
Subject: Weighting for same cohort across multiple surveys

Posted by [cportner](#) on Mon, 14 Jun 2021 16:53:53 GMT

[View Forum Message](#) <> [Reply to Message](#)

I am looking at outcomes for women, aged 40+, who are born within a given time period. To get a sufficient sample size, I pool the four NFHS surveys. Say, I am looking at women born 1958-1967. In that case, the women cover three different surveys (a simple tabulation of number of women by year of birth and survey using only urban women is below), with some cohorts drawing women from one survey and some cohorts drawing women from two surveys. My problem is what to do with the weights across the surveys since the weight are specific to each survey. I assume I need to de-normalize by survey, but once that is done would I need to calculate new weights to combine the surveys?

```
> table(india_40$year_of_birth, india_40$version_number)
```

```
      23  42  52  74
1942  36   0   0   0
1943 376   0   0   0
1944 515   0   0   0
1945 493   0   0   0
1946 494   0   0   0
1947 742   0   0   0
1948 598   0   0   0
1949 564 271   0   0
1950 684 565   0   0
1951 628 555   0   0
1952 773 564   0   0
1953 205 745   0   0
1954   0 702   0   0
1955   0 643   0   0
1956   0 794 423   0
1957   0 712 835   0
1958   0 893 811   0
1959   0 262 817   0
1960   0   5 1258   0
1961   0   0 1009   0
1962   0   0 1041   0
1963   0   0 1185   0
1964   0   0 1183   0
1965   0   0 1539 824
1966   0   0  529 2807
1967   0   0   0 3958
1968   0   0   0 3682
1969   0   0   0 4231
1970   0   0   0 5710
1971   0   0   0 4037
1972   0   0   0 4215
1973   0   0   0 4316
```

1974	0	0	0	4670
1975	0	0	0	4876
1976	0	0	0	1541

Thank you for the help,

Claus

Subject: Re: Weighting for same cohort across multiple surveys

Posted by [Bridgette-DHS](#) on Mon, 14 Jun 2021 19:37:15 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

When you do this, with pooled data, you are simply recoding v010 (year of birth). You are not following individuals. You still have different cases (independent cases, in fact), with different weights within and across surveys. I would just stay with the original weights. I suggest that you first conceptualize the analysis without worrying about the weights, and then consider whether to re-calculate the weights.

For svyset, you need to construct unique id codes for clusters and strata, as has been described in earlier forum postings.
