ICPSR
Inter-university Consortium for Political and Social Research

Survey of Income and Program Participation (SIPP) 1984 Full Panel Research File

User Notes

U.S. Dept. of Commerce
Bureau of the Census

ICPSR 9331

This document was previously available in paper format only. It was converted to Portable Document Format (PDF), with no manual editing, on the date below as part of ICPSR's electronic document conversion project. The document may not be completely searchable. No additional updating of this collection has been performed (pagination, missing pages, etc.).

May 2002

ICPSR
Inter-university Consortium for Political and Social Research

Survey of Income and Program Participation (SIPP) 1984 Full Panel Research File

User Notes

U.S. Dept. of Commerce
Bureau of the Census

ICPSR 9331

SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP)
1984 FULL PANEL RESEARCH FILE: USER NOTES

(ICPSR 9365)

Principal Investigator

U.S. Dept. of Commerce
Bureau of the Census

Inter-university Consortium for Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106

First ICPSR Release
February 1991

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, D.C. 20233

SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP)
1984 FULL PANEL
MICRODATA FILE

USER NOTE NO. 1

Subject: Source and Accuracy Statement for the Family Disruption and Economic Hardship Report

Attached is the Source and Accuracy Statement for the report, "Family Disruption and Economic Hardship: The Short-Run Picture for Children" based on data collected in the 1984 Full Panel of the Survey of Income and Program Participation (SIPP).

November 1990

SOURCE AND ACCURACY STATEMENT FOR THE REPORT TITLED "FAMILY DISRUPTION AND ECONOMIC HARDSHIP: THE SHORT-RUN PICTURE FOR CHILDREN" FROM THE SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP) 1984 LONGITUDINAL PANEL FILE

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.

1984 SIPP Panel. The 1984 panel SIPP sample is located in 174 areas comprising 450 counties (including one partial county) and independent cities. Within these areas, clusters of two to four LQs were systematically selected from lists of addresses prepared for the 1970 decennial census to form the bulk of the sample. In addition, the sample was updated to account for new residential construction since the 1970 census.

In jurisdictions that do not issue building permits, small land areas were sampled and the LQs within were listed by field personnel and then subsampled. In addition, sample LQs were selected from a supplemental frame that included LQs identified as missed in the 1980 census and group quarters.

The first cycle (i.e., wave) of interviewing of this panel was conducted during October, November, and December 1983, and January 1984. Approximately one-fourth of the sample was interviewed in each of these months. Each sample person was visited every four months thereafter. At each interview the reference period was the four months preceding the interview month.

Approximately 26,000 LQs were originally designated for the sample. At the first contact, interviews were obtained from the occupants of about 19,900 of the 26,000 designated LQs. Most of the remaining 6,100 LQs were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. However, approximately 1,000 of the 6,100 LQs were not interviewed because the occupants refused to be interviewed, could not be found at home, were temporarily absent, or were otherwise unavailable. Thus, occupants of about 95 percent of all eligible LQs 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 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 sample persons were defined to be everyone else.

ESTIMATION

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 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.

The data for the longitudinal estimates provided in the report are obtained from eight interviews of the 1984 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 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 Nonblacks. 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.

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 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, 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. 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 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. The Census Bureau created generalized variance parameters (denoted as "a" and "b") for use in the calculation of 1984 longitudinal panel estimates. However, for this report, the parameters were not used. Instead, the author used a variance generating program called VPLX, written by Robert Fay of the Census Bureau. VPLX was used to calculate the standard errors independently of Bureau estimates through the use of replication methods. The tables of standard errors are provided at the end of the report for user convenience. For example, standard errors for table 2. are provided in table A-2.

In using VPLX to calculate standard errors, individual correlations between related items were also calculated and used to estimate the standard errors of differences. Due to space limitations we are not providing the correlations in this document. [If users wish to obtain correlations for specific items call the author at (301) 763-8354.] As a result, the user may generate estimates that are not exactly the same as those given in the text.

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

$$s_{(x-y)} = \sqrt{s_x^2 + s_y^2} \qquad (1)$$

where s_x and s_y are the standard errors of the estimates x and y. The estimates can be numbers, averages, percents, ratios, etc. The above formula assumes that the correlation coefficient, r, between the characteristics estimated by x and y is zero. If r is really positive (negative), then this assumption will result in a tendency towards overestimates (underestimates) of the true standard error.

If users obtain the correlation coefficient between the characteristics estimated by x and y, the standard error of a difference is estimated by

$$s_{(x-y)} = \sqrt{S_x^2 + S_y^2 - 2xS_xS_y} \qquad (2)$$

Illustration. Table 4 of the report shows that in the eighth interview 24.1 percent of children whose father left during the panel received food stamps and 47 percent of children who lived with their mother only received food stamps. Using table A-4, the standard errors of these percentages are approximately 3.46 percent and 2.14 percent respectively.

The standard error of the difference is computed using formula (1):

$$\sqrt{(3.46)^2 + (2.14)^2} = 4.1 \text{ percent}$$

Suppose that it is desired to test at the 10 percent significance level whether the above two percentages differ significantly. To perform the test, compare the difference of 22.9 percent to the product of 1.6 x 4.1 = 6.6 percent. Since the difference is larger than 1.6 times the standard error of the difference, the data supports the hypothesis that the two percent estimates are significantly different at the 10 percent level.

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

$$s_{\frac{x}{y}} = \sqrt{\left[\frac{x}{y}\right]^2 \left[\left[\frac{s_y}{y}\right]^2 + \left[\frac{s_x}{x}\right]^2\right]} \qquad (3)$$

where x and y are means or medians, and s_x and s_y are their associated standard errors. Formula 3 assumes that the means or medians are not correlated. If the correlation between the population means or medians 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 or medians.

Illustration. Suppose the SIPP estimate of "Mother Only" to "Always Two Parents" mean family income at the first interview is 0.40. Also, suppose that the mean family income and its standard error are $1,132 and $49, respectively for "Mother Only" families; and $2,834

and $56, respectively for "Always Two Parent" families.

Using formula (3), the standard error of this ratio is approximated by:

$$s_{\frac{x}{y}} = \sqrt{\left[\frac{1132}{2834}\right]^2 \left[\left[\frac{49}{1132}\right]^2 + \left[\frac{56}{2834}\right]^2\right]}$$

$$= 0.019$$