Appendix C. Source and Accuracy of Estimates

SOURCE OF DATA

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. 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. With the exceptions noted above, persons who were at least 15 years of age at the time of the interview were eligible to be interviewed in the survey.

The 1985, 1986, and 1987 panel SIPP samples are located in 230 primary sampling units (PSU's) each consisting of a county or a group of contiguous counties. Within these PSU's, expected clusters of two to four living quarters (LQ's) were systematically selected from lists of addresses prepared for the 1980 decennial census to form the bulk of the sample. To account for LQ's built within each of the sample areas after the 1980 census, a sample was drawn of permits issued for construction of residential LQ's up until shortly before the beginning of the panel. In jurisdictions that do not issue building permits, small land areas were sampled and the LQ's within were listed by field personnel and then subsampled. In addition, sample LQ's were selected from a supplemental frame that included LQ's identified as missed in the 1980 census and group quarters.

The first interview of each panel was conducted during February, March, April, and May of that particular year. Approximately one-fourth of the sample was interviewed in each of these months. Each sample person was visited every 4 months thereafter. At each interview the reference period was the 4 months preceding the interview month.

Approximately 17,800, 16,300, and 16,700 living quarters were originally designated for the 1985, 1986, and 1987 samples, respectively. At the first interview, interviews were obtained from the occupants of about 13,400 of the 17,800 designated LQ's for the 1985 panel, 11,500 of the 16,300 designated LQ's for the 1986 panel, and 11,700 for the 16,700 designated LQ's for the 1987 panel. Most of the remaining 4,400, 4,800, and 5,000 LQ's in the 1985, 1986, and 1987 panels respectively, were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. However, approximately 1,000 of the 4,400 LQ's for the 1985 panel, 900 of the 4,800 LQ's for the 1986 panel, and 800 of the 5,000 LQ's for the 1987 panel 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 93 percent of all eligible living quarters for all three panels participated in the first interview of the survey.

For subsequent interviews, only original sample persons (those interviewed in the first interview) and persons living with them were eligible to be interviewed. Original sample persons were followed if they moved to a new address, unless the new address was more than 100 miles from a SIPP sample area. Then, telephone interviews were attempted. All first interview noninterviewed households were automatically designated as noninterviews for all subsequent interviews. When original sample persons moved to remote parts of the country, moved without leaving a forwarding address or refused to be interviewed, additional noninterviews resulted.

A person was classified as interviewed or noninterviewed for the entire 1985 panel based on the following definitions. Interviewed sample persons were defined to be 1) those for whom self or proxy responses were obtained for each reference month of all eight interviews or 2) those for whom self or proxy responses were obtained for the first reference month of the panel and for each subsequent reference month until they were known to have died or moved to an ineligible address (foreign living quarters, institutions, or military barracks). Noninterviewed persons were defined to be those for whom neither self nor proxy responses were obtained for one or more reference months of the eight interviews (but not because they were deceased or moved to an ineligible address).

For longitudinal estimates, approximately 36,500 persons were counted as initially interviewed in the 1985 panel. In the 1985 panel weighting procedure, approximately 23,000 persons were classified as interviewed. Persons who missed interviews due to the February 1986 sample cut were not classified as noninterviews but were adjusted for in the weighting procedure by a special factor (see "Estimation"). The person nonresponse rate is estimated to be 37 percent for the entire

1985 panel. Some respondents did not respond to some of the questions; therefore, the overall nonresponse rate for some items, especially sensitive income and money related items, is higher than the person nonresponse rate.

ESTIMATION

The estimation procedure used to derive SIPP cross-sectional person weights for each panel involve 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. Then, a noninterview factor was applied to the weight of every occupant of interviewed households to account for persons in noninterviewed occupied households which were eligible for the sample. (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.

Similarly, several stages of weight adjustments were involved in the estimation procedure used to derive the SIPP longitudinal person weights. Each person received a base weight equal to the inverse of his/her probability of selection. Two noninterview adjustment factors were applied. One adjusted the weights of interviewed persons in interviewed households to account for households which were eligible for the sample but could not be interviewed at the first interview. The second was applied to compensate for person noninterviews occurring in subsequent interviews. The Bureau has used complex techniques to adjust the weights for nonresponse, but the success of these techniques in avoiding bias is unknown. Another 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 cross-sectional and longitudinal person weights was performed to reduce the mean square error of the survey estimates. This was accomplished by bringing the sample estimates into agreement with monthly Current Population Survey (CPS) type estimates of the civilian (and some military) noninstitutional population of the United States by demographic characteristics including age, sex, race, and Hispanic ethnicity as of the specified control date. The CPS estimates by age, race, sex, and Hispanic origin were themselves brought into agreement with estimates from the 1980 decennial census which have been adjusted to reflect births, deaths, immigration, emigration, and changes in the Armed Forces since 1980. In cross-sectional weighting only, husbands and wives are assigned equal weights.

Two sources of error were identified in weighting of the 1985 panel. Two first stage factors were incorrect and inconsistent independent controls (independent estimates) were used during the second stage ratio adjustment procedure. The impact of these two error sources on primary SIPP estimates is believed to be minimal.

The data for the quarterly estimates provided in the report were obtained from the 1985, 1986, and 1987 SIPP panels. The quarterly estimates are based on combined data from two overlapping SIPP samples when available. Otherwise, they are based on data from only one sample. The data for the longitudinal estimates provided in the report were obtained from all eight interviews of the 1985 panel of the Survey of Income and Program Participation (SIPP).

ACCURACY OF 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 use in data analysis.

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

Table C-1. **SIPP Generalized Variance Parameters for Estimates Using Panel Weights from the 1985 Longitudinal Panel File**

(Use for longitudinal estimates section)

Characteristic	a	b	f factor
Total or White Persons			
All characteristics:			
Both sexes	-0.0001800	41,497	1.00
Male	-0.0003722	41,497	1.00
Female	-0.0003485	41,497	1.00
Black Persons			
All characteristics:			
Both sexes	-0.0005564	15,351	0.61
Male	-0.0011956	15,351	0.61
Female	-0.0010406	15,351	0.61

Table C-2. **Standard Errors of Estimated Numbers of Persons for 1985 Longitudinal Panel File**

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
200	91	22,000	909
300	112	26,000	978
600	158	30,000	1041
1,000	203	50,000	1275
2,000	287	80,000	1472
5,000	451	100,000	1533
8,000	566	130,000	1534
11,000	659	150,000	1475
13,000	713	200,000	1049
15,000	763	220,000	646
17,000	808	230,000	149

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 interviewing pattern used, and failure of all units in the universe to have some probability of being selected for the sample (undercoverage). Quality control and edit procedures were used to reduce errors made by respondents, coders, and interviewers.

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 characteristics different from those of interviewed persons in the same age-race-sex group. Further, the independent population controls used have not been adjusted for undercoverage.

Unique to the 1986 panel, maximum telephone interviewing was tested in waves 2, 3, and 4. Specifically, half of the sample in rotations 4 and 1 of wave 2 and rotations 2 and 3 of wave 3 (Phase I) and rotations 2, 3, and 4 of wave 4 (Phase II) were designated for telephone interviews. Analysis (done by designated mode) of household nonresponse, item nonresponse rates for labor force and income core items, and selected cross-sectional estimates of recipiency, income, low income status, and selected topical module items gave no indication of an overall significant mode effect. However, analysis was restricted to a limited number and type of estimates. If differences between two time periods or differences in characteristics for demographic groups result in borderline significant differences, the significance may be due to bias from the use of the telephone mode. Similarly, borderline insignificant differences may also be due to this bias. Thus, although no overall significant mode effect was detected, the user should consider the possibility of mode effects while analyzing exclusively the 1986 panel data or combined data involving the 1986 panel after wave 1, especially results based on waves 2 through 4 data. Details on analyses are in "Preliminary Evaluation of Maximum Telephone Interviewing on the SIPP" (paper by Gbur and Petroni in the forthcoming 1989 Proceedings of the Survey Research Methods Section, American Statistical Association) and "SIPP 86: Telephone Experiment-Preliminary Analysis" (internal Census Bureau draft memorandum from Waite to Davey, August 21, 1989).

Table C-3. **Standard Errors of Estimated Percentages of Persons for 1985 Longitudinal Panel File**

Base of estimated percentage (thousands)	Estimated percentage					
	≤1 or 99≥	2 or 98	5 or 95	10 or 90	25 or 75	50
200	4.5	6.4	9.9	13.7	19.7	22.8
300	3.7	5.2	8.1	11.2	16.1	18.6
600	2.6	3.7	5.7	7.9	11.4	13.1
1,000	2.0	2.9	4.4	6.1	8.8	10.2
2,000	1.4	2.0	3.1	4.3	6.2	7.2
5,000	0.9	1.3	2.0	2.7	3.9	4.6
8,000	0.7	1.0	1.6	2.2	3.1	3.6
11,000	0.6	0.9	1.3	1.8	2.7	3.1
13,000	0.6	0.8	1.2	1.7	2.4	2.8
17,000	0.5	0.7	1.1	1.5	2.1	2.5
22,000	0.4	0.6	0.9	1.3	1.9	2.2
26,000	0.4	0.6	0.9	1.2	1.7	2.0
30,000	0.4	0.5	0.8	1.1	1.6	1.9
50,000	0.3	0.4	0.6	0.9	1.2	1.4
80,000	0.2	0.3	0.5	0.7	1.0	1.1
100,000	0.2	0.3	0.4	0.6	0.9	1.0
130,000	0.2	0.3	0.4	0.5	0.8	0.9
180,000	0.2	0.2	0.3	0.5	0.7	0.8
200,000	0.1	0.2	0.3	0.4	0.6	0.7
220,000	0.1	0.2	0.3	0.4	0.6	0.7
230,000	0.1	0.2	0.3	0.4	0.6	0.7

Table C-4. **SIPP Generalized Variance Parameters for 1985 Panel Cross-Sectional Quarterly Estimates**

(Use for 1st quarter 1986 estimates)

Characteristic	a	b	f factor
Total or White Persons			
All characteristics:			
Both sexes	-0.0001621	37,372	1.00
Male	-0.0003352	37,372	1.00
Female	-0.0003139	37,372	1.00
Black Persons			
All characteristics:			
Both Sexes	-0.0005011	13,825	0.61
Male	-0.0010768	13,825	0.61
Female	-0.0009372	13,825	0.61

Comparability with other estimates. Caution should be exercised when comparing data from this report with data from other SIPP publications 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.

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.

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.

All statements of comparison in the report have passed a hypothesis test at the 0.10 level of significance or better. This means that, for differences cited in the report, the estimated absolute difference between parameters is greater than 1.6 times the standard error of the difference.

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

Table C-5. **Standard Errors of Estimated Numbers of Persons for 1985 Panel Cross-Sectional File**

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
200	86	22,000	862
300	106	26,000	928
600	150	30,000	988
1,000	193	50,000	1210
2,000	272	80,000	1397
5,000	428	100,000	1455
8,000	537	130,000	1456
11,000	626	150,000	1399
13,000	677	200,000	995
15,000	724	220,000	613
17,000	767	230,000	143

Table C-6. **Standard Errors of Estimated Percentages of Persons for 1985 Panel Cross-Sectional File**

Base of estimated percentage (thousands)	Estimated percentage					
	≤1 or 99≥	2 or 98	5 or 95	10 or 90	25 or 75	50
200	4.3	6.1	9.4	13.0	18.7	21.6
300	3.5	5.0	7.7	10.6	15.3	17.6
600	2.5	3.5	5.4	7.5	10.8	12.5
1,000	1.9	2.7	4.2	5.8	8.4	9.7
2,000	1.4	1.9	3.0	4.1	5.9	6.8
5,000	0.9	1.2	1.9	2.6	3.7	4.3
8,000	0.7	1.0	1.5	2.1	3.0	3.4
11,000	0.6	0.8	1.3	1.7	2.5	2.9
13,000	0.5	0.8	1.2	1.6	2.3	2.7
17,000	0.5	0.7	1.0	1.4	2.0	2.3
22,000	0.4	0.6	0.9	1.2	1.8	2.1
26,000	0.4	0.5	0.8	1.1	1.6	1.9
30,000	0.4	0.5	0.8	1.1	1.5	1.7
50,000	0.3	0.4	0.6	0.8	1.2	1.4
80,000	0.2	0.3	0.5	0.6	0.9	1.1
100,000	0.2	0.3	0.4	0.6	0.8	1.0
130,000	0.2	0.2	0.4	0.5	0.7	0.8
180,000	0.1	0.2	0.3	0.4	0.6	0.7
200,000	0.1	0.2	0.3	0.4	0.6	0.7
230,000	0.1	0.2	0.3	0.4	0.6	0.6
250,000	0.1	0.2	0.3	0.4	0.5	0.6

characteristics are different. Of course, sometimes this conclusion will be wrong. When the characteristics are, in fact, 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, 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. Summary measures 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 estimates will reveal useful information when computed on a base smaller than 200,000. Also, nonsampling error in one or more of the small number of cases providing the estimate can cause large relative error in that particular estimate. 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. Therefore, 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.

Standard error parameters and tables and their use. Most SIPP estimates have greater standard errors than those obtained through a simple random sample because clusters of living quarters are sampled for the SIPP. 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 and two parameters (denoted "a" and "b") were developed to approximate the standard error behavior of each group of estimates. 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. These "a" and "b" parameters vary by characteristic and by demographic subgroup to which the

Table C-7. **SIPP Generalized Variance Parameters for 1986 and 1987 Panels Cross-Sectional Quarterly Estimates**

(Use for 1st quarter 1988 through 4th quarter 1988 estimates)

Characteristic	a1	b1	f factor
Total or White Persons			
All characteristics:			
Both sexes	-0.0001356	31,260	1.00
Male	-0.0002804	31,260	1.00
Female	-0.0002625	31,260	1.00
Black Persons			
All characteristics:			
Both sexes	-0.0004192	11,565	0.61
Male	-0.0009007	11,565	0.61
Female	-0.0007839	11,565	0.61

[1]To account for sample attrition, multiply the "a" and "b" parameters by 1.09 for estimates which include data from wave 5 and beyond.

Table C-8. **Standard Errors of Estimated Numbers of Persons for 1986 Panel and 1987 Panel Cross-Sectional File**

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
200	79	22,000	789
300	97	26,000	849
600	137	30,000	903
1,000	176	50,000	1106
2,000	249	80,000	1278
5,000	391	100,000	1330
8,000	491	130,000	1331
11,000	572	150,000	1280
13,000	619	200,000	910
15,000	662	220,000	560
17,000	702	230,000	129

estimate applies. Table C-1 provides base "a" and "b" parameters to be used for 1985 longitudinal panel estimates. Table C-4 provides base "a" and "b" parameters for 1985 panel quarterly estimates. Table C-7 provides base "a" and "b" parameters for 1986 panel and 1987 panel quarterly estimates. Table C-10 provides parameters for 1985-86 combined panel quarterly estimates. Table C-13 provides parameters for 1986-87 combined panel quarterly estimates. For those users who wish further simplification, we have also provided general standard errors in tables C-2, C-3, C-5, C-6, C-8, C-9, C-11, C-12, C-14, and C-15. Note that these standard errors must be adjusted by a factor from table C-1, C-4, C-7, C-10, or C-13. 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.

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

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

where f is the appropriate "f" factor from table C-1, C-4, C-7, C-10, or C-13 and s is the standard error of the estimate obtained by interpolation from table C-2, C-5, C-8, C-11, or C-14. Alternatively, s_x may be approximated by the formula

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

Here x is the estimated number and "a" and "b" are the parameters associated with the particular type of characteristic. Use of formula (2) will provide more accurate results than the use of formula (1).

Illustration. The SIPP estimate of total number of persons lacking health insurance coverage for the first 28 months covered by the 1985 SIPP longitudinal panel is 9,739,000 as indicated in table C of the report. The appropriate "a" and "b" parameters to use in calculating a standard error for the estimate are obtained from table C-1. They are a= -0.0001800 and b = 41,497, respectively. Using formula (2), the approximate standard error is

$$\sqrt{(-0.0001800)(9{,}739{,}000)^2 + (41{,}497)(9{,}739{,}000)} = 622{,}000$$

Table C-9. **Standard Errors of Estimated Percentages of Persons for 1986 Panel and 1987 Panel Cross-Sectional File**

Base of estimated percentage (thousands)	Estimated percentage					
	≤1 or 99≥	2 or 98	5 or 95	10 or 90	25 or 75	50
200	3.9	5.5	8.6	11.9	17.1	19.8
300	3.2	4.5	7.0	9.7	14.0	16.1
600	2.3	3.2	5.0	6.8	9.9	11.4
1,000	1.8	2.5	3.9	5.3	7.7	8.8
2,000	1.2	1.8	2.7	3.8	5.4	6.3
5,000	0.8	1.1	1.7	2.4	3.4	4.0
8,000	0.6	0.9	1.4	1.9	2.7	3.1
11,000	0.5	0.7	1.2	1.6	2.3	2.7
13,000	0.5	0.7	1.1	1.5	2.1	2.5
17,000	0.4	0.6	0.9	1.3	1.9	2.1
22,000	0.4	0.5	0.8	1.1	1.6	1.9
26,000	0.3	0.5	0.8	1.0	1.5	1.7
30,000	0.3	0.5	0.7	1.0	1.4	1.6
50,000	0.2	0.4	0.5	0.8	1.1	1.3
80,000	0.2	0.3	0.4	0.6	0.9	1.0
100,000	0.2	0.2	0.4	0.5	0.8	0.9
130,000	0.2	0.2	0.3	0.5	0.7	0.8
180,000	0.1	0.2	0.3	0.4	0.6	0.7
200,000	0.1	0.2	0.3	0.4	0.5	0.6
230,000	0.1	0.2	0.3	0.3	0.5	0.6
250,000	0.1	0.2	0.2	0.3	0.5	0.6

Table C-10. **SIPP 1985, 1986 Combined Panels Generalized Variance Parameters for Cross-Sectional Quarterly Estimates**

(Use for 2nd quarter 1986 through 1st quarter 1987)

Characteristic	a	a	f factor
Total or White Persons			
All characteristics:			
Both sexes	-0.0000814	18,775	1.00
Male	-0.0001684	18,775	1.00
Female	-0.0001576	18,775	1.00
Black Persons			
All characteristics:			
Both Sexes	-0.0002517	6,946	0.61
Male	-0.0005409	6,946	0.61
Female	-0.0004708	6,946	0.61

Table C-11. **Standard Errors of Estimated Numbers of Persons for 1985, 1986 Combined Panels Cross-Sectional File**

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
200	61	22,000	611
300	75	26,000	658
600	106	30,000	700
1,000	137	50,000	857
2,000	193	80,000	990
5,000	303	100,000	1031
8,000	381	130,000	1032
11,000	443	150,000	992
13,000	480	200,000	706
15,000	513	220,000	437
17,000	544	230,000	110

The 90-percent confidence interval as shown by the data is from 8,744,000 to 10,734,000. 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 90 percent of all samples.

Using formula (1), the appropriate "f" factor (f=1.00) from table C-1, and the standard error of the estimate by interpolation using table C-2, the approximate standard error is

$$s_x = (1.00)(620{,}000) = 620{,}000$$

The 90 percent confidence interval as shown by the data is from 8,747,000 to 10,731,000.

Standard errors of estimated percentages. This section refers to the type of percentages presented in this report. These are the percentages of a group of persons possessing a particular attribute. For example, the percentage of persons who were covered by health insurance for the entire SIPP 1985 longitudinal panel. 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 over 50 percent. For example, the percent of persons covered by health insurance is more reliable than the estimated number of persons covered by health insurance. When the numerator and denominator of the percentage have

Table C-12. **Standard Errors of Estimated Percentages of Persons for 1985, 1986 Combined Panels Cross-Sectional File**

Base of estimated percentage (thousands)	Estimated percentage					
	≤1 or 99≥	2 or 98	5 or 95	10 or 90	25 or 75	50
200	3.0	4.3	6.7	9.2	13.3	15.3
300	2.5	3.5	5.5	7.5	10.8	12.5
600	1.8	2.5	3.9	5.3	7.7	8.8
1,000	1.4	1.9	3.0	4.1	5.9	6.9
2,000	1.0	1.4	2.1	2.9	4.2	4.8
5,000	0.6	0.9	1.3	1.8	2.7	3.1
8,000	0.5	0.7	1.1	1.5	2.1	2.4
11,000	0.4	0.6	0.9	1.2	1.8	2.1
13,000	0.4	0.5	0.8	1.1	1.6	1.9
17,000	0.3	0.5	0.7	1.0	1.4	1.7
22,000	0.3	0.4	0.6	0.9	1.3	1.5
26,000	0.3	0.4	0.6	0.8	1.2	1.3
30,000	0.2	0.4	0.5	0.8	1.1	1.3
50,000	0.2	0.3	0.4	0.6	0.8	1.0
80,000	0.2	0.2	0.3	0.5	0.7	0.8
100,000	0.1	0.2	0.3	0.4	0.6	0.7
130,000	0.1	0.2	0.3	0.4	0.5	0.6
180,000	0.1	0.1	0.2	0.3	0.4	0.5
200,000	0.1	0.1	0.2	0.3	0.4	0.5
230,000	0.1	0.1	0.2	0.3	0.4	0.5
250,000	0.1	0.1	0.2	0.3	0.4	0.4

Table C-13. **SIPP 1986, 1987 Combined Panels Generalized Variance Parameters for Cross-Sectional Quarterly Estimates**

(Use for 2nd quarter 1987 through 4th quarter 1987)

Characteristic	a	b	f factor
Total or White Persons			
All characteristics:			
Both sexes	-0.0000802	18,492	1.00
Male	-0.0001658	18,492	1.00
Female	-0.0001552	18,492	1.00
Black Persons			
All characteristics:			
Both sexes	-0.0002479	6,842	0.61
Male	-0.0005328	6,842	0.61
Female	-0.0004637	6,842	0.61

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.

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

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

where f is the appropriate "f" factor from table C-1, C-4, C-7, C-10, or C-13 and s is the standard error of the estimate obtained by interpolation from table C-3, C-6, C-9, C-12, or C-15. Alternatively, it may be approximated by the formula

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

Here x is the base of the percentage, p is the percentage ($0 < p < 100$) and b is the "b" parameter associated with the characteristic in the numerator. Use of this formula will give more accurate results than use of formula (3).

Illustration. Suppose that the SIPP estimate for the number of persons in the second quarter of 1986 is 237,184,000 as indicated in table B of the report. Of these, 85.7 percent were covered by private or government health insurance. Using formula (4) and the "b" parameter of 18,775 (from table 10), the approximate standard error is

$$\sqrt{\frac{(18{,}775)}{(237{,}184{,}000)}(85.7)(100.-85.7)} = 0.3 \text{ percent}$$

Consequently, the 90-percent confidence interval as shown by these data is from 85.2 to 86.2 percent.

Using formula (3), the appropriate "f" factor (f=1.00) from table C-10, and the appropriate s by interpolation using table C-12, the approximate standard error is

$$s_x = (1.0)(0.40) = 0.4 \text{ percent}$$

The 90 percent confidence interval shown by these data is from 85.0 to 86.4 percent.

Standard error of a difference. The standard error of a difference between two sample estimates, x and y, is equal to

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

where s_x and s_y are the standard errors of the estimates x and y, and r is the correlation coefficient between the characteristics estimated by x and y. The estimates can be numbers, averages, percents, ratios, etc. Underestimates or overestimates of standard error of differences result if the estimated correlation coefficient is overestimated or underestimated, respectively.

Illustration. Suppose that we are interested in the difference in the percentage of Blacks and Whites who did not receive health insurance coverage for at least one month during the 1985 SIPP panel. Of the 26,954,000 Blacks and 192,194,000 Whites, 37.7 percent and 26.4 percent, respectively, did not receive health insurance coverage for at least 1 month (see table C of the report). Using parameters and factors from table C-1, the standard errors of these percentages are approximately 1.9 percent and 1.1 percent for Blacks and Whites, respectively.

Now, the standard error of the difference is computed using the above two standard errors. The correlation between these estimates is assumed to be zero. Therefore, standard error of the difference is computed by formula (9):

$$\sqrt{(1.9)^2 + (1.1)^2} = 2.2 \text{ percent}$$

Suppose that it is desired to test at the 10 percent significance level whether the percentage of Blacks differs significantly from the number of Whites. To

Table C-14. **Standard Errors of Estimated Numbers of Persons for 1986, 1987 Combined Panels Cross-Sectional File**

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
200	61	22,000	607
300	74	26,000	653
600	105	30,000	695
1,000	136	50,000	851
2,000	191	80,000	983
5,000	301	100,000	1023
8,000	378	130,000	1024
11,000	440	150,000	985
13,000	476	200,000	700
15,000	509	220,000	432
17,000	540	230,000	103

Table C-15. **Standard Errors of Estimated Percentages of Persons for 1986, 1987 Combined Panels Cross-Sectional File**

Base of estimated percentage (thousands)	Estimated percentage					
	≤1 or 99≥	2 or 98	5 or 95	10 or 90	25 or 75	50
200	3.0	4.3	6.6	9.1	13.2	15.2
300	2.5	3.5	5.4	7.4	10.8	12.4
600	1.7	2.5	3.8	5.3	7.6	8.8
1,000	1.4	1.9	3.0	4.1	5.9	6.8
2,000	1.0	1.3	2.1	2.9	4.2	4.8
5,000	0.6	0.9	1.3	1.8	2.6	3.0
8,000	0.5	0.7	1.0	1.4	2.1	2.4
11,000	0.4	0.6	0.9	1.2	1.8	2.1
13,000	0.4	0.5	0.8	1.1	1.6	1.9
17,000	0.3	0.5	0.7	1.0	1.4	1.6
22,000	0.3	0.4	0.6	0.9	1.3	1.4
26,000	0.3	0.4	0.6	0.8	1.2	1.3
30,000	0.2	0.3	0.5	0.7	1.1	1.2
50,000	0.2	0.3	0.4	0.6	0.8	1.0
80,000	0.2	0.2	0.3	0.5	0.7	0.7
100,000	0.1	0.2	0.3	0.4	0.6	0.7
130,000	0.1	0.2	0.3	0.4	0.5	0.6
180,000	0.1	0.1	0.2	0.3	0.4	0.5
200,000	0.1	0.1	0.2	0.3	0.4	0.5
230,000	0.1	0.1	0.2	0.3	0.4	0.4
250,000	0.1	0.1	0.2	0.3	0.4	0.4

perform the test, compare the difference of 11.3 percent to the product 1.6 x 2.2 percent = 3.6 percent. Since the difference is larger than 1.6 times the standard error of the difference, the data show that the estimates for the percentage of Blacks and Whites who did not receive health insurance coverage for at least 1 month differ significantly at the 10 percent level.