operatorname{Ln}\left(\frac{\$2{,}500}{\$2{,}000}\right)\Big/\operatorname{Ln}\left(\frac{16{,}307{,}000}{22{,}106{,}000}\right)\right] \doteq \$2{,}197$$

Also by examining table C-2, we see that the percentage of 51.2 falls in the same income interval. Thus, A_1, A_2, N_1, and N_2 are the same. So the lower bound of a 68-percent confidence interval for the median is

$$(\$2{,}000)\exp\left[\operatorname{Ln}\left(\frac{(.512)(39{,}851{,}000)}{22{,}106{,}000}\right)\operatorname{Ln}\left(\frac{\$2{,}500}{\$2{,}000}\right)\Big/\operatorname{Ln}\left(\frac{16{,}307{,}000}{22{,}106{,}000}\right)\right] \doteq \$2{,}121$$

Thus, the 68-percent confidence interval on the estimated median is from $2,121 to $2,197. An approximate standard error is $\frac{\$2{,}197 - \$2{,}121}{2} = \$38$

Standard errors of ratios of means and medians. The standard error for a ratio of means or medians is approximated by formula (10):

$$S_{\left(\frac{x}{y}\right)} = \sqrt{\left(\frac{x}{y}\right)^2\left[\left(\frac{S_y}{y}\right)^2 + \left(\frac{S_x}{x}\right)^2\right]} \tag{10}$$

where x and y are the means or medians, and S_x and S_y are their associated standard errors. Formula (10) assumes that the means or medians are not correlated. If the correlation between the two means or medians is actually positive (negative), then this procedure will provide an overestimate (underestimate) of the standard error for the ratio of means and medians.