Survey of Income and Program Participation

Weighting

What weights are and why they should be used

The weight for a responding unit in a survey data set is an estimate of the number of units in the target population that the responding unit represents. In general, since population units may be sampled with different selection probabilities and since response rates and coverage rates may vary across subpopulations, different responding units represent different numbers of units in the population. The use of weights in survey analysis compensates for this differential representation, thus producing estimates that relate to the target population.

SIPP weights vary due to differential sampling rates as a result of oversampling and because response and coverage rates vary across subpopulations. For example, in Wave 1 of the 2004 Panel, the final person lower quartile weight is 1,682 and the upper

quartile weight is 3,429 (the maximum weight is 16,482). A respondent with a final person weight of 1,682 represents 1,682 people in the U.S. population for the reference month, whereas a respondent with a weight of 3,429 represents 3,429 people. Because weights in SIPP vary over a sufficiently large range of values, performing unweighted analyses may produce appreciably biased estimates for the U.S. population.

Weights available in SIPP Files

Choosing a weight

How weights are constructed

Reference month weights

Full Panel and Calendar Year Weights

Weights available in SIPP Files

SIPP cross sectional data files contain reference month weights for each person, household, head of family and related head of subfamily. SIPP longitudinal data files contain calendar year weights and panel weights for each person.

Back to top

Choosing a weight

The decision of which weight to use for a given analysis depends on the population of interest for that analysis. Useful guidance for choosing the correct set of weights is to consider to what population the results are intended to apply.

The weights in the SIPP files are constructed for sample cohorts defined by:

- Month (e.g., the reference month weights in the core wave files and interview month weights in topical module files);
- Year (e.g., the calendar year weights in the full panel file); and
- Panel (e.g., the full panel weight in the full panel file).

Users can choose to base their analyses on:

- A cross-sectional sample at a given month;
- A longitudinal sample that provides continuous monthly data over a year;
- A longitudinal sample that provides monthly data over the life of a panel (48 months with the 1996 and 2004 Panels, 67 months with the 2008 Panel); or
- A subset of the sample and/or the period in any of the above.

Monthly (cross-sectional) weights allow the use of all available data for a given month. For this type of analysis, users can choose among the following units of analysis:

- Person (e.g., WPFINWGT);
- Household (e.g., WHFNWGT);
- Family (e.g., WFFINWGT); and
- Related subfamily (e.g., WSFINWGT).

Analysts can use longitudinal samples to follow the same people over time and hence study such issues as the dynamics of program participation, lengths of poverty spells, and changes in other circumstances (e.g., household composition). The longitudinal weights allow the inclusion of all people for whom data were collected for every month of the period involved (calendar year or full panel period), including those who left the target population through death or because they moved to an ineligible address (institution, foreign living quarters, military barracks), as well as those for whom data were imputed for missing months. The Census Bureau makes nonresponse adjustments to the longitudinal weights to compensate for panel attrition and poststratification adjustments to make the weighted sample totals conform to population totals for key variables.

Back to top

How weights are constructed

This section describes how the weights are constructed. The basic components for all the different sets of weights are the same, namely:

- A base weight that reflects the probability of selection for a sample unit;
- An adjustment for subsampling within clusters;
- An adjustment for movers (in Waves 2 and beyond);

- A nonresponse adjustment to compensate for sample nonresponse; and
- A poststratification (second-stage calibration) adjustment to correct for departures from known population totals.

Back to top

Reference Month Weights

Reference month final weights are provided on the SIPP core wave files for persons, households, families, and subfamilies. The special weights for persons are constructed first. The household, family, and related subfamily final weights are derived from the final person weights. This section summarizes the steps involved in constructing the various sets of weights, starting with the final person weights for a reference month. The reference and interview month weights[1] for people on the core wave files are computed (i.e., are nonzero) for all responding sample members who are in scope. (i.e., a part of the survey's universe.the resident, noninstitutional population of the United States) in the specified month.[2] A number of factors lead to fluctuations in sample size from month to month. They include births, deaths, immigration, and emigration from the population (and therefore from the sample). In addition to those population dynamics, people move into and out of the sample as a result of the changing household composition of sample members. In Wave 1, the weight for each sample person per month is a product of four components:

1. **Wave 1 base weight.** This weight is the inverse of the probability of a sample person's address being selected.
2. **Duplication-control factor.** This factor adjusts for the occasional subsampling of clusters. Clusters are occasionally subsampled in the field when they turn out to be much larger than expected.[3]
3. **Wave 1 nonresponse adjustment.** This adjustment compensates for different rates of household noninterview within adjustment classes. More than 500 nonresponse adjustment classes are defined based on a cross-classification of characteristics. Those characteristics include Census Region;

MSA/Place Status (MSA-central city, MSA-non-central city, other place); race of reference person (black, nonblack); household tenure (owner, renter); household size (1, 2, 3, 4+ people). In addition, the within-primary-sampling-unit poverty stratum (high poverty, low poverty) was added for the 1996 Panel.

4. **Wave 1 second-stage calibration.** This adjustment brings the sample estimates into agreement with independent monthly estimates of population totals. The characteristics used for calibration include age, race, sex, Hispanic origin, family relationship, and household type. A raking procedure is used to ensure that the weights agree with all the control totals included for calibration. The adjustment is done by rotation group, with each group assigned one-fourth of the population total for the month.

In subsequent waves, each person receives an initial weight that is carried over from the preceding wave. This weight is adjusted to compensate for changes in the sample between waves resulting from movers and nonresponse, and then it is realigned to match the population totals for the reference or interview month:

- **Wave 2+ initial weight.** This is the weight from the previous wave before the second-stage calibration for each original sample person who is a reference person or is in group quarters for the current wave.

- **Wave 2+ mover's adjustment.** This adjustment is made to compensate for including people who were not in the original sample but were in the SIPP universe in Wave 1 and who moved into a sample household after Wave 1. For people in housing units that contain adult members who were not part of the original sample but were in the SIPP universe at Wave 1, the weights are decreased. For example, if a third adult moves into a household occupied by two original sample persons, all three adults would receive the initial weight of the original sample persons multiplied by a factor of two-thirds.

- **Wave 2+ nonresponse adjustment.** The nonresponse adjustment for Waves 2 and beyond is used to compensate for household nonresponse after the first interview. The nonresponse adjustment classes are defined on the basis of sample unit characteristics and personal demographic characteristics[4] from the most recent wave. The information used consists of household characteristics. Reference person characteristics are used to define some of the household characteristics. Tenure (owner/renter occupied), household type (female householder, no spouse present; 65+; other), race and Hispanic origin, and education level are defined at the household level by using reference person data. Other household characteristics include size, poverty status, type of income, type of financial assets, census division, and number of imputed items. Poverty threshold, census division, and number of imputed items are new to the 1996 Panel. Some adjustment classes are combined to ensure that the adjustment for each class does not exceed a factor of 2, and each class contains at least 30 unweighted sample households.

- **Wave 2+ second-stage calibration.** To derive this adjustment, use the same procedure as in Wave 1; that is, use the appropriate population control totals by reference month. The reference month final weights for households, families, and subfamilies are derived from the person weights:

- The household weight is the person weight of the household reference person (renter/owner of housing unit).

- The family weight is the person weight of the family reference person.

- The subfamily weight for a related subfamily is the person weight of the related subfamily reference person.

Back to top

Full Panel and Calendar Year Weights

Final full panel and final calendar year weights are provided on the full panel files for eligible sample members. There is one set of final panel weights and generally more than one set of calendar year weights, one for each calendar year covered by the panel. Final panel weights are computed only for people who are in the sample at Wave 1 of the panel and for whom data are obtained (either reported or imputed) for every month of the panel for which they were in scope for the survey. Other people in the panel file are assigned weights of zero. Most people with nonzero final panel weights have provided data for all months of the panel. However, people who missed a wave and whose missing wave data were imputed and people who provided data up to the point that they left the survey (through death or because they moved to an ineligible address) are also assigned nonzero final panel weights. Final calendar year weights are computed only for people who had an interview covering the control date[5] and for whom data are obtained (either reported or imputed) for every month of the calendar year for which they were in scope for the survey. Other people are assigned final calendar year weights of zero. Some people who joined the household of an original sample person after the start of the panel are assigned nonzero calendar year weights for the second calendar year, if data are obtained for that period.

The full panel weighting scheme does not assign weights to people who enter the sample universe after Wave 1. Similarly, the calendar year weighting scheme does not assign weights to people who do not have an interview covering the control date. This group consists of (a) people who enter the sample universe after the first wave of interviewing for the calendar year and (b) people who were in the sample universe in the first wave of interviewing in the calendar year but did not have an interview covering the control date. For example, newborn infants and people leaving institutions who are entering the sample universe after Wave 1 are assigned full panel and calendar year 1 weights of zero. Note that the same people will receive positive calendar year 2 (CY2) weights if they are in the sample universe in the first wave of interviewing for CY2 and have an interview covering the control date for CY2.

The final panel and calendar year weights are constructed from the following three components:

1. **Initial weight.** This weight is constructed from the components of the cross-sectional weights at the start of the panel and calendar year weighting periods before the second-stage calibration adjustment.
2. **Nonresponse adjustment factors.** These factors account for noninterviewed eligible sample persons not already accounted for in the noninterview adjustment component of the initial weight. The adjustment classes are similar to those used in the Wave 2+ nonresponse adjustment factors.
3. **Second-stage calibration factors.** These factors are determined by a process similar to that used for reference and interview month weighting. The control totals used for the calendar year weights are the population estimates for the control date of the relevant year. Those for the full panel weight are the population estimates for a designated date in the first wave of the panel.

Footnotes

[1] Interview month weights were not computed for the 1996 Panel.

[2] Persons subjected to Type Z imputation receive weights, although they are not respondents.

[3] This adjustment has been used since Wave 5 of the 1984 Panel.

[4] Known as the control card information before the 1996 Panel, when computer-assisted interviewing (CAI) began.

[5] The calendar year control dates are January 1 for the given calendar year. The exception is calendar year 1996 for the 1996 Panel. Its control date is currently March 1, 1996. This would change to January 1 should there be imputation for January and February data.

Last Revised: March 1, 2016