UnderStandingAmericaStudy

PRODUCING NATIONALLY REPRESENTATIVE YEARLY ESTIMATES OF VARIABLES

Marco Angrisani, Francisco Arce-Perez

11/17/2023

OVERVIEW

In September 2023, the UAS team added a new datafile containing weights that can be used to create nationally representative values for the core surveys in the UAS.

The UAS core surveys are fielded on a rolling basis for periods that span approximately two years. Hence, the standard weights produced for those surveys (and the ones produced for the Comprehensive File (CF)) can be used to calculate estimates of population quantities for a 2-year period.

This note describes the newly created CF-yearly weight file, which can be used to produce nationally representative statistics by calendar year when used in conjunction with the Comprehensive File. The file can be accessed in the UAS file [here](#).

DESCRIPTION

Respondents answer the core surveys in the Comprehensive File on a rolling basis. They are invited to complete new versions of the surveys approximately two years after their invitation to the prior round. Because panel members join the panel at different times, responses to the core surveys are continually added. Consequently, using observations collected in a given year, it is possible to produce population-level statistics for each calendar year.

The Comprehensive File includes wave-specific weights (r12final_weight, r13final_weight, r14final_weight, r15final_weight), which are constructed to make the sample of all respondents who participated in a given wave representative of the adult population in the U.S. The dimensions used to compute these weights are sex, race/ethnicity, age, education, and geographic location (Census regions). A complete description of the UAS weighting procedure can be found here. By construction, these weights are designed to produce population-level estimates spanning the (approximately) two-year period of each wave.

Yearly weights were created to generate population-level statistics for each calendar year. To this end, we adopted the general UAS weighting procedure, also used to create Comprehensive File's wave-specific weights, which we applied to make the sample of respondents who completed a certain core survey in a given year representative of the U.S. adult population in that year. The yearly weights dataset can be downloaded here and linked to the Comprehensive File using the individual identifier variable: *uasid*.

The yearly weights are constructed separately for each "topic" in the Comprehensive File. There are distinct weight variables for each topic because different panelists respond to different surveys each calendar year. For example, a respondent might have answered the "What do people know about Social Security" survey (the source for the k-prefix variables) in 2015 and 2017, but answered the HRS surveys (the foundation for the r-prefix variables) in 2016 and 2018. As a result, that respondent will have a non-zero weight for kweight in 2015 and 2017 and a non-zero rweight in 2016 and 2018. As a result, that respondent will have a non-zero weight for kweight in 2015 and 2017 – allowing to produce population-level statistics using variables from the "What do people know about Social Security" survey for the years 2015 and 2017 – and a non-zero rweight in 2016 and 2018 – allowing to produce population-level statistics using variables from the HRS surveys for the years 2015 and 2017.

As mentioned above, the yearly weights were constructed using the general UAS weighting procedure described in detail [here](#). This consists of a two-step process. In the first step, we compute base weights, which correct for unequal probabilities of sampling UAS members. In the second step, we start from the individual-specific base weights and implement an iterative raking/trimming algorithm to generate post-stratification weights. The final weights align the sample of **each core survey in each year** to the U.S. adult population along the following dimensions: sex (Male/Female), race/ethnicity (White/Black/Other/Hispanic/Native American), age (18-39/40-49/50-59/60+), education (High School or Less/Some College/Bachelor or More), and residence in Census regions (Northeast/Midwest/South/West, excluding California/California, excluding Los Angeles County/Los Angeles County). Note that we defined the variables used for post-stratification to account for the over-representation of Native Americans, California residents, and Los Angeles County residents in the UAS. Population benchmarks for each year are obtained from the Basic Monthly Current Population Survey, combining all the months within a year.

> **Warning!**
>
> Before drawing conclusions about comparisons across time, it is important to ensure that the variables are indeed comparable across time.
>
> As described in the Comprehensive File codebook, whenever there is a change in the wording of a question or there is a change in the response options in the surveys of topics "k", "l", "f", "c", "p", "w", "a", "v", n", we add a suffix to the variable name in the Comprehensive File (such as "ans13"). This helps the user to avoid making comparisons of variables that may not be properly comparable.