UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

February 14, 2005

DSMD

MEMORANDUM FOR Walter C. Odom, Jr.
Chief, Administrative and Customer Services Division

From: Alan R. Tupek
Chief, Demographic Statistical Methods Division

Subject: Survey of Income and Program Participation (SIPP) 2001 Panel: Source and Accuracy Statement for Wave 1 - Wave 9 Public Use Files (S&A-2)[1]

The attached document is the Source and Accuracy Statement for the SIPP 2001 Panel Wave 1 - Wave 9 Public Use Files.

Attachment

cc:

A. Shields	(ACSD)	C. Nelson	(HHES)
T. Blatt	(DSD)	C. Popoff	
N. McKee		T. Palumbo	
Z. McBride		J. Hisnanick	
J. Eargle		A. Gottschalck	
P. Benton		Q. Wang	
R. Kominski	(POP)	A. Jones Jr.	
M. O'Connell		J. Tin	
K. Bauman		S. Stern	
J. Day		M. Weismantle	
P. Salopek		B. Downs	
J. Fields		L. Cahoon	(DSMD)
		SIPPB	

DSMD/SIPPB/CSISSEL/ S:\SIPPB\FINAL MEMOS\S&A\S&A-2_SIPP2001_W1toW9_20050214.wpd PUBLIC USE FILES(S&A-2)WPD

[1]This source and accuracy statement can also be accessed through the U.S. Census Bureau website at "http://www.sipp.census.gov/sipp/sourceac/S&A-2_SIPP2001_W1toW9_20050214.pdf."

SOURCE AND ACCURACY STATEMENT FOR THE WAVE 1 - WAVE 9 PUBLIC USE FILES FROM THE SURVEY OF INCOME AND PROGRAM PARTICIPATION 2001 PANEL[1]

SOURCE OF DATA

The data were collected in the 2001 panel of the Survey of Income and Program Participation (SIPP). The population represented (the population universe) in the 2001 SIPP is the civilian noninstitutionalized population living in the United States. The institutionalized population, which is excluded from the population universe, is composed primarily of the population in correctional institutions and nursing homes (91 percent of the 4.1 million institutionalized people in Census 2000). The 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. Also, United States citizens residing abroad were not eligible to be in the survey. Foreign visitors who work or attend school in this country and their families were eligible; all others were not eligible to be in the survey. With the exceptions noted above, persons who were at least 15 years of age at the time of the interview were eligible to be in the survey.

The 2001 panel of the SIPP sample is located in 322 Primary Sampling Units (PSUs), each consisting of a county or a group of contiguous counties. Within these PSUs, living quarters (LQs) were systematically selected from lists of addresses prepared for the 1990 decennial census to form the bulk of the sample. To account for LQs built within each of the sample areas after the 1990 census, a sample containing clusters of four LQs was drawn of permits issued for construction of residential LQs up until shortly before the beginning of the panel.

In jurisdictions that do not issue building permits or have incomplete addresses, we systematically sampled expected clusters of four LQs which were listed by field personnel and then subsampled in the field. In addition, we selected sample LQs from a supplemental frame that included LQs identified as missed in the 1990 census.

Sample households within a given panel are divided into four random subsamples of nearly equal size. These subsamples are called rotation groups and one rotation group is interviewed each month. Each household in the sample was scheduled to be interviewed at 4 month intervals over a period of roughly 3 years beginning in February 2001. The reference period for the questions is the 4-month period preceding the interview month. In general, one cycle of four interviews covering the entire sample, using the same questionnaire, is called a wave.

---

[1]For questions or further assistance with the information provided in this document contact: Tracy Mattingly of the Demographic Statistical Methods Division on 301/763-6445 or via the email at Tracy.L.Mattingly@census.gov.

In Wave 1, we fielded a sample consisting of 88 reduction groups (88 comparable representative subsamples) which resulted in an average sampling interval of approximately 2,420 housing units. In this wave, we obtained interviews from occupants of about 35,100 of the 40,500 eligible living quarters. We found most of the remaining 15,400 living quarters in the panel to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. However, we did not interview approximately 5,400 of the 15,400 living quarters in the panel because the occupants, (1) refused to be interviewed, (2) could not be found at home, (3) were temporarily absent, or (4) were otherwise unavailable. Thus, occupants of about 87 percent of all eligible living quarters participated in the first interview of the panel.

Due to budget constraint, we cut the sample in Wave 2 by 13 reduction groups which resulted in an average sampling interval of approximately 2,840 housing units. We did not cut the sample in the remaining waves (Wave 3 to Wave 9). For interviews in Wave 2 to Wave 9, only original sample persons (those in Wave 1 sample households which survived the sample cut in Wave 2 and interviewed in Wave 1) and persons living with them were eligible to be interviewed. We followed original sample persons if they moved to a new address, unless the new address was more than 100 miles from a SIPP sample area. Then, we attempted telephone interviews. Based on these follow-up criteria, we interviewed about 28,100 living quarters of the approximately 30,500 eligible living quarters for Wave 2, about 27,500 living quarters of the approximately 30,900 eligible living quarters for Wave 3, about 27,200 living quarters of the approximately 31,100 eligible living quarters for Wave 4, about 26,800 living quarters of the approximately 31,300 eligible living quarters for Wave 5, about 26,600 living quarters of the approximately 31,400 eligible living quarters for Wave 6, about 26,500 living quarters of the approximately 31,500 eligible living quarters for Wave 7, about 26,000 living quarters of the approximately 31,600 eligible living quarters for Wave 8, about 25,500 living quarters of the approximately 31,700 eligible living quarters for Wave 9. In each of these waves, we did not interview some of the eligible living quarters because the occupants either directly or indirectly refused our interview in the same manner described for Wave 1 or moved to an unknown address. The rates of non-interviewed living quarters due to direct or indirect refusal were 6.2% for Wave 2, 8.4% for Wave 3, 9.5% for Wave 4, 10.9% for Wave 5, 11.6% for Wave 6, 12.3% for Wave 7, 13.3% for Wave 8, and 14.7% for Wave 9. The rates of non-interviewed living quarters due to moving to an unknown address were 1.7% for Wave 2, 2.7% for Wave 3, 3.2% for Wave 4, 3.6% for Wave 5, 3.7% for Wave 6, 3.8% for Wave 7, 4.5% for Wave 8, and 4.8% for Wave 9.

The public use files include core and supplemental (topical module) data. Core questions are repeated at each interview over the life of the panel. Topical modules include questions which are asked only in certain waves. The 2001 panel topical modules are given in Table 1.

Table 2 indicates the reference months and interview months for the collection of data from each rotation group for the 2001 panel. For example, Wave 1 rotation group 1 of the 2001 panel was interviewed in February 2001 and data for the reference months October 2000 through January 2001 were collected.

This source and accuracy statement can also be accessed through the U.S. Census Bureau website at "http://www.sipp.census.gov/sipp/sourceac/S&A01_w1tow9_cross_puf.pdf."

Estimation. We used several stages of weight adjustments in the estimation procedure to derive the SIPP cross-sectional person level weights. We gave each person a base weight **(BW)** equal to the inverse of probability of selection of a person's household. We applied two noninterview adjustment factors. One factor adjusted the weights of interviewed persons in interviewed households to account for households which were eligible for the sample but which field representatives could not interview at the first interview (F_{N1}). The second factor compensated for person noninterviews occurring in subsequent interviews (F_{N2}). We used a Duplication Control Factor **(DCF)** which adjusts for subsampling done in the field when the number of sample units is much larger than expected. We applied a Mover's Weight **(MW)**, which adjusts for persons in the SIPP universe who move into sample households after Wave 1. The last factor applied is the Second Stage Adjustment Factor (F_{2s}). This factor adjusts estimates to population controls and causes husbands' and wives' weights to be equal. See the next section on population controls for more information on how they are obtained.

Population Controls. This survey's estimation procedure adjusts weighted sample results to agree with independently derived population estimates of the civilian noninstitutional population of the United States. We control to independent population estimates in an attempt to reduce our mean square error by partially correcting for undercoverage. To obtain the controls, we take the CPS weights and do a "March type" family equalization. That is, we assign wives' weights to husbands and then proportionally adjust the weights of persons by month, rotation group, race, sex, age, and by the marital and family status of householders. Using these weights with CPS data, the controls for SIPP are obtained. These are prepared annually to agree with the most current set of population estimates that are released as part of the Census Bureau's population estimates and projections program.

The population controls for the nation are distributed by demographic characteristics in two ways:

- age, sex, and race (Non Black, Black) and
- age, sex, and Hispanic origin.

The estimates begin with the latest decennial census as the base and incorporate the latest available information on births and deaths along with the latest estimates of net international migration.

The net international migration component in the population estimates includes a combination of:

- legal migration to the U.S.,
- emigration of foreign born and native people from the U.S.,
- net movement between the U.S. and Puerto Rico,
- estimates of temporary migration, and
- estimates of net residual foreign-born population, which include unauthorized migration.

Because the latest available information on these components lag the survey date, to develop the estimate for the survey date, it is necessary to make short-term projections of these components. The final cross-sectional weight is $\mathbf{Fw_c = BW \times DCF \times F_{n1} \times F_{2S}}$ for Wave 1 and is $\mathbf{Fw_c = IW \times F_{n2} \times F_{2S}}$ for Waves 2+ , where **IW** is either $\mathbf{BW \times DCF \times F_{n1}}$ or **MW**. James (1995) and Siegel (1995a) describe SIPP cross-sectional weighting in greater detail.

Researchers both inside and outside the Census Bureau conducted evaluations of SIPP weighting methodology and researched alternative methodologies. Several improvements to SIPP weighting methods were implemented beginning with the 1996 panel. They are described below.

- We dropped the first stage factor (F_{1s}) from cross-sectional weighting. This factor adjusted for differences between the Census count of population and an estimate of that count based on Census data for sample PSUs. James (1994) found that it did not reduce variance as was previously believed. Jabine, et al (1990) describe the first stage factor used in earlier panels.

- We are using additional variables in nonresponse adjustment. We added high/low poverty stratum code to the Wave 1 nonresponse adjustment, and we added household income, geographic division, and number of imputations for selected income and asset items to the nonresponse adjustment for Waves 2+. Research by Rizzo, et al (1994) and by Folsom and Witt (1994) pointed out the potential of the latter three variables in reducing nonresponse bias.

- We redefined nonresponse adjustment cells for Waves 2+ weighting. We formed the nonresponse cells by successively partitioning data from five panels by whichever variable most reduced the bias of the household income to poverty threshold ratio. We used data from a sixth panel to evaluate the results. We calculated the nonresponse bias of six variables at Waves 2 and 7 for both the new cells and the original cells using initial weights and data from the most recent interview in the calculations. The new cells had lower bias for five of the six variables (Siegel, 1995b).

Research was conducted on a number of promising weighting improvements. Allen and Petroni (1994) reported on an adjustment for mover attrition. Folsom and Witt (1994) and Rizzo, et al (1994) studied alternative nonresponse adjustments using response propensity models. Each study computed weights using an alternative methodology. The researchers then compared estimates of various items to benchmarks. The benchmarks came from administrative records and survey data with less nonresponse than the SIPP. The comparisons did not provide strong evidence of lower bias using the alternative weighting methods.

Additional Methodology

Use of Weights. Each household and each person within each household, on each core wave file has four weights. These four weights are reference month specific and therefore can be used only to form reference month estimates. Reference month estimates can be averaged to form estimates of monthly averages over some period of time.

> **Example**, using the proper weights, one can estimate the monthly average number of households in a specified income range over November and December 2001. To estimate monthly averages of a given measure (such as, total, mean) over a number of consecutive months, sum the monthly estimates and divide by the number of months.

To form an estimate for a particular month, use the reference month weight for the month of interest, summing over all persons or households with the characteristic of interest whose reference period includes the month of interest. Multiply the sum by a factor to account for the number of rotations contributing data for the month. This factor equals four divided by the number of rotations contributing data for the month. For example, December 2000 data is only available from rotations 1, 2, and 3 for Wave 1 of the 2001 panel (See Table 2), so a factor of 4/3 must be applied.

When estimates for months with less than four rotations worth of data are constructed from a wave file, factors greater than 1 must be applied, as above. However, when core data from consecutive waves are used together, data from all four rotations may be available, in which case the factors are equal to 1.

These core wave files contain no weight for characteristics that involve a persons's or household's status over two or more months (such as, number of households with a 50 percent increase in income between December 2000 and January 2001).

Producing Estimates for Census Regions and States. The total estimate for a region is the sum of the state estimates in that region. Using this sample, estimates for individual states are subject to very high variance and may not be state representative due to the nature of the sample design. Therefore, estimates for individual states are not recommended. The state codes on the file are primarily of use in linking respondent characteristics with appropriate contextual variables (for example, state-specific welfare criteria) and for tabulating data by user-defined groupings of states.

ESTIMATES

SIPP estimates 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. Found in the next sections are descriptions of sources of SIPP nonsampling error, followed by a discussion of sampling error, its estimation, and its effect in data analyses.

Nonsampling Error. Nonsampling errors can be attributed to many sources:

- 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 the following: collection such as in recording or coding the data, processing the data, estimating values for missing data
- biases resulting from the differing recall periods caused by the interviewing pattern used
- and undercoverage.

Quality control and edit procedures were used to reduce errors made by respondents, coders and interviewers. More detailed discussions of the existence and control of nonsampling errors in the SIPP can be found in the *SIPP Quality Profile, 1998 SIPP Working Paper Number 230, issued May 1999*.

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 (second stage weight adjustment) 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 characteristics different from those of interviewed persons in the same age-race-sex group. Further, the independent population controls used have been adjusted for undercoverage in the Census.

A common measure of survey coverage is the coverage ratio, the estimated population before ratio adjustment divided by the independent population control. The Table below shows SIPP coverage ratios for age-sex-race groups for one month-February 2001 prior to the weighting adjustment. The SIPP coverage ratios exhibit some variability from month to month, but these are a typical set of coverage ratios. Other Census Bureau household surveys (like the Current Population Survey) experience similar coverage.

Comparability with Other Estimates. Caution should be exercised when comparing data from this with data from other SIPP products or with data from other surveys. The comparability problems are caused by such sources as the seasonal patterns for many characteristics, different nonsampling errors, and different concepts and procedures. Refer to the *SIPP Quality Profile* for known differences with data from other sources and further discussions.

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.

SIPP Coverage Ratios for February 2001
Age by Non-Black/Black Status and Sex

	Non-Black		Black	
Age	**M**	**F**	**M**	**F**
15	0.9175	1.1235	0.7044	0.7749
16-17	0.8640	0.9289	0.8826	0.9433
18-19	0.8620	0.8647	0.8274	0.8339
20-21	0.8848	0.8041	0.6255	0.9596
22-24	0.7859	0.8692	0.5857	0.6705
25-29	0.8022	0.8254	0.8504	0.8386
30-34	0.8721	0.9063	0.8792	0.7991
35-39	0.9212	0.9855	0.7119	0.8982
40-44	0.9058	0.9321	0.8059	0.9653
45-49	0.9009	0.9761	0.6856	0.7758
50-54	0.9667	0.9181	0.8993	1.2103
60-61	0.8405	0.8961	1.0210	0.9877
62-64	0.9866	1.0698	0.9914	0.9618
65-69	0.9304	0.9423	1.0646	0.7759
70-74	0.8836	0.9362	0.7896	1.3338
75-79	0.8952	1.0046	--------	0.9104
80-84	0.8974	0.9651	--------	--------
85+	0.9558	0.9669	--------	--------

USES AND COMPUTATION OF STANDARD ERRORS

Confidence Intervals. 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.

Hypothesis Testing. Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population characteristics using sample estimates. The most common types of hypotheses tested are 1) the population characteristics are identical versus 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 characteristics are different when, in fact, they are identical.

To perform the most common test, compute the difference $X_A - X_B$, where X_A and X_B are sample estimates of the characteristics of interest. A later section explains how to derive an estimate of the standard error of the difference $X_A - X_B$. Let that standard error be S_{DIFF}. If $X_A - X_B$ is between -1.6 times S_{DIFF} and +1.6 times S_{DIFF}, no conclusion about the characteristics is justified at the 10 percent significance level. If, on the other hand, $X_A - X_B$ is smaller than -1.6 times S_{DIFF} or larger than +1.6 times S_{DIFF}, the observed difference is significant at the 10 percent level. In this event, it is commonly accepted practice to say that the characteristics are different. Of course, sometimes this conclusion will be wrong. When the characteristics are the same, there is a 10 percent chance of concluding that they are different.

Note that as more tests are performed, more erroneous significant differences will occur. For example, at the 10 percent significance level, if 100 independent hypothesis tests are performed in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, the significance of any single test should be interpreted cautiously.

Note Concerning Small Estimates and Small Differences. Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a base smaller than 200,000. Care must be taken in the interpretation of small differences since even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Calculating Standard Errors for SIPP Estimates. There are three main ways we calculate the Standard Errors for SIPP Estimates. They are as follows:

- Replicate Weighting Methods,
- Generalized Variance parameters (denoted as *a* and *b*),
- Simplified tables using the *a* and *b* parameters.

  SIPP uses the Replicate Weighting Method to produce Generalized Variance parameters. Using the Generalized Variance parameters, we create simplified tables.

Standard Error Parameters and Tables and Their Use. Most SIPP estimates have greater standard errors than those obtained through a simple random sample because PSUs are sampled and clusters of living quarters are sampled for the SIPP in the area and new construction frames. To derive standard errors that would be applicable to a wide variety of estimates and could be prepared at a moderate cost, a number of approximations were required. Estimates with similar standard error behavior were grouped together by characteristics at the person level and characteristics of households (including unrelated persons). Two parameters (denoted *a* and *b*) were computed for each characteristic in order to approximate the standard error behavior. These *a* and *b* parameters vary according to wave and characteristic as well as the demographic subgroup of the group to which the estimate applies. Because the actual standard error behavior was not identical for all characteristics and groups, the standard errors computed using these parameters provide an indication of the order of magnitude of the standard error estimate for a specific group. Table 3 provides tables of base *a* and *b* parameters by wave to be used for the 2001 panel estimates. There are four sets of parameters in Table 3: the first set of parameters per item is given to be used for calculations based on persons or households interviewed during Wave 1 the second set is for Waves 2 and 3, the third set is for Wave 4 to Wave 6, and the fourth set is for Wave 7 to Wave 9. Table 9 provides the base generalized variance a and b parameters for calculating 2001 topical module variances.

Table 2 lists the reference months for each interview month. Use Table 4 (if needed) to select the adjustment factor appropriate to the wave. Multiply this factor by the *a* and *b* base parameters of Table 3 to produce *a* and *b* parameters for the variance estimate for a specific subgroup and reference period. For example, the base *a* and *b* parameters for total number of households are -0.00003286 and 3546, respectively. Using Table 4 for Wave 1, the factor for November 2000 is 2 *since only 2 rotation months of data are available*. So the *a* and *b* parameters for the variance estimate of a white household characteristic in November 2000 based on Wave 1 are $-0.00003286 \times 2 = -0.00006572$ and $3546 \times 2 = 7{,}092$, respectively.

Similarly, the factor for the last quarter of 2000 is 1.8519 (Table 4) since the only data available are the 6 rotation months from Wave 1 (namely, as indicated in Table 2, rotation 1 provides three rotation months, rotation 2 provides two rotation months, and rotation 3 provides one rotation month of data.) So the *a* and *b* parameters for the variance estimate of a white household characteristic in the last quarter of 2000 are $-0.00003286 \times 1.8519 = -0.00006085$ and $3546 \times 1.8519 = 6{,}567$, respectively.

The *a* and *b* parameters may be used to calculate the standard error for estimated numbers and percentages. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. Methods for using these parameters for computation of approximate standard errors are given in the following sections.

For those users who wish further simplification, we have also provided base standard errors for estimates of total and estimates of percentages in Tables 5 through 8. Note that these base standard errors only apply when data from all four rotations are used and must be adjusted by an f factor provided in Table 3. The standard errors resulting from this simplified approach are less accurate. Methods for using these parameters and tables for computation of standard errors are given in the following sections.

The procedures described below apply only to reference month estimates or averages of reference month estimates. Refer to the section "Use of Weights" for a more detailed discussion of the construction of estimates.

Variance stratum codes and half sample codes are included on the tapes (data sets) to enable the user to compute the variances directly and more accurately by methods such as balanced repeated replications (BRR). William G. Cochran provides a list of references discussing the application of this technique. (See Sampling Techniques, 3rd Ed., New York: John Wiley and Sons, 1977, p. 321.)

Standard Errors of Estimated Numbers. The approximate standard error, s_x, of an estimated number of persons, households, families, unrelated individuals and so forth, can be obtained in two ways. Both apply when data from all four rotations are used to make the estimate. However, only the second method (formula 2) should be used when less than four rotations of data are available for the estimate. Note that neither method should be applied to dollar values.

The standard error may be obtained by the use of the formula

$$s_x = fs \tag{1}$$

where f is the appropriate f factor from Table 3, and s is the base standard error on the estimate obtained by interpolation from Table 5 or 6. Alternatively, s_x may be approximated by the formula

$$s_x = \sqrt{ax^2 + bx} \tag{2}$$

from which the base standard errors in Tables 7 and 8 were calculated. Here x is the size of the estimate and a and b are the parameters from Table 4 which are associated with the characteristic being estimated (and the wave which applies). Use of formula 2 will generally provide more accurate results than the use of formula 1.

Illustration.

Suppose SIPP estimates based on Wave 1 of the 2001 panel show that there were 1,700,000 black households with monthly household income above $4,000 in January 2001. The appropriate parameters and factor from Table 3 and the appropriate general standard error from Table 5 are

$$a = -0.00019168 \qquad b = 2{,}495 \qquad f = 0.84 \qquad s = 76{,}800$$

Using formula 1, the approximate standard error is

$$s_x = (0.84)(76{,}800) = 64{,}512$$

Using formula 2, the approximate standard error is

$$\sqrt{(-0.00019168)(1{,}700{,}000)^2 + (2{,}495)(1{,}700{,}000)} = 60{,}725$$

Using the standard error based on formula 2, the approximate 90-percent confidence interval as shown by the data is from 1,600,107 to 1,799,893. 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 90% of all samples.

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 3 below. Because of the approximations used in developing formula 3, an estimate of the standard error of the mean obtained from this formula will generally underestimate the true standard error. The formula used to estimate the standard error of a mean $\overline{x}$ is

$$s_{\bar{x}} = \sqrt{\left(\frac{b}{y}\right) s^2} \qquad (3)$$

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 population variance s^2 may be estimated by one of two methods. In both methods, we assume x_i is the value of the item for unit "i." (Unit may be person, family, or household). To use the first method, the range of values for the item is divided into "c" intervals. The upper and lower boundaries of interval j are Z_{j-1} and Z_j, respectively. Each unit is placed into one of "c" groups such that $Z_{j-1} < x_i \leq Z_j$.

The estimated population variance, s^2, is given by the formula:

$$s^2 = \sum_{j=1}^{c} p_j m_j^2 - \bar{x}^2, \qquad (4)$$

where p_j is the estimated proportion of units in group j, and $m_j = (Z_{j-1} + Z_j) / 2$. The most representative value of the item in group j is assumed to be m_j. If group "c" is open-ended, or there is no upper interval boundary exists, then an approximate value for m_c is

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

The mean, $\bar{x}$ can be obtained using the following formula:

$$\bar{x} = \sum_{j=1}^{c} p_j m_j$$

In the second method, the estimated population mean, $\bar{x}$, and variance, s^2 are given by

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

$$s^2 = \frac{\sum_{i=1}^{n} w_i x_i^2}{\sum_{i=1}^{n} w_i} - \bar{x}^2, \tag{5}$$

where there are n units with the item of interest and w_i is the final weight for unit "I". (Note that $\sum w_i = y$ in formula 3.)

Illustration.

Suppose that based on Wave 1 data, the distribution of monthly cash income for persons age 25 to 34 during the month of January 2001 is given in Table 10.

Using formula 4 and the mean monthly cash income of $2,530 the approximate population variance, s^2, is

$$s^2 = \left(\frac{1,371}{39,851}\right)(150)^2 + \left(\frac{1,651}{39,851}\right)(450)^2 + \ldots\ldots + \left(\frac{1,493}{39,851}\right)(9,000)^2 - (2,530)^2 = 3,159,887.$$

Using formula 3 and the appropriate base b parameter from Table 3, the estimated standard error of a mean $\bar{x}$ is

$$s_{\bar{x}} = \sqrt{\left(\frac{4,263}{39,851,000}\right)(3,159,887)} = \$18.39$$

Standard error of an aggregate. An aggregate is defined to be the total quantity of an item summed over all the units in a group. The standard error of an aggregate can be approximated using formula 6.

As with the estimate of the standard error of a mean, the estimate of the standard error of an aggregate will generally underestimate the true standard error. Let y be the size of the base, s^2 be the estimated population variance of the item obtained using formula (4) or (5) and b be the parameter associated with the particular type of item. The standard error of an aggregate is

$$s_x = \sqrt{(b)\ (y) s^2} \tag{6}$$

Standard Errors of Estimated Percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more, e.g., the percent of people employed is more reliable than the estimated number of people employed. When the numerator and denominator of the percentage have different parameters, use the parameter (and appropriate factor) of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

There are two types of percentages commonly estimated. The first is the percentage of persons, families or households sharing a particular characteristic such as the percent of persons owning their own home. The second type is the percentage of money or some similar concept held by a particular group of persons or held in a particular form. Examples are the percent of total wealth held by persons with high income and the percent of total income received by persons on welfare.

For the percentage of persons, families, or households, the approximate standard error, $s_{(x,p)}$, of the estimated percentage p can be obtained by the formula

$$s_{(x,p)} = fs \tag{7}$$

when data from all four rotations are used to estimate p.

In this formula, f is the appropriate f factor from Table 3 (for the appropriate wave) and s is the base standard error of the estimate from Table 7 or 8.

Alternatively, it may be approximated by the formula

$$s_{(x,p)} = \sqrt{\frac{b}{x}\,(p)\,(100-p)} \qquad (8)$$

from which the standard errors in Tables 7 and 8 were calculated. Here x is the size of the subclass of social units which is the base of the percentage, p is the percentage ($0 < p < 100$), and b is the parameter associated with the characteristic in the numerator. Use of this formula will give more accurate results than use of formula 7 above and should be used when data from less than four rotations are used to estimate p.

Illustration.

Suppose that, in the month of January 2001, 6.7 percent of the 16,812,000 persons in nonfarm households with a mean monthly household cash income of $4,000 to $4,999, were black. Using formula 8 and the b parameter of 4,475 from Table 3 and a factor of 1 for the month of January 2001 from Table 4, the approximate standard error is

$$\sqrt{\frac{4{,}475}{(16{,}812{,}000)}\,(6.7)\,(100-6.7)} = 0.41 \text{ percent}$$

Consequently, the 90 percent confidence interval as shown by these data is from 6.03 to 7.37 percent.

For percentages of money, a more complicated formula is required. A percentage of money will usually be estimated in one of two ways. It may be the ratio of two aggregates:

$$p_I = 100\ (X_A \ / \ X_N)$$

or it may be the ratio of two means with an adjustment for different bases:

$$p_I = 100\ (\hat{p}_A \, \bar{X}_A \ / \ \bar{X}_N)$$

where x_A and x_N are aggregate money figures, $\overline{x}_A$ and $\overline{x}_N$ are mean money figures, and $\hat{p}_A$ is the estimated number in group A divided by the estimated number in group *N*. In either case, we estimate the standard error as

$$s_I = \sqrt{\left(\frac{\hat{p}_A \bar{x}_A}{\bar{x}_N}\right)^2 \left[\left(\frac{s_p}{\hat{p}_A}\right)^2 + \left(\frac{s_A}{\bar{x}_A}\right)^2 + \left(\frac{s_B}{\bar{x}_N}\right)^2\right]} \quad , \tag{9}$$

where s_p is the standard error of $\hat{p}_A$, s_A is the standard error of $\bar{x}_A$ and s_B is the standard error of $\bar{x}_N$. To calculate s_p, use formula 8. The standard errors of $\bar{x}_N$ and $\bar{x}_A$ may be calculated using formula 3.

It should be noted that there is frequently some correlation between $\hat{p}_A$, $\bar{x}_N$, and $\bar{x}_A$. Depending on the magnitude and sign of the correlations, the standard error will be over or underestimated.

Illustration.

Suppose that in January 2001, 9.8% of the households own rental property, the mean value of rental property is \$72,121, the mean value of assets is \$78,734, and the corresponding standard errors are 0.19 %, \$5799, and \$2867, respectively. In total there are 86,790,000 households. Then, the percent of all household assets held in rental property is

$$= 100 \left((0.098) \frac{72121}{78734} \right) = 9.0\%$$

Using formula (9), the appropriate standard error is

$$s_I = \sqrt{\left(\frac{(0.098)(72121)}{78734}\right)^2 \left[\left(\frac{0.0019}{0.098}\right)^2 + \left(\frac{5799}{72121}\right)^2 + \left(\frac{2867}{78734}\right)^2\right]}$$

= 0.008 = 0.8%

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

$$s_{(x-y)} = \sqrt{s_x^2 + s_y^2} \tag{10}$$

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 correlation coefficient between the characteristics estimated by x and y is zero. If the correlation is really positive (negative), then this assumption will tend to cause overestimates (underestimates) of the true standard error.

Illustration.

Suppose that SIPP estimates show the number of persons age 35-44 years with monthly cash income of $4,000 to $4,999 was 3,186,000 in the month of January 2001 and the number of persons age 25-34 years with monthly cash income of $4,000 to $4,999 in the same time period was 2,619,000. Then, using parameters from Table 3 and formula 2, the standard errors of these numbers are approximately 115,689 and 105,029, respectively. The difference in sample estimates is 567,000 and using formula 10, the approximate standard error of the difference is

$$\sqrt{(115{,}689)^2 + (105{,}029)^2} = 156{,}253$$

Suppose that it is desired to test at the 10 percent significance level whether the number of persons with monthly cash income of $4,000 to $4,999 was different for persons age 35-44 years than for persons age 25-34 years. To perform the test, compare the difference of 567,000 to the product $1.645 \times 156{,}253 = 257{,}036$. Since the difference is greater than 1.645 times the standard error of the difference, the data show that the two age groups are significantly different at the 10 percent significance level.

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. To calculate standard errors on medians, the procedure described below may be used.

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 either formula 7 or formula 8, 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 with more of the item 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 with more of the item 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. If density is declining in the area, then we recommend Pareto interpolation. If density is fairly constant in the area, then we recommend linear interpolation. Note, however, that Pareto interpolation can never be used if the interval contains zero or negative measures of the item of interest. Interpolation is used as follows. The quantity of the item such that *p* percent have more of the item is

$$X_{pN} = \exp\left[\left(\mathrm{Ln}\left(\frac{pN}{N_1}\right) / \mathrm{Ln}\left(\frac{N_2}{N_1}\right)\right) \mathrm{Ln}\left(\frac{A_2}{A_1}\right)\right] A_1 \tag{11}$$

if Pareto Interpolation is indicated and

$$X_{pN} = \left[\frac{PN - N_1}{N_2 - N_1} (A_2 - A_1) + A_1\right] \tag{12}$$

if linear interpolation is indicated, where

N	is the size of the group,
A_1 and A_2	are the lower and upper bounds, respectively, of the interval in which X_{pN} falls
N_1 and N_2	are the estimated number of group members owning more than A_1 and A_2, respectively
exp	refers to the exponential function and
Ln	refers to the natural logarithm function

<u>Illustration</u>.

To illustrate the calculations for the sampling error on a median, we return to Table 10, and suppose that the income tabulated for this group is for January 2001. The median monthly income for this group is \$2,158 in January 2001. The size of the group is 39,851,000.

1. Using formula 8 (with b = 4,263 for Wave 1), the standard error of 50 percent on a base of 39,851,000 is about 0.5 percentage points.

2. Following step 2, the two percentages of interest are 49.5 and 50.5.

3. By examining Table 10, we see that the percentage 49.5 falls in the income interval from 2000 to 2499. (Since 55.5% receive more than \$2,000 per month, the dollar value corresponding to 49.5 must be between \$2,000 and \$2,500). Thus, A_1 = \$2,000, A_2 = \$2,500, N_1 = 22,106,000, and N_2 = 16,307,000.

In this case, we decided to use Pareto interpolation. Therefore, the upper bound of a 68% confidence interval for the median is

$$\$2{,}000 \exp\left[\left(\mathrm{Ln}\left(\frac{(.495)(39{,}851{,}000)}{22{,}106{,}000}\right) \Big/ \mathrm{Ln}\left(\frac{16{,}307{,}000}{22{,}106{,}000}\right)\right)\mathrm{Ln}\left(\frac{2{,}500}{2{,}000}\right)\right] = \$2174$$

Also by examining Table 10, we see that 50.5 falls in the same income interval. Thus, A_1, A_2, N_1 and N_2 are the same. We also use Pareto interpolation for this case. So the lower bound of a 68% confidence interval for the median is

$$\$2{,}000 \exp\left[\left(\mathrm{Ln}\left(\frac{(.505)(39{,}851{,}000)}{22{,}106{,}000}\right) \Big/ \mathrm{Ln}\left(\frac{16{,}307{,}000}{22{,}106{,}000}\right)\right)\mathrm{Ln}\left(\frac{2{,}500}{2{,}000}\right)\right] = \$2142$$

Thus, the 68-percent confidence interval on the estimated median is from \$2142 to \$2174. An approximate standard error is

$$\frac{\$2174 - \$2142}{2} = \$16$$

Standard Errors of Ratios of Means and Medians. The standard error for a ratio of means or medians is approximated by:

$$s_{\frac{x}{y}} = \sqrt{\left(\frac{x}{y}\right)^2 \left[\left(\frac{s_y}{y}\right)^2 + \left(\frac{s_x}{x}\right)^2\right]} \qquad (13)$$

where x and y are the means or medians, and s_x and s_y are their associated standard errors. Formula 13 assumes that the means are not correlated. If the correlation between the population means estimated by x and y are actually positive (negative), then this procedure will tend to produce overestimates (underestimates) of the true standard error for the ratio of means.

Standard Errors Using SAS or SPSS. Standard errors and their associated variance, calculated by SAS or SPSS statistical software package, do not accurately reflect the SIPP's complex sample design. Erroneous conclusions will result if these standard errors are used directly. We provide adjustment factors by characteristics that should be used to correctly compensate for likely under-estimates. The factors called DEFF available in Table 3, must be applied to SAS or SPSS generated variances. The square root of DEFF can be directly applied to similarly generated standard errors. These factors approximate design effects which adjust statistical measures for sample designs more complex than simple random sample.

Table 1 - 2001 Panel Topical Modules

W 1	‣ Recipiency History ‣ Employment History	W6	‣ Assets, Liabilities, Eligibility ‣ Medical Expenses/Health Care Usage ‣ Work-related Expenses ‣ Child Support Paid ‣ Child Care Poverty
W 2	‣ Work Disability ‣ Education & Training History ‣ Marital History ‣ Migration History ‣ Fertility ‣ Household Relationships	W7	‣ Annual Income & Retirement Accounts ‣ Taxes ‣ Retirement & Pension Plan ‣ Home Health Care ‣ Child Well-Being
W 3	‣ Assets, Liabilities, Eligibility ‣ Medical Expenses/Health Care Usage ‣ Work-related Expenses ‣ Child Support Paid ‣ Child Care Poverty	W8	‣ Adult Well-Being ‣ Child Support Agreements ‣ Support for Non-household members ‣ Functional Limitations/Disabilities-Adult ‣ Functional Limitations/Disabilities-Child ‣ Welfare Reform
W 4	‣ Annual Income & Retirement Accounts ‣ Taxes ‣ Work Schedule ‣ Child Care	W9	‣ Assets, Liabilities, Eligibility ‣ Medical Expenses/Health Care Usage ‣ Work-related Expenses ‣ Child Support Paid ‣ Child Care Poverty
W 5	‣ School Enrollment & Financing ‣ Child Support Agreements ‣ Support for Non-household members ‣ Functional Limitations/Disabilities-Adult ‣ Functional Limitations/Disabilities-Child ‣ Employer-Provided Health Benefits		

Table 2 - SIPP Panel 2001 Reference Months (horizontal) for Each Interview Month (vertical)

		2000			2001												2002												2003											
		4th Quarter			1St Quarter			2nd Quarter			3rd Quarter			4th Quarter			1St Quarter			2nd Quarter			3rd Quarter			4th Quarter			1St Quarter			2nd Quarter			3rd Quarter			4th Quarter		
	Month of Wave/Rotation	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Spt	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Spt	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Spt	Oct	Nov	Dec
Feb 01	1/1	1	2	3	4																																			
Mar	1/2		1	2	3	4																																		
Apr	1/3			1	2	3	4																																	
May	1/4				1	2	3	4																																
Jun	2/1					1	2	3	4																															
July	2/2						1	2	3	4																														
Aug	2/3							1	2	3	4																													
Sept	2/4								1	2	3	4																												
Oct	3/1									1	2	3	4																											
Nov	3/2										1	2	3	4																										
Dec	3/3											1	2	3	4																									
Jan 02	3/4												1	2	3	4																								
Feb	4/1													1	2	3	4																							
Mar	4/2														1	2	3	4																						
Apr	4/3															1	2	3	4																					
May	4/4																1	2	3	4																				
Jun	5/1																	1	2	3	4																			
July	5/2																		1	2	3	4																		
Aug	5/3																			1	2	3	4																	
Sept	5/4																				1	2	3	4																
Oct	6/1																					1	2	3	4															
Nov	6/2																						1	2	3	4														
Dec	6/3																							1	2	3	4													
Jan 03	6/4																								1	2	3	4												
Feb	7/1																									1	2	3	4											
Mar	7/2																										1	2	3	4										
Apr	7/3																											1	2	3	4									
May	7/4																												1	2	3	4								
Jun	8/1																													1	2	3	4							
July	8/2																														1	2	3	4						
Aug	8/3																															1	2	3	4					
Sep	8/4																																1	2	3	4				
Oct	9/1																																	1	2	3	4			
Nov	9/2																																		1	2	3	4		
Dec	9/3																																			1	2	3	4	
Jan 04	9/4																																				1	2	3	4

Table 3[2] - SIPP Panel 2001 - Indirect Generalized Variance Base Parameters for Wave 1

Characteristics	Parameters			
PERSONS	**a**	**b**	**DEFF**	**f**
Total or White				
16+ Poverty and Program Participation				
Both Sexes	-0.00002444	5,342	2.21	0.87
Male	-0.00005077	5,342	2.21	0.87
Female	-0.00004712	5,342	2.21	0.87
16+ Income and Labor Force				
Both Sexes	-0.00001950	4,263	1.76	0.78
Male	-0.00004051	4,263	1.76	0.78
Female	-0.00003760	4,263	1.76	0.78
Other Person Items				
Both Sexes	-0.00002511	7,002	2.89	1.00
Male	-0.00005145	7,002	2.89	1.00
Female	-0.00004903	7,002	2.89	1.00
Black				
Person Items				
Both Sexes	-0.00012805	4,475	1.85	0.80
Male	-0.00027985	4,475	1.85	0.80
Female	-0.00023605	4,475	1.85	0.80
Hispanic				
Person Items				
Both Sexes	-0.00019658	6,515	2.69	0.96
Male	-0.00038425	6,515	2.69	0.96
Female	-0.00040250	6,515	2.69	0.96
HOUSEHOLDS				
Total or White	-0.00003286	3,546	1.47	1.00
Black	-0.00019168	2,495	1.03	0.84
Hispanic	-0.00035803	3,323	1.37	0.97

[2] Use the “Total or White Other Person Items” parameters for (1) tabulations of people aged 0+ in labor force, (2) retirement tabulations, (3) tabulations of Combined who are: aged 0+ in program participation, benefits, and income, and (4) tabulation of characteristics not specifically specified in this table, for the total or white population.

Table 3 (Continued) - SIPP Panel 2001 - Indirect Generalized Variance Base Parameters for Wave 2 and Wave 3

Characteristics	Parameters			
PERSONS	**a**	**b**	**DEFF**	**f**
Total or White				
16+ Poverty and Program Participation				
Both Sexes	-0.00003113	6,828	2.40	0.81
Male	-0.00006469	6,828	2.40	0.81
Female	-0.00006001	6,828	2.40	0.81
16+ Income and Labor Force				
Both Sexes	-0.00002458	5,391	1.90	0.72
Male	-0.00005108	5,391	1.90	0.72
Female	-0.00004738	5,391	1.90	0.72
Other Person Items				
Both Sexes	-0.00003130	8,753	3.08	0.92
Male	-0.00006415	8,753	3.08	0.92
Female	-0.00006112	8,753	3.08	0.92
Black				
Person Items				
Both Sexes	-0.00019935	7,002	2.47	0.82
Male	-0.00043655	7,002	2.47	0.82
Female	-0.00036690	7,002	2.47	0.82
Hispanic				
Person Items				
Both Sexes	-0.00030514	10,371	3.65	1.00
Male	-0.00059697	10,371	3.65	1.00
Female	-0.00062417	10,371	3.65	1.00
HOUSEHOLDS				
Total or White	-0.00003723	4,028	1.42	0.93
Black	-0.00028036	3,618	1.27	0.88
Hispanic	-0.00047316	4,626	1.63	1.00

Table 3 (Continued) - SIPP Panel 2001 - Indirect Generalized Variance Base Parameters for Wave 4 to Wave 6

Characteristics	Parameters			
PERSONS	**a**	**b**	**DEFF**	**f**
Total or White				
16+ Poverty and Program Participation				
Both Sexes	-0.00003417	7,517	2.65	0.84
Male	-0.00007096	7,517	2.65	0.84
Female	-0.00006591	7,517	2.65	0.84
16+ Income and Labor Force				
Both Sexes	-0.00002684	5,905	2.08	0.75
Male	-0.00005574	5,905	2.08	0.75
Female	-0.00005178	5,905	2.08	0.75
Other Person Items				
Both Sexes	-0.00003322	9,359	3.30	0.94
Male	-0.00006786	9,359	3.30	0.94
Female	-0.00006506	9,359	3.30	0.94
Black				
Person Items				
Both Sexes	-0.00020885	7,354	2.59	0.83
Male	-0.00045725	7,354	2.59	0.83
Female	-0.00038444	7,354	2.59	0.83
Hispanic				
Person Items				
Both Sexes	-0.00029967	10,568	3.72	1.00
Male	-0.00058335	10,568	3.72	1.00
Female	-0.00061623	10,568	3.72	1.00
HOUSEHOLDS				
Total or White	-0.00003787	4,122	1.45	0.88
Black	-0.00027786	3,789	1.33	0.84
Hispanic	-0.00049604	5,322	1.87	1.00

Table 3 (Continued) - SIPP Panel 2001 - Indirect Generalized Variance Base Parameters for Wave 7 to Wave 9

Characteristics	Parameters			
PERSONS	**a**	**b**	**DEFF**	**f**
Total or White				
16+ Poverty and Program Participation				
Both Sexes	-0.00003367	7,581	2.67	0.77
Male	-0.00006944	7,581	2.67	0.77
Female	-0.00006537	7,581	2.67	0.77
16+ Income and Labor Force				
Both Sexes	-0.00002657	5,983	2.11	0.69
Male	-0.00005480	5,983	2.11	0.69
Female	-0.00005159	5,983	2.11	0.69
Other Person Items				
Both Sexes	-0.00003508	10,020	3.53	0.89
Male	-0.00007151	10,020	3.53	0.89
Female	-0.00006885	10,020	3.53	0.89
Black				
Person Items				
Both Sexes	-0.00022157	7,953	2.80	0.79
Male	-0.00048801	7,953	2.80	0.79
Female	-0.00040583	7,953	2.80	0.79
Hispanic				
Person Items				
Both Sexes	-0.00034664	12,746	4.49	1.00
Male	-0.00067557	12,746	4.49	1.00
Female	-0.00071195	12,746	4.49	1.00
HOUSEHOLDS				
Total or White	-0.00004011	4,502	1.59	0.85
Black	-0.00030905	4,350	1.53	0.84
Hispanic	-0.00055052	6,204	2.18	1.00

Table 4 - Factors to be Applied to Table 3 Base Parameters to Obtain Parameters for Various Reference Periods

Number of Available Rotation Months[3]	Factor
Monthly Estimate	
1	4.0000
2	2.0000
3	1.3333
4	1.0000
Quarterly Estimate	
6	1.8519
8	1.4074
9	1.2222
10	1.0494
11	1.0370
12	1.0000

[3] The number of available rotation months for a given estimate is the sum of the number of rotations available for each month of the estimates.

Table 5 - Base Standard Errors of Estimated Numbers (in thousands) of Households, Families, and Households of Unrelated Residents

Size of Estimate	Base Standard Error	Size of Estimate	Base Standard Error
200	27	25,000	264
300	33	30,000	281
500	42	40,000	303
750	52	50,000	314
1,000	60	60,000	314
2,000	84	70,000	303
3,000	103	75,000	293
5,000	131	80,000	280
7,500	159	90,000	242
10,000	181	100,000	180
15,000	216	105,000	129

Notes: (1) This table is developed based on Wave 1. To account for sample attrition, multiply the base standard error by a factor of 1.09 for estimates including data from Wave 2 and/or Wave 3, a factor of 1.13 for estimates including data from Wave3 and/or Wave 4 and/or Wave 6, and a factor of 1.17 for estimates including data from Wave 7 and/or Wave 8 and/or Wave 9.

(2) Multiply the base standard error in this table by an appropriate f factor provided in Table 3 to obtain the final standard error estimate.

Table 6 - Base Standard Errors of Estimated Numbers (in Thousands) of People

Size of Estimate	Base Standard Errors	Size of Estimate	Base Standard Errors
200	38	90,000	657
300	46	100,000	675
500	59	110,000	688
750	73	120,000	697
1,000	84	130,000	703
2,000	118	140,000	705
3,000	145	150,000	703
5,000	186	160,000	698
7,500	227	170,000	690
10,000	261	180,000	677
15,000	316	190,000	661
25,000	401	200,000	640
30,000	435	210,000	614
40,000	492	220,000	583
50,000	539	230,000	546
60,000	577	240,000	501
70,000	609	250,000	446
75,000	623	260,000	376
80,000	636	275,500	208

Notes: (1) This table is developed based on Wave 1. To account for sample attrition, multiply the base standard error by a factor of 1.09 for estimates including data from Wave 2 and/or Wave 3, a factor of 1.13 for estimates including data from Wave3 and/or Wave 4 and/or Wave 6, and a factor of 1.17 for estimates including data from Wave 7 and/or Wave 8 and/or Wave 9.

(2) Multiply the base standard error in this table by an appropriate f factor provided in Table 3 to obtain the final standard error estimate.

Table 7 - Base Standard Errors of Estimated Percentages of Households, Families, and Households of Unrelated Residents

Base of Estimated Percentage (in Thousands)	Estimated Percentages					
	≤1 or ≥99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	1.34	1.88	2.93	4.03	5.82	6.72
300	1.09	1.54	2.39	3.29	4.75	5.49
500	0.85	1.19	1.85	2.55	3.68	4.25
750	0.69	0.97	1.51	2.08	3.00	3.47
1,000	0.60	0.84	1.31	1.80	2.60	3.00
2,000	0.42	0.59	0.93	1.27	1.84	2.12
3,000	0.35	0.49	0.76	1.04	1.50	1.73
5,000	0.27	0.38	0.59	0.81	1.16	1.34
7,500	0.22	0.31	0.48	0.66	0.95	1.10
10,000	0.19	0.27	0.41	0.57	0.82	0.95
15,000	0.15	0.22	0.34	0.47	0.67	0.78
25,000	0.12	0.17	0.26	0.36	0.52	0.60
30,000	0.11	0.15	0.24	0.33	0.48	0.55
40,000	0.09	0.13	0.21	0.29	0.41	0.48
50,000	0.08	0.12	0.19	0.25	0.37	0.42
60,000	0.08	0.11	0.17	0.23	0.34	0.39
70,000	0.07	0.10	0.16	0.22	0.31	0.36
75,000	0.07	0.10	0.15	0.21	0.30	0.35
80,000	0.07	0.09	0.15	0.20	0.29	0.34
90,000	0.06	0.09	0.14	0.19	0.27	0.32
100,000	0.06	0.08	0.13	0.18	0.26	0.30
105,000	0.06	0.08	0.13	0.18	0.25	0.29

Notes: (1) This table is developed based on Wave 1. To account for sample attrition, multiply the base standard error by a factor of 1.09 for estimates including data from Wave 2 and/or Wave 3, a factor of 1.13 for estimates including data from Wave3 and/or Wave 4 and/or Wave 6, and a factor of 1.17 for estimates including data from Wave 7 and/or Wave 8 and/or Wave 9..

(2) Multiply the base standard error in this table by an appropriate f factor provided in Table 3 to obtain the final standard error estimate.

Table 8 - Base Standard Errors of Estimated Percentages of People

Base of Estimated Percentage (in Thousands)	Estimated Percentages					
	≤1 or ≥99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	1.87	2.63	4.09	5.63	8.13	9.39
300	1.53	2.15	3.34	4.60	6.64	7.67
600	1.08	1.52	2.36	3.25	4.69	5.42
1,000	0.84	1.18	1.83	2.52	3.64	4.20
2,000	0.59	0.83	1.29	1.78	2.57	2.97
5,000	0.37	0.53	0.82	1.13	1.63	1.88
7,500	0.31	0.43	0.67	0.92	1.33	1.53
10,000	0.26	0.37	0.58	0.80	1.15	1.33
15,000	0.22	0.30	0.47	0.65	0.94	1.08
20,000	0.19	0.26	0.41	0.56	0.81	0.94
25,000	0.17	0.24	0.37	0.50	0.73	0.84
30,000	0.15	0.21	0.33	0.46	0.66	0.77
50,000	0.12	0.17	0.26	0.36	0.51	0.59
75,000	0.10	0.14	0.21	0.29	0.42	0.48
100,000	0.08	0.12	0.18	0.25	0.36	0.42
125,000	0.07	0.11	0.16	0.23	0.33	0.38
150,000	0.07	0.10	0.15	0.21	0.30	0.34
200,000	0.06	0.08	0.13	0.18	0.26	0.30
225,000	0.06	0.08	0.12	0.17	0.24	0.28
250,000	0.05	0.07	0.12	0.16	0.23	0.27
260,000	0.05	0.07	0.11	0.16	0.23	0.26
275,500	0.05	0.07	0.11	0.15	0.22	0.25

Notes: (1) This table is developed based on Wave 1. To account for sample attrition, multiply the base standard error by a factor of 1.09 for estimates including data from Wave 2 and/or Wave 3, a factor of 1.13 for estimates including data from Wave3 and/or Wave 4 and/or Wave 6, and a factor of 1.17 for estimates including data from Wave 7 and/or Wave 8 and/or Wave 9.

(2) Multiply the base standard error in this table by an appropriate f factor provided in Table 3 to obtain the final standard error estimate.

Table 9 - Topical Module Generalized Variance Parameters for the SIPP Panel 2001

Characteristics	Parameters a	Parameters b
Employment History, Wave 1		
Both Sexes 18+	-0.00001950	4,263
Males 18+	-0.00004051	4,263
Females 18+	-0.00003760	4,263
Recipiency History, Wave 1		
Both Sexes 18+	-0.00002444	5,342
Males 18+	-0.00005077	5,342
Females 18+	-0.00004712	5,342
Fertility History, Wave 2		
Women	-0.00003819	4,349
Births	-0.00006964	7,929
Education Attainment, Wave 2	-0.00002699	5,923
Marital Status and Person's Family Characteristics, Wave 2		
Some Household Members	-0.00004087	8,963
All Household Members	-0.00003773	10,892
Child Support		
Wave 5	-0.00006353	7,283
Wave 8	-0.00007893	9,245
Support for Non-Household Members		
Wave 5	-0.00003295	7,283
Wave 8	-0.00004094	9,245

Characteristics	Parameters	
	a	b
Health and Disability		
Wave 5	-0.00003139	9,113
Wave 8	-0.00002892	8,446
Child Care, Age 0 to 15, Wave 4	-0.00009227	6,437
Welfare History and AFDC		
Both Sexes 18+ (Wave 5)	-0.00007451	15,858
Males 18+ (Wave 5)	-0.00015497	15,858
Females 18+ (Wave 5)	-0.00014375	15,858
Both Sexes 18+ (Wave 8)	-0.00007804	16,849
Males 18+ (Wave 8)	-0.00016172	16,849
Females 18+ (Wave 8)	-0.00015088	16,849
Assets and Liabilities		
Wave 3	-0.00002722	5,980
Wave 6	-0.00002723	6,039
Wave 9	-0.00002943	6,637
2001 Migration History, Wave 2	-0.00002570	5,666

Table 10 - Distribution of Monthly Cash Income Among People 25 to 34 Years Old (Not Actual Data and to Be Used for Only Calculation Illustrations)

	Interval of Monthly Cash Income												
	Under $300	$300 to $599	$600 to $899	$900 to $1,119	$1,200 to $1,499	$1,500 to $1,999	$2,000 to $2,499	$2,500 to $2,999	$3,000 to $3,499	$3,500 to $3,999	$4,000 to $4,999	$5,000 to $5,999	$6,000 and Over
Number of People in Each Interval (in thousands)	1,371	1,651	2,259	2,734	3,452	6,278	5,799	4,730	3,723	2,519	2,619	1,223	1,493
Cumulative of People with at Least as Much as Lower Bound of Each Interval (in thousands)	39,851 (Total People)	38,480	36,829	34,570	31,836	28,384	22,106	16,307	11,577	7,854	5,335	2,716	1,493
Percent of People with at Least as Much as Lower Bound of Each Interval	100	96.6	92.4	86.7	79.9	71.2	55.5	40.9	29.1	19.7	13.4	6.8	3.7